Contents of all Volumes

Contributors to all Volumes xvii

Preface xxxvii

Introduction xxxix

Editors-in-Chief xliii

Volume Editors xlv

Volume 1 Natural Products Structural Diversity-I Secondary Metabolites: Organization and Biosynthesis

01 Overview and Introduction 1
Craig A. Townsend, *Johns Hopkins University, Baltimore, MD, USA*

02 Unsaturated Fatty Acids 5
Peter H. Buist, *Carleton University, ON, Canada*

03 Prostaglandin Endoperoxide Synthases: Structure, Function, and Synthesis of Novel Lipid Signaling Molecules 35
Melissa V. Turman and Lawrence J. Marnett, *Vanderbilt University School of Medicine, Nashville, TN, USA*

04 Mycolic Acid/Cyclopropane Fatty Acid/Fatty Acid Biosynthesis and Health Relations 65
David B. Kastrinsky, Nicholas S. McBride, Keriann M. Backus, Jason J. LeBlanc, and Clifton E. Barry, III, *National Institutes of Health, Bethesda, MD, USA*

05 Microbial Type III Polyketide Synthases 147
Yohei Katsuyama and Sueharu Horinouchi, *The University of Tokyo, Bunkyo-ku, Tokyo, Japan*

06 Plant Type III PKS 171
Hiroyuki Morita and Ikuro Abe, *The University of Tokyo, Hongo, Tokyo, Japan*
Hiroshi Noguchi, *University of Shizuoka, Yada, Shizuoka, Japan*

07 Type II PKS 227
Jürgen Rohr, *University of Kentucky, Lexington, KY, USA*
Christian Hertweck, *Leibniz Institute for Natural Product Research and Infection Biology, HKI, Jena, Germany*

08 Structural Enzymology of Polyketide Synthase: The Structure–Sequence–Function Correlation 305
Tyler Paz Korman, Brian Ames, and Shiou-Chuan (Sheryl) Tsai, *University of California, Irvine, CA, USA*

09 Fungal Type I Polyketides 347
Russell J. Cox and Thomas J. Simpson, *University of Bristol, Bristol, UK*

10 Type I Modular PKS 385
Alison M. Hill, *University of Exeter, Exeter, UK*
James Staunton, *University of Cambridge, Cambridge, UK*

11 NRPS/PKS Hybrid Enzymes and Their Natural Products 453
Christopher M. Rath, Jamie B. Scaglione, Jeffrey D. Kittendorf, and David H. Sherman, *Life Sciences Institute University of Michigan, Ann Arbor, MI, USA*

12 Mevalonate Pathway in Bacteria and Archaea 493
Tomohisa Kuzuyama, *The University of Tokyo, Tokyo, Japan*
Hisashi Hemmi, *Nagoya University, Nagoya, Japan*
Shunji Takahashi, *RIKEN, Advanced Science Institute, Saitama, Japan*

13 Methylerythritol Phosphate Pathway 517
Michel Rohmer, *Université de Strasbourg/CNRS, Strasbourg, France*

14 Prenyltransferase 557
Hirofumi Kurokawa and Tanetoshi Koyama, *Tohoku University, Sendai, Japan*

15 Advances in the Enzymology of Monoterpene Cyclization Reactions 585
Edward M. Davis, *Washington State University, Pullman, WA, USA*

16 Sesquiterpenes 609
Joe Chappell, *University of Kentucky, Lexington, KY, USA*
Robert M. Coates, *University of Illinois, Urbana, IL, USA*

17 Diterpenes 643
Tomonobu Toyomasu and Takeshi Sassa, *Yamagata University, Tsuruoka, Japan*

18 Triterpenes 673
Tetsuo Kushiro and Yutaka Ebizuka, *The University of Tokyo, Tokyo, Japan*

19 Bacterial Squalene Cyclase 709
Ikuro Abe, *The University of Tokyo, Hongo, Tokyo, Japan*

20 Carotenoids 733
Norihiko Misawa, *Kirin Holdings Co. Ltd., Nonoichi-machi, Ishikawa, Japan*

21 Sterol and Steroid Biosynthesis and Metabolism in Plants and Microorganisms 755
Hubert Schaller, *Université de Strasbourg and CNRS, Strasbourg, France*

22 Isoprenoid in Actinomycetes 789
Tohru Dairi, *Toyama Prefectural University, Toyama, Japan*

23 Lignans (Neolignans) and Allyl/Propenyl Phenols: Biogenesis, Structural Biology, and Biological/Human Health Considerations 815
Daniel G. Vassão, Kye-Won Kim, Laurence B. Davin, and Norman G. Lewis, *Washington State University, Pullman, WA, USA*

24 Plant Phenolics: Phenylpropanoids 929
Shin-ichi Ayabe, Hiroshi Uchiyama, Toshio Aoki, and Tomoyoshi Akashi, *Nihon University, Fujisawa, Japan*

25 Alkaloids 977
Sarah E. O'Connor, *Massachusetts Institute of Technology, Cambridge, MA, USA*

Volume 2 Natural Products Structural Diversity-II Secondary Metabolites: Sources, Structures and Chemical Biology

01 Overview and Introduction 1
Bradley S. Moore, *University of California at San Diego, La Jolla, CA, USA*
Phillip Crews, *University of California at Santa Cruz, Santa Cruz, CA, USA*

02 Terrestrial Plants as a Source of Novel Pharmaceutical Agents 5
Gordon M. Cragg and David J. Newman, *NCI-Frederick, Frederick, MD, USA*
David G. I. Kingston, *Virginia Polytechnic Institute & State University, Blacksburg, VA, USA*

03 Marine Macroalgal Natural Products 41
E. Paige Stout and Julia Kubanek, *Georgia Institute of Technology, Atlanta, GA, USA*

04 Insect Natural Products 67
Matthew Gronquist, *SUNY College at Fredonia, Fredonia, NY, USA*
Frank C. Schroeder, *Cornell University, Ithaca, NY, USA*

05 Terrestrial Microorganisms – Filamentous Bacteria 109
Sheo B. Singh, *Merck Research Laboratories, Rahway, NJ, USA*
Olga Genilloud and Fernando Peláez, *CIBE, Merck Sharp & Dohme de España, Madrid, Spain*

06 The Natural Products Chemistry of Cyanobacteria 141
Kevin Tidgewell, Benjamin R. Clark, and William H. Gerwick, *University of California San Diego, La Jolla, CA, USA*

07 Myxobacteria – Unique Microbial Secondary Metabolite Factories 189
Silke C.Wenzel and Rolf Müller, *Saarland University, Saarbrücken, Germany*

08 Natural Product Diversity from Marine Fungi 223
Rainer Ebel, *University of Aberdeen, Aberdeen, UK*

09 Bioactive Metabolites from Marine Dinoflagellates 263
Jun'ichi Kobayashi and Takaaki Kubota, *Hokkaido University, Sapporo, Japan*

10 Marine Invertebrates: Sponges 327
Yoichi Nakao, *Waseda University, Tokyo, Japan*
Nobuhiro Fusetani, *Hokkaido University, Hakodate, Japan*

11 The Natural Products Chemistry of the Gorgonian Genus Pseudopterogorgia (Octocorallia: Gorgoniidae) 363
Jeffrey Marrero, Ileana I. Rodríguez, and Abimael D. Rodríguez, *University of Puerto Rico, San Juan, PR, USA*

12 Exploiting Genomics for New Natural Product Discovery in Prokaryotes 429
Christophe Corre and Gregory L. Challis, *University of Warwick, Coventry, UK*

13 Unlocking Environmental DNA Derived Gene Clusters Using a Metagenomics Approach 455
Jeffrey H. Kim, T. Luke Simmons, and Sean F. Brady, *The Rockefeller University, New York, NY, USA*

14 The Chemistry of Symbiotic Interactions 475
Jörn Piel, *University of Bonn, Bonn, Germany*

15 Natural Peptide Toxins 511
Frank Marí, *Florida Atlantic University, Boca Raton, FL, USA*
Jan Tytgat, *University of Leuven, Leuven, Belgium*

16 Cyanobactins – Ubiquitous Cyanobacterial Ribosomal Peptide Metabolites 539
Mohamed S. Donia and Eric W. Schmidt, *University of Utah, Salt Lake City, UT, USA*

17 The Role of Synthesis and Biosynthetic Logic 559
Yousong Ding and David H. Sherman, *The University of Michigan, Ann Arbor, MI, USA*

18 Missasigned Structures: Case Examples from the Past Decade 581
Taro Amagata, *San Francisco State University, San Francisco, CA, USA*

19 Natural Products of Therapeutic Importance 623
David J. Newman and Gordon M. Cragg, *NCI-Frederick, Frederick, MD, USA*

20 Natural Products as Probes of Selected Targets in Tumor Cell Biology and Hypoxic Signaling 651
Dale G. Nagle and Yu-Dong Zhou, *University of Mississippi, University, MS, USA*

Volume 3 Development & Modification of Bioactivity

01 Overview and Introduction 1
Rob Verpoorte, *Leiden University, Leiden, The Netherlands*

02 Natural Products as Lead Sources for Drug Development 5
Stefan Wetzel, Hugo Lachance, and Herbert Waldmann, *Max Planck Institute of Molecular Physiology, Dortmund, Germany*

03 Topical Chemical Space Relation to Biological Space 47
Anders Backlund, *Uppsala University, Uppsala, Sweden*

04 The NAPRALERT Database as an Aid for Discovery of Novel Bioactive Compounds 81
James G. Graham and Norman R. Farnsworth, *University of Illinois at Chicago, Chicago, IL, USA*

05 Plant Diversity from Brazilian Cerrado and Atlantic Forest as a Tool for Prospecting Potential Therapeutic Drugs 95
Dulce Helena Siqueira Silva, Ian Castro-Gamboa, and Vanderlan Da Silva Bolzani, *São Paulo State University, Araraquara, São Paulo, Brazil*

06 Nature as Source of Medicines; Novel Drugs from Nature; Screening for Antitumor Activity 135
Gordon M. Cragg and David J. Newman, *NCI-Frederick, Frederick, MD, USA*

07 The Identification of Bioactive Natural Products by High Throughput Screening (HTS) 177
Vicky M. Avery, David Camp, Anthony R. Carroll, Ian D. Jenkins, and Ronald J. Quinn, *Griffith University, Brisbane, QLD, Australia*

08 Natural Products Drug Discovery 205
Giovanni Appendino, *Università del Piemonte Orientale, Novara, Italy*
Gabriele Fontana, *Indena S.p.A., Milano, Italy*
Federica Pollastro, *Università del Piemonte Orientale, Novara, Italy*

09 Natural Product-Based Biopesticides for Insect Control 237
Azucena Gonzalez-Coloma, *Instituto de Ciencias Agrarias-CCMA, Madrid, Spain*
Matìas Reina, Carmen E. Diaz, and Braulio M. Fraga, *Instituto de Productos Naturales y Agrobiologia, Tenerife, Spain*

10 Natural Products as Sweeteners and Sweetness Modifiers 269
A. Douglas Kinghorn, Young-Won Chin, and Li Pan, *The Ohio State University, Columbus, OH, USA*
Zhonghua Jia, *Givaudan Flavors Corporation, Cincinnati, OH, USA*

11 Chemistry of Cosmetics 317
Masahiro Ota and Mineyuki Yokoyama, *Shiseido Co., Ltd., Yokohama, Japan*

12 Ethnopharmacology and Drug Discovery 351
Michael Heinrich, *University of London, London, UK*

13 Chinese Traditional Medicine 383
Min Yang, Sijia Tao, Shuhong Guan, Xiaohui Wu, Pingping Xu, and De-an Guo, *Shanghai Institute of Materia Medica, Shanghai, China*

14 Ayurveda in Modern Medicine: Development and Modification of Bioactivity 479
Pulok K. Mukherjee, M. Venkatesh, and Arunava Gantait, *Jadavpur University, Kolkata, India*

15 Biologically Active Compounds in Food Products and Their Effects on Obesity and Diabetes 509
Renger F. Witkamp, *Wageningen University and TNO Quality of Life, Wageningen, The Netherlands*

16 Chemistry of Flavonoid-Based Colors in Plants 547
Øyvind M. Andersen and Monica Jordheim, *University of Bergen, Bergen, Norway*

17 Production of Pharmaceuticals by Plant Tissue Cultures 615
Toshiya Muranaka, *Yokohama City University, Yokohama, Japan, RIKEN Plant Science Center, Yokohama, Japan*
Kazuki Saito, *Chiba University, Chiba, Japan, RIKEN Plant Science Center, Yokohama, Japan*

18 Plant Secondary Metabolism Engineering: Methods, Strategies, Advances, and Omics 629
Rafael Zárate, *The University of La Laguna, Tenerife, Spain*

19 Biotransformation of Monoterpenoids 669
Yoshiaki Noma and Yoshinori Asakawa, *Tokushima Bunri University, Yamashiro-cho, Tokushima, Japan*

20 Biotransformation of Sesquiterpenoids 803
Yoshinori Asakawa and Yoshiaki Noma, *Tokushima Bunri University, Yamashiro-cho, Tokushima, Japan*

21 Biotransformation of Di- and Triterpenoids, Steroids, and Miscellaneous Synthetic Substrates 893
Yoshinori Asakawa and Yoshiaki Noma, *Tokushima Bunri University, Yamashiro-cho, Tokushima, Japan*

22 Beer Flavor 967
Leen C. Verhagen, *University of Leiden, Leiden, The Netherlands*

23 Chemistry of Tea 999
Ulrich H. Engelhardt, *Institut für Lebensmittelchemie, Braunschweig, Germany*

24 Chemistry of Cannabis 1033
Arno Hazekamp, Justin T. Fischedick, Mónica Llano Díez, Andrea Lubbe, and Renee L. Ruhaak, *Leiden University, Leiden, The Netherlands*

25 Chemistry of Coffee 1085
S. Oestreich-Janzen, *CAFEA GmbH, Hamburg, Germany*

26 Chemistry of Wine 1119
Véronique Cheynier, Rémi Schneider, Jean-Michel Salmon, and Hélène Fulcrand, *Montpellier Research Centre, Montpellier, France*

27 Trees: A Remarkable Biochemical Bounty 1173
Ann M. Patten, Daniel G. Vassão, Michael P. Wolcott, Laurence B. Davin, and Norman G. Lewis, *Washington State University, Pullman, WA, USA*

28 The Chemistry of *Arabidopsis thaliana* 1297
M. Soledade, C. Pedras and Qingan Zheng, *University of Saskatchewan, Saskatoon, SK, Canada*

Volume 4 Chemical Ecology

01 Overview and Introduction 1
Kenji Mori, *The University of Tokyo, Tokyo, Japan*

02 Plant Hormones 9
Isomaro Yamaguchi, *Maebashi Institute of Technology, Gunma, Japan*
Jerry D. Cohen and Angela H. Culler, *University of Minnesota, Minneapolis, USA*
Marcel Quint, *Leibniz Institute of Plant Biochemistry, Halle, Germany*
Janet P. Slovin, *United States Department of Agriculture, Washington, DC, USA*
Masatoshi Nakajima, *The University of Tokyo, Tokyo, Japan*
Shinjiro Yamaguchi, Hitoshi Sakakibara, and Takeshi Kuroha, *RIKEN Plant Science Center, Kanagawa, Japan*
Nobuhiro Hirai, *Kyoto University, Kyoto, Japan*
Takao Yokota, *Teikyo University, Utsunomiya, Japan*
Hiroyuki Ohta and Yuichi Kobayashi, *Tokyo Institute of Technology, Tokyo, Japan*
Hitoshi Mori and Yoji Sakagami, *Nagoya University, Nagoya, Japan*

03 Insect Hormones 127
E. David Morgan, *Keele University, Staffordshire, UK*

04 Pheromones of Terrestrial Invertebrates 153
Wittko Francke, *University of Hamburg, Hamburg, Germany*
Stefan Schulz, *Technische Universität Braunschweig, Braunschweig, Germany*

05 Pheromones in Vertebrates 225
Peter W. Sorensen and Thomas R. Hoye, *University of Minnesota, Minneapolis, MN, USA*

06 Pheromones of Marine Invertebrates and Algae 263
Masaki Kita, *University of Tsukuba, Tsukuba, Japan*
Makoto Kitamura and Daisuke Uemura, *Keio University, Yokohama, Japan*

07 Cell-to-Cell Communications among Microorganisms 283
Sueharu Horinouchi, *The University of Tokyo, Tokyo, Japan*
Kenji Ueda, *Nihon University, Kanagawa, Japan*
Jiro Nakayama, *Kyushu University, Fukuoka, Japan*
Tsukasa Ikeda, *Utsunomiya University, Tochigi, Japan*

08 Chemical Defence and Toxins of Plants 339
Hisakazu Yamane, *The University of Tokyo, Tokyo, Japan*
Kotaro Konno, *National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan*
Maurice Sabelis, *University of Amsterdam, Amsterdam, The Netherlands*
Junji Takabayashi, *Kyoto University, Otsu, Shiga, Japan*
Takeshi Sassa, *Yamagata University, Tsuruoka, Japan*
Hideaki Oikawa, *Hokkaido University, Sapporo, Japan*

09 Chemical Defense and Toxins of Lower Terrestrial and Freshwater Animals 387
Konrad Dettner, *University of Bayreuth, Bayreuth, Germany*

10 Toxins of Microorganisms 411
Shohei Sakuda, *University of Tokyo, Tokyo, Japan*
Makoto Kimura, *RIKEN, Saitama, Japan*

11 Terrestrial Natural Products as Antifeedants 457
Falko P. Drijfhout and E. David Morgan, *Keele University, Staffordshire, UK*

12 Marine Natural Products as Antifeedants 503
Mary J. Garson, *The University of Queensland, Brisbane, QLD, Australia*

13 Allelochemicals for Plant–Plant and Plant–Microbe Interactions 539
Koichi Yoneyama, *Utsunomiya University, Utsunomiya, Japan*
Masahiro Natsume, *Tokyo University of Agriculture and Technology, Tokyo, Japan*

14 Allelochemicals in Plant–Insect Interactions 563
Keiichi Honda and Hisashi Ômura, *Hiroshima University, Higashihiroshima, Japan*
Masatoshi Hori, *Tohoku University, Sendai, Japan*
Yooichi Kainoh, *University of Tsukuba, Tsukuba, Japan*

15 Human–Environment Interactions (1): Flavor and Fragrance 595
Masashi Ishikawa, Yasuhiro Warita, Eisuke Takahisa, and Yasutaka Ohkubo, *T. Hasegawa Co., Ltd., Tokyo, Japan*

16 Human–Environment Interactions – Taste 631
Kunisuke Izawa, Yusuke Amino, Masanori Kohmura, Yoichi Ueda, and Motonaka Kuroda, *Ajinomoto Co., Inc., Kawasaki, Japan*

Volume 5 Amino Acids, Peptides and Proteins

01 Overview and Introduction 1
John C. Vederas, *University of Alberta, Edmonton, AB, Canada*

02 Nonprotein L-Amino Acids 5
Sabesan Yoganathan and John C. Vederas, *University of Alberta, Edmonton, AB, Canada*

03 Novel Enzymes for Biotransformation and Resolution of Alpha-Amino Acids 71
Paul C. Engel and Francesca Paradisi, *University College Dublin, Dublin, Ireland*

04 1-Aminocyclopropane-1-Carboxylate Synthase, an Enzyme of Ethylene Biosynthesis 91
Małgorzata Jakubowicz and Witold Nowak, *Adam Mickiewicz University, Poznan, Poland*

05 Specific and Nonspecific Incorporation of Selenium into Macromolecules 121
William T. Self, *University of Central Florida, Orlando, FL, USA*

06 Protein Toxins from Bacteria 149
Holger Barth, *University of Ulm Medical Center, Ulm, Germany*
Bradley G. Stiles, *Wilson College, Chambersburg, PA, USA*; *US Army Medical Research Institute of Infectious Diseases, Fort Detrick, MD, USA*

07 Host Defense Peptides: Bridging Antimicrobial and Immunomodulatory Activities 175
Jason Kindrachuk, Anastasia Nijnik, and R. E. W. Hancock, *University of British Columbia, Vancouver, BC, Canada*

08 Biosynthesis and Mode of Action of Lantibiotics 217
Lisa E. Cooper, Bo Li, and Wilfred A. van der Donk, *University of Illinois at Urbana-Champaign, Urbana, IL, USA*

09 Plant Peptide Toxins from Nonmarine Environments 257
Jan-Christoph Westermann and David J. Craik, *University of Queensland, Brisbane, QLD, Australia*

10 Therapeutic Value of Peptides from Animal Venoms 287
Nicolas Andreotti, Besma Jouirou, Stephanie Mouhat, Ludovic Mouhat, and Jean-Marc Sabatier, *Université de la Méditerranée – Ambrilia Biopharma Inc., Marseille Cedex, France*

11 Signal Transduction in Gram-Positive Bacteria by Bacterial Peptides 305
Ingolf F. Nes, Ola Johnsborg, and Dzung Bao Diep, *Norwegian University of Life Sciences, Aas, Norway*

12 A New Generation of Artificial Enzymes: Catalytic Antibodies or 'Abzymes' 323
Rémy Ricoux and Jean-Pierre Mahy, *Université de Paris-sud 11, Orsay Cedex, France*

13 Recent Progress on Understanding Ribosomal Protein Synthesis 353
Ute Kothe, *University of Lethbridge, Lethbridge, AB, Canada*

14 Glutaminyl-tRNA and Asparaginyl-tRNA Biosynthetic Pathways 383
Jonathan L. Huot, Jacques Lapointe, and Robert Chênevert, *Université Laval, Québec, QC, Canada*
Marc Bailly and Daniel Kern, *Institut de Biologie Moléculaire et Cellulaire, Strasbourg, France*

15 Posttranslational Modification of Proteins 433
Keith D. Green and Sylvie Garneau-Tsodikova, *University of Michigan, Ann Arbor, MI, USA*

16 Collagen Formation and Structure 469
Hans Peter Bächinger, Kazunori Mizuno, Janice A. Vranka, and Sergei P. Boudko, *Oregon Health & Science University, Portland, OR, USA*

17 Lipidation of Peptides and Proteins 531
Kristina Görmer and Herbert Waldmann, *Max Planck Institute of Molecular Physiology and Technische Universität Dortmund, Dortmund, Germany*
Luc Brunsveld, *Max Planck Institute of Molecular Physiology and Chemical Genomics Centre of the Max Planck Society, Dortmund, Germany*

18 Genetic Incorporation of Unnatural Amino Acids into Proteins 587
Angela R. Parrish and Lei Wang, *The Salk Institute for Biological Studies, La Jolla, CA, USA*

19 Nonribosomal Peptide Synthetases 619
Timothy J. Montavon and Steven D. Bruner, *University of Florida, Gainesville, FL, USA*

20 The Properties, Formation, and Biological Activity of 2,5-Diketopiperazines 657
P. J. Milne and G. Kilian, *Nelson Mandela Metropolitan University, Port Elizabeth, South Africa*

21 Ubiquitin-Dependent Protein Degradation 699
Ashok N. Hegde, *Wake Forest University Health Sciences, Winston Salem, NC, USA*

Volume 6 Carbohydrates, Nucleosides & Nucleic Acids

01 Overview and Introduction 1
Robert Woodward and Peng George Wang, *The Ohio State University, Columbus, OH, USA*

02 Enzymatic Synthesis of Complex Carbohydrates 5
Wei Zhao and Tiehai Li, *Nankai University, Tianjin, China*
Robert Woodward, Chengfeng Xia, and Peng George Wang, *The Ohio State University, Columbus, OH, USA*
Wanyi Guan, *Shandong University, Shandong, China*

03 New Strategies for Glycopeptide, Neoglycopeptide, and Glycoprotein Synthesis 55
Ashraf Brik, *Ben-Gurion University of the Negev, Beer Sheva, Israel*

04 Carbohydrate Vaccines 91
Qianli Wang and Zhongwu Guo, *Wayne State University, Detroit, MI, USA*

05 MS-Based Glycoanalysis 123
Kay-Hooi Khoo, *Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan*

06 Glycoanalysis of Bacterial Glycome 157
Chuan Wang, Camille L. A. Hamula, and Xing-Fang Li, *University of Alberta, Edmonton, AB, Canada*
Xin Liu and Jianjun Li, *National Research Council of Canada, Ottawa, ON, Canada*

07 Chemical Glycobiology 175
Chad M. Whitman and Michelle R. Bond, *Stanford University, Stanford, CA, USA*
Jennifer J. Kohler, *University of Texas Southwestern Medical Center, Dallas, TX, USA*

08 Alkaloid Glycosidase Inhibitors 225
Alan D. Elbein, *University of Arkansas for Medical Sciences, Little Rock, AR, USA*
Russell J. Molyneux, *Western Regional Research Center, ARS-USDA, Albany, CA, USA*

09 Molecular Probes for Protein Glycosylation 261
Howard C. Hang, *The Rockefeller University, New York, NY, USA*

10 O Antigen Biosynthesis 297
Miguel A. Valvano, *The University of Western Ontario, London, ON, Canada*

11 Biosynthesis of Complex Mucin-Type *O*-Glycans 315
Inka Brockhausen, *Queen's University, Kingston, ON, Canada*

12 Bacterial Protein Glycosylation 351
Amirreza Faridmoayer and Mario F. Feldman, *University of Alberta, Edmonton, AB, Canada*

13 Structure and Biosynthesis of the Mycobacterial Cell Wall 381
Dean C. Crick, Delphi Chatterjee, Michael S. Scherman, and Michael R. McNeil, *Colorado State University, Fort Collins, CO, USA*

14 Structure, Biosynthesis, and Function of Glycosaminoglycans 407
Courtney L. Jones and Jian Liu, *University of North Carolina, Chapel Hill, NC, USA*
Ding Xu, *University of California, San Diego, CA, USA*

15 Biochemistry and Molecular Biology of Glycogen Synthesis in Bacteria and Mammals and Starch Synthesis in Plants 429
Jack Preiss, *Michigan State University, East Lansing, MI, USA*

16 Celluloses 493
Rajai H. Atalla, *Cellulose Sciences International, Madison, WI, USA*
Akira Isogai, *University of Tokyo, Tokyo, Japan*

17 Vascular Plant Lignification: Biochemical/Structural Biology Considerations of Upstream Aromatic Amino Acid and Monolignol Pathways 541
Dhrubojyoti D. Laskar, Oliver R. A. Corea, Ann M. Patten, ChulHee Kang, Laurence B. Davin, and Norman G. Lewis, *Washington State University, Pullman, WA, USA*

18 Proanthocyanidins: Chemistry and Biology 605
Daneel Ferreira, Jannie P. J. Marais, Christina M. Coleman, and Desmond Slade, *University of Mississippi, Oxford, MS, USA*

19 Nucleoside Analogues 663
Darrell R. Davis, *The University of Utah, Salt Lake City, UT, USA*

20 RNA Modifying Enzymes 683
George A. Garcia, Julie K. Hurt, and Yi-Chen Chen, *University of Michigan, Ann Arbor, MI, USA*

21 Riboswitches 743
Tina M. Henkin, *The Ohio State University, Columbus, OH, USA*

Volume 7 Cofactors

01 Overview and Introduction 1
Tadhg P. Begley, *Texas A&M University, College Station, TX, USA*

02 Riboflavin Biosynthesis 3
Markus Fischer, *University of Hamburg, Hamburg, Germany*
Adelbert Bacher, *Technical University of Munich, Garching, Germany*

03 Flavin-Dependent Enzymes 37
Rebecca L. Fagan and Bruce A. Palfey, *University of Michigan Medical School, Ann Arbor, MI, USA*

04 Biotechnology of Riboflavin Production 115
Hans-Peter Hohmann, *DSM Nutritional Products, Kaiseraugst, Switzerland*
Klaus-Peter Stahmann, *Fachhochschule Lausitz, Senftenberg, Germany*

05 The Use of Subsystems to Encode Biosynthesis of Vitamins and Cofactors 141
Andrei L. Osterman, *Burnham Institute for Medical Research, La Jolla, CA, USA*
Ross Overbeek, *Fellowship for Interpretation of Genomes, Burr Ridge, IL, USA*
Dmitry A. Rodionov, *Burnham Institute for Medical Research, La Jolla, CA, USA*

06 Biosynthesis of Biotin 161
Andrée Marquet, *Université Pierre et Marie Curie, Paris, France*

07 Lipoic Acid Biosynthesis and Enzymology 181
Elizabeth S. Billgren, Robert M. Cicchillo, Natasha M. Nesbitt, and Squire J. Booker, *The Pennsylvania State University, University Park, PA, USA*

08 Genomics and Enzymology of NAD Biosynthesis 213
Leonardo Sorci, Oleg Kurnasov, Dmitry A. Rodionov, and Andrei L. Osterman, *Burnham Institute for Medical Research, La Jolla, CA, USA*

09 Pyridoxal Phosphate Biosynthesis 259
Jeremiah W. Hanes, Steven E. Ealick, and Tadhg P. Begley, *Cornell University, Ithaca, NY, USA*
Ivo Tews, *Heidelberg University Biochemistry Center, Heidelberg, Germany*

10 Pyridoxal 5′-Phosphate-Dependent Enzymes: Catalysis, Conformation, and Genomics 273
Samanta Raboni, Francesca Spyrakis, Barbara Campanini, Alessio Amadasi, Stefano Bettati, Alessio Peracchi, and Andrea Mozzarelli, *University of Parma, Parma, Italy*; *Present address: University of Padua, Padua, Italy*
Roberto Contestabile, *University of Rome 'La Sapienza,' Rome, Italy*

11 Coenzyme A Biosynthesis and Enzymology 351
Erick Strauss, *Stellenbosch University, Stellenbosch, South Africa*

12 Menaquinone/Ubiquinone Biosynthesis and Enzymology 411
R. Meganathan, *Northern Illinois University, DeKalb, IL, USA*

13 Biosynthesis of Heme and Vitamin B_{12} 445
Gunhild Layer and Dieter Jahn, *Technical University of Braunschweig, Braunschweig, Germany*
Evelyne Deery, Andrew D. Lawrence, and Martin J. Warren, *University of Kent, Kent, UK*

14 Cobalamin Coenzymes in Enzymology 501
Perry Allen Frey, *University of Wisconsin-Madison, Madison, WI, USA*

15 Thiamin Biosynthesis 547
Tadhg P. Begley and Steven E. Ealick, *Cornell University, Ithaca, NY, USA*

16 Thiamin Enzymology 561
Frank Jordan, *Rutgers, The State University of New Jersey, Newark, NJ, USA*

17 The Biosynthesis of Folate and Pterins and Their Enzymology 599
Markus Fischer, *University of Hamburg, Hamburg, Germany*
Beat Thöny, *University Children's Hospital, Zürich, Switzerland*
Silke Leimkühler, *University of Potsdam, Potsdam, Germany*

18 Cofactor Catabolism 649
Tathagata Mukherjee, Kathryn M. McCulloch, Steven E. Ealick, and Tadhg P. Begley, *Cornell University, Ithaca, NY, USA*

19 Protein-Derived Cofactors 675
Victor L. Davidson, *University of Mississippi Medical Center, Jackson, MS, USA*

20 Biosynthesis of the Methanogenic Coenzymes 711
Laura L. Grochowski and Robert H. White, *Virginia Polytechnic Institute and State University, Blacksburg, VA, USA*

Volume 8 Enzymes and Enzyme Mechanisms

01 Overview and Introduction 1
Christian P. Whitman, *The University of Texas, Austin, TX, USA*

02 Evolution and the Enzyme 9
Shelley D. Copley, *University of Colorado at Boulder, Boulder, CO, USA*

03 Enzyme Promiscuity – Evolutionary and Mechanistic Aspects 47
Olga Khersonsky and Dan S. Tawfik, *Weizmann Institute of Science, Rehovot, Israel*

04 Mechanistic and Structural Studies of Microbial Dehalogenases: How Nature Cleaves a Carbon–Halogen Bond 89
Gerrit J. Poelarends, *University of Groningen, Groningen, The Netherlands*
Christian P. Whitman, *The University of Texas, Austin, TX, USA*

05 Guanidine-Modifying Enzymes in the Pentein Superfamily 125
Thomas W. Linsky and Walter Fast, *University of Texas, Austin, TX, USA*

06 Tunnels and Intermediates in the Glutamine-Dependent Amidotransferases 161
Nigel G. J. Richards, Robert N. Humkey, Kai Li, and Megan E. Meyer, *University of Florida, Gainesville, FL, USA*
Tania C. Córdova de Sintjago, *Universidad Central de Venezuela, Caracas, Venezuela*

07 Fatty Acid Biosynthesis and Oxidation 231
Huaning Zhang, Carl A. Machutta, and Peter J. Tonge, *Stony Brook University, Stony Brook, NY, USA*

08 Diels–Alderases 277
Hideaki Oikawa, *Hokkaido University, Sapporo, Japan*

09 Phosphoryl and Sulfuryl Transfer 315
Tiago A. S. Brandao and Alvan C. Hengge, *Utah State University, Logan, UT, USA*

10 Catalytic Mechanism of DNA Polymerases 349
Michelle P. Roettger, Marina Bakhtina, and Sandeep Kumar, *The Ohio State University, Columbus, OH, USA*
Ming-Daw Tsai, *Academia Sinica, Taipei, Taiwan*

11 Mechanisms of Enzymatic Glycosyl Transfer 385
Ran Zhang, Vivian L.Y. Yip, and Stephen G. Withers, *University of British Columbia, Vancouver, BC, Canada*

12 Synthesis of Alginate in Bacteria 423
Peter A. Tipton, *University of Missouri, Columbia, MO, USA*

13 Enzymology of Bacterial Resistance 443
Jed F. Fisher and Shahriar Mobashery, *University of Notre Dame, IN, USA*

14 Copper Metalloenzymes 489
Albrecht Messerschmidt, *Max-Planck-Institute of Biochemistry, Martinsried, Germany*

15 Mechanisms of Metal-Dependent Hydrolases in Metabolism 547
Marcy Hernick, *Virginia Polytechnic Institute and State University, Blacksburg, VA, USA*
Carol Fierke, *University of Michigan, Ann Arbor, MI, USA*

16 Dioxygenase Enzymes and Oxidative Cleavage Pathways 583
Timothy D. H. Bugg, *University of Warwick, Coventry, UK*

17 *S*-Adenosylmethionine and Iron–Sulfur Clusters in Biological Radical Reactions: The Radical SAM Superfamily 625
Eric M. Shepard and Joan B. Broderick, *Montana State University, Bozeman, MT, USA*

18 Detection of Novel Enzyme Intermediates 663
Cristina M. Furdui, *Wake Forest University School of Medicine, Winston-Salem, NC, USA*
Karen S. Anderson, *Yale University School of Medicine, New Haven, CT, USA*

19 Bisubstrate Analog Inhibitors 689
Patrick A. Frantom and John S. Blanchard, *Albert Einstein College of Medicine, Bronx, NY, USA*

20 Quantum Chemical Modeling of Enzymatic Reactions – Applications to Epoxide-Transforming Enzymes 719
Kathrin H. Hopmann and Fahmi Himo, *Royal Institute of Technology, Stockholm, Sweden*

Volume 9 Modern Methods for the Investigation of Natural Products

01 Overview and Introduction 1
Lew Mander, *Australia National University, Canberra, ACT, Australia*

02 High Performance Liquid Chromatographic Separation Methods 5
Reinhard I. Boysen and Milton T. W. Hearn, *Monash University, Melbourne, VIC, Australia*

03 Introduction to Macromolecular X-Ray Crystallography 51
Esko Oksanen and Adrian Goldman, *University of Helsinki, Helsinki, Finland*

04 Characterization by Circular Dichroism Spectroscopy 91
Nina Berova, George A. Ellestad, and Nobuyuki Harada, *Columbia University, New York, NY, USA*

05 Determination of Structure including Absolute Configuration of Bioactive Natural Products 147
Kenji Mori, *The University of Tokyo, Tokyo, Japan*

06 NMR – Small Molecules and Analysis of Complex Mixtures 169
Arthur S. Edison, *University of Florida, Gainesville, FL, USA*
Frank C. Schroeder, *Cornell University, Ithaca, NY, USA*

07 Biomolecular Recognition by Oligosaccharides and Glycopeptides: The NMR Point of View 197
Katalin E. Kövér, László Szilágyi, and Gyula Batta, *University of Debrecen, Debrecen, Hungary*
Dušan Uhrín, *University of Edinburgh, Edinburgh, UK*
Jesús Jiménez-Barbero, *Centro de Investigaciones Biológicas, Madrid, Spain*

08 Determination of Three-Dimensional Structures of Nucleic Acids by NMR 247
Nikolai B. Ulyanov and Thomas L. James, *University of California, San Francisco, San Francisco, CA, USA*

09 Derivation of Peptide and Protein Structure using NMR Spectroscopy 279
Glenn F. King and Mehdi Mobli, *The University of Queensland, St. Lucia, QLD, Australia*

10 Mass Spectrometry: An Essential Tool for Trace Identification and Quantification 327
Charles H. Hocart, *Australian National University, Canberra, ACT, Australia*

11 Applications of Modern Mass Spectrometry Techniques in Natural Products Chemistry 389
Roland D. Kersten, Michael J. Meehan, and Pieter C. Dorrestein, *University of California, San Diego, La Jolla, CA, USA*

12 Mass Spectrometry: Structure Determination of Proteins and Peptides 457
Chhabil Dass, *The University of Memphis, Memphis, TN, USA*

13 Application of Mass Spectrometry to Rapid Analysis of Bacterial Polysaccharides 497
Jianjun Li and Eleonora Altman, *Institute for Biological Sciences, Ottawa, ON, Canada*

14 Modern Methods for the Isolation of Natural Product Receptors 513
Peter Karuso, *Macquarie University, Sydney, NSW, Australia*

15 Bioinformatics 569
Yi-Ping Phoebe Chen, *Deakin University, Melbourne, VIC, Australia*
Elena P. Ivanova and Feng Wang, *Swinburne University of Technology, Melbourne, VIC, Australia*
Paolo Carloni, *International School for Advanced Studies, Trieste, Italy*

16 Natural Products Research and Metabolomics 595
V. Craige Trenerry and Simone J. Rochfort, *DPI-Werribee Centre, Werribee, VIC, Australia*

17 Small Molecules as Versatile Tools for Activity-Based Protein Profiling Experiments 629
Stephan A. Sieber, Thomas Böttcher, Isabell Staub, and Ronald Orth, *Ludwig-Maximilians-Universität München, Munich, Germany*

18 Metabolic Studies Using the Retrobiosynthesis Concept – Theory, Technology, and Examples 675
Adelbert Bacher and Wolfgang Eisenreich, *Technische Universität München, Garching, Germany*

19 Bacterial Protein Overexpression Systems and Strategies 695
C. Kinsland, *Cornell University, Ithaca, NY, USA*

20 Directed Evolution of Enzymes 723
Colin J. Jackson, *CSIRO Entomology, Black Mountain, ACT, Australia*
Elizabeth M. J. Gillam, *University of Queensland, St. Lucia, QLD, Australia*
David L. Ollis, *Australian National University, Canberra, ACT, Australia*

21 Single Molecule Fluorescence Methods in Enzymology 751
Peng Chen and Nesha May Andoy, *Cornell University, Ithaca, NY, USA*

Volume 10 Cumulative Index **1**

1.01 Overview and Introduction

C. A. Townsend, Johns Hopkins University, Baltimore, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

In the decade since the publication of the first edition of *Comprehensive Natural Products Chemistry*, progress in understanding the biosynthesis of the principal natural product groups has advanced greatly. This effort has been carried forward by an exponential growth in genome sequence information, the widespread adoption of molecular biological and bioinformatics methods, detailed mechanistic studies often coupled with X-ray crystal structures of key biosynthetic enzymes and, of course, improved analytical methods, notably electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) mass spectrometry. The breadth of these achievements is highlighted in 24 chapters spanning fatty acids to alkaloids. The further connectedness of this research can be glimpsed to chemical medicine and the rapidly evolving application of fermentation technology and pathway engineering to practical syntheses of natural and unnatural products for human health and commerce.

The border between primary and secondary metabolism has never been clearly drawn as is most obvious in fatty acids. Chapter 1.02 by Buist reviews unsaturated fatty acids whose physical properties not only affect membrane fluidity, but also effect species-specific roles in signaling. The means by which desaturation occurs to introduce isolated cis-double bonds is examined in detail to create sites of chemical potential that are realized, for example, by lipoxygenases and cyclooxygenases of the eicosanoid cascade. Turman and Marnett (Chapter 1.03) give an authoritative overview of this field and its intersection with nitrosation and physiology, but concentrate then on arachidonate metabolism by prostaglandin endoperoxide synthases and the multi-faceted roles played by these enzymes. The fatty acid section of this volume is rounded out in Chapter 1.04 by Barry and has team at the Tuberculosis Section of the NIH in a wide-ranging presentation of lipid-containing virulence factors in the mycobacterial cell envelope. Mycobacteria are characterized by a disproportionately high number of genes associated with lipid synthesis and biochemistry. These lipid classes are comprehensively reviewed with attention to their complex structures and as drug targets.

Evolutionarily related to the fatty acid synthases and sharing many of their catalytic functions are the type I and type II polyketide synthases (PKSs). Contemporaneous with, and absent from, the publication of the first edition of *Comprehensive Natural Products Chemistry* was the exciting discovery of the structurally and mechanistically distant family of type III PKSs from plants and bacteria. These remarkably simplified enzymes are discussed in chapters by Horinouchi with Katsuyama and Takahashi (Chapter 1.05), and by Morita, Abe, and Noguchi (Chapter 1.06). Rohr and Hertweck (Chapter 1.07) summarize the large body of work now published on type II PKSs and recent progress toward engineering the synthesis of new products. Important crystal structures of type II enzymes have appeared in the past 10 years and biochemical information reviewed by Rohr and Hertweck can be visualized in the context of dramatic structural information now available and treated by Tsai with Korman and Ames (Chapter 1.08) in the following chapter. Cox and Simpson (Chapter 1.09) discuss the iterative type I polyketide synthases (IPKSs) and relate the considerable advances that have been made in the last decade with these historically intractable enzymes. Hill and Staunton (Chapter 1.10) survey the very great accomplishments that have been achieved in understanding the reactions and global structures of type I modular PKSs and some of the applications of this knowledge in engineered systems. While nonribosomal peptide synthetases (NRPSs) are covered in Volume 5, Chapter 5.20, interesting products arise from hybrid synthetases typically linked with modular type I PKS domains. In Chapter 1.11, Sherman and members of his research group present an encompassing review of the field.

The classical mevalonate pathway and the more recently elucidated methylerythritol phosphate (MEP) pathway to the C_5-building blocks of isoprenoid biosynthesis are treated in a pair of chapters by Kuzuyama, Hemmi, and Takahashi (Chapter 1.12), and by Rohmer (Chapter 1.13), respectively, to open the discussion of the terpene classes and their biosynthesis. Considerable progress has been made in elucidating the MEP

pathway since the first edition. The assembly of these building blocks by the prenyltransferases is fundamental to all the downstream terpene classes and is reviewed by Koyama and Kurokawa in Chapter 1.14. Davis begins a discussion of the terpenes in a broad chapter (Chapter 1.15) outlining the principles of monoterpene cyclases and ending with detailed summaries of three cyclase crystal structures and their relation to mechanisms of reactions catalyzed and their stereochemical features. Mechanistic considerations and the relevant enzymes contributing to the generation of diverse C_{15} sesquiterpene skeletons are discussed by Chappell and Coates in Chapter 1.16, together with some emerging information on tailoring enzymes. In a similar manner, elaboration of diverse C_{20} diterpene structures is summarized by Toyomasu and Sassa in Chapter 1.17. Enzymatic construction of C_{30} triterpene scaffolds are discussed in a pair of chapters by Kushiro and Ebizuka (Chapter 1.18), and by Abe (Chapter 1.19), the former introducing functional diversity of oxidosqualene cyclases, whereas the latter discusses detailed structure–activity relationships among bacterial squalene cyclases. Misawa reviews the current understanding of carotenoid biosynthetic genes and their product enzymes with their potentials for industrial production of carotenoids by pathway-engineering approach (Chapter 1.20). Biosynthesis of plant sterols and their metabolism is updated by Schaller (Chapter 1.21). Isoprenoids from actinomycetes are summarized by Dairi in Chapter 1.22, illustrating the feasibility of genome mining for finding new structures.

Lewis and Davin wrote a chapter in the first edition of *Comprehensive Natural Products Chemistry* on lignans. He returns with members of his research group with a thorough treatment of lignans, neolignans, and allyl/propenyl phenols with a number of perspectives on human health (Chapter 1.23). He is joined by Ayabe and his team contributing a broad chapter (Chapter 1.24) on diverse plant phenolics including the flavonoids.

This volume is concluded by a single chapter (Chapter 1.25) on the alkaloids. Activity is building in this field as especially the genes encodings key biosynthetic enzymes are found, cloned, and overexpressed for biochemical, engineering, and structural studies. O'Connor reviews the advances of the last decade, notably, in benzoisoquinoline and the monoterpene alkaloid families.

Looking to the future it can be seen that burgeoning numbers of whole genome sequences, further development of bioinformatics, predictive tools and genome mining methods, increased knowledge of techniques to culture DNA from 'unculturable' organisms, growth in understanding enzyme function and structure, directed evolution of biosynthetic enzymes, pathway engineering, and applications of fermentation technology can all be seen to converge on new approaches for the generation of chemical diversity, stimulation of drug discovery, and the rational modification of natural products and their efficient large-scale production. The discovery and isolation of new natural product pathways, and the elucidation of the detailed mechanisms and structures of their encoded biosynthetic proteins will continue to lie at the heart of these developments by generating the fundamental knowledge on which engineering advances depend.

Biographical Sketch

Craig A. Townsend is the Alsoph H. Corwin Professor of Chemistry and holds joint appointments in the Departments of Biology and Biophysics. He received his BA with

Honors in chemistry from Williams College and his Ph.D. in organic chemistry from Yale University, where he was an NIH Predoctoral Fellow and won the Richard L. Wolfgang Prize. He was then an International Exchange Postdoctoral Fellow of the Swiss National Science Foundation at the ETH in Zurich, and joined the faculty at Johns Hopkins in 1976. His research interests are in the areas of biosynthesis and chemistry of natural products; stereochemical and mechanistic studies of enzyme action; drug design; small molecule DNA interactions; enzymology and molecular biology of secondary metabolism; and the clinical applications of fatty acid synthesis inhibition in the treatment of cancer, infectious diseases, and obesity. Professor Townsend has been a Research Fellow of the A. P. Sloan Foundation, a Camille and Henry Dreyfus Teacher-Scholar; he has received the Maryland Chemist of the Year award and an Arthur C. Cope Scholar Award from the American Chemical Society, and the Stuart Pharmaceuticals (now Astra/Zeneca) Award in Chemistry. He has served on NIH Study Sections as a regular and ad hoc member, and is on the editorial board for Bioorganic Chemistry. He was cochair and cofounder of the Bioorganic Chemistry Gordon Conference and currently serves as an at-large member on the governing Council of the Gordon Research Conferences, and has advised the Office of Technology Assessment, the American Chemical Society, and the National Institutes of Health.

1.02 Unsaturated Fatty Acids

Peter H. Buist, Carleton University, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

1.02.1	**Introduction**	5
1.02.2	**Monounsaturated Fatty Acids**	7
1.02.2.1	Oleic Acid	7
1.02.2.1.1	Soluble (plant) stearoyl ACP Δ9 desaturase	8
1.02.2.1.2	Membrane-bound (microbial and animal) Δ9 desaturases	14
1.02.2.2	**Regiochemical Variants of Δ9 Desaturation**	16
1.02.2.2.1	(*Z*)-4-Hexadecenoic acid	17
1.02.2.2.2	(*Z*)-5-Hexadecenoic acid	18
1.02.2.2.3	(*Z*)-6-Hexadecenoic acid	18
1.02.2.2.4	(*Z/E*)-11-Tetradecenoic acid	19
1.02.2.2.5	Sphingolipids	20
1.02.3	**Methylene-Interrupted Polyunsaturated Fatty Acids**	20
1.02.3.1	Linoleic Acid	22
1.02.3.2	α-Linolenic Acid	22
1.02.3.3	γ-Linolenic Acid	24
1.02.4	**Unusual Unsaturated Fatty Acids**	24
1.02.4.1	Ricinoleic Acid	24
1.02.4.2	Vernolic Acid	25
1.02.4.3	Crepenynoic Acid	25
1.02.4.4	Conjugated Fatty Acids	26
1.02.4.5	Allenic Fatty Acids	27
1.02.4.6	Sterculic Acid	27
1.02.5	**Summary and Future Prospects**	28
References		29

1.02.1 Introduction

Unsaturated fatty acid derivatives play a critically important role in virtually every life form. The most universal of these functions relates to the fact that the fluidity of biomembrane phospholipids is optimized by the presence of (*Z*)-olefinic hydrocarbon side-chains. That is, the presence of the latter characteristic prevents overly rigid packing of the linear hydrocarbon portions of the membrane and thus promotes the biologically viable liquid crystalline phase rather than the solid gel phase.[1] In addition to this structural aspect, unsaturated triglyceride esters are important cellular reservoirs of chemical potential energy that can be easily mobilized for catabolic use as required.[2] Interestingly, seed oils of many plants harbor various desaturated fatty acid triglycerides with additional exotic functionality that appears to confer antifeedant properties.[3] Probably the most subtle role of the C–C double bond in long-chain lipidic molecules has to do with their signaling properties; the corresponding saturated analogue is devoid of this capability.[4] The classic example of lipid-based signaling phenomenon is exhibited by insect pheromones where the number, position, and stereochemistry of the olefinic recognition element constitutes a species-specific, semiochemical language.[5] More recently, the importance of regiospecific double bond position in sphingolipid (cell messenger) function has also begun to be appreciated.[6] Finally, it should be noted that fatty acids with multiple double bonds serve as substrates for oxidizing enzymes such as lipoxygenases and cyclooxygenases to generate an ensemble of highly bioactive compounds in plants and animals.[7,8] Ironically, the presence of polyunsaturated fatty acids also promotes deleterious autoxidation processes that have been implicated in various disease states.

The most abundant unsaturated fatty acids have a chain length of 18 carbons. The structural relationships between various classes of C18 olefinic fatty acids are displayed in **Figure 1**. These include the common

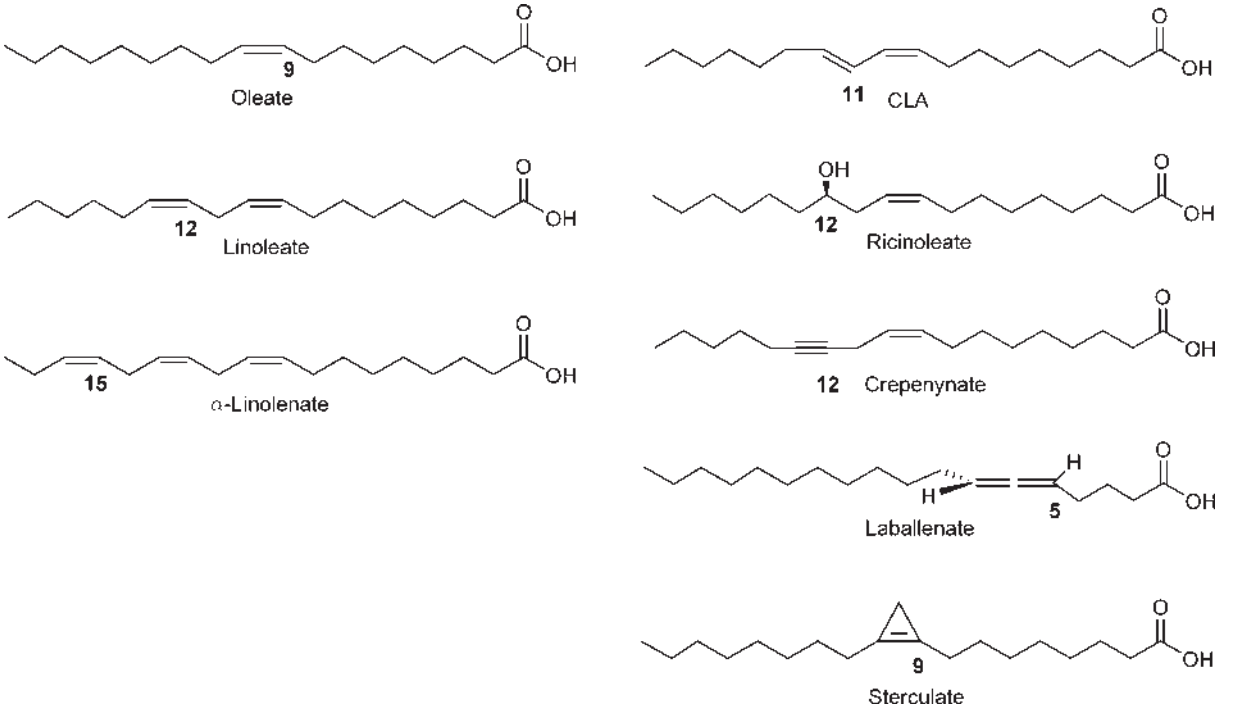


Figure 1 Structural relationships between unsaturated fatty acids.

(a) monounsaturated fatty acids (oleate); (b) methylene-interrupted, polyunsaturated fatty acids (linoleate and α-linolenate) as well as the unusual; (c) conjugated fatty acids such as the conjugated isomers of linoleic acid (9,11-CLA); (d) oxygenated fatty acids (ricinoleate); (e) acetylenic fatty acids (crepenynate); (f) allenic fatty acids (laballenate); and (g) cyclopropenyl fatty acids (sterculate).

The isolation and characterization of many of the genes responsible for the formation of the compounds listed above has revitalized the study of unsaturated fatty acid biosynthesis and made possible a series of fundamental mechanistic and structural studies. Application of this knowledge has helped to open new doors in more applied areas of research such as plant biotechnology. The current concept is one of the plant as a green factory capable of generating petroleum-like feedstocks for the production of high volume compounds such as lubricants, detergents, cosmetics, surfactants, and polymers.[9] In addition, the idea of adding value to plant seed oils for the growing nutritional supplement market through genetic manipulation remains a viable vision.[10] Closely related to a more sophisticated understanding of how the lipodome impacts human health is the search by medicinal chemists for new therapeutic targets to manage symptoms of the metabolic syndrome and associated metabolic diseases such as diabetes.[11] Unsaturated lipid-based targets are also being considered in the global effort to eradicate tuberculosis[12] and to understand the role of sphingolipids in apoptosis.[13]

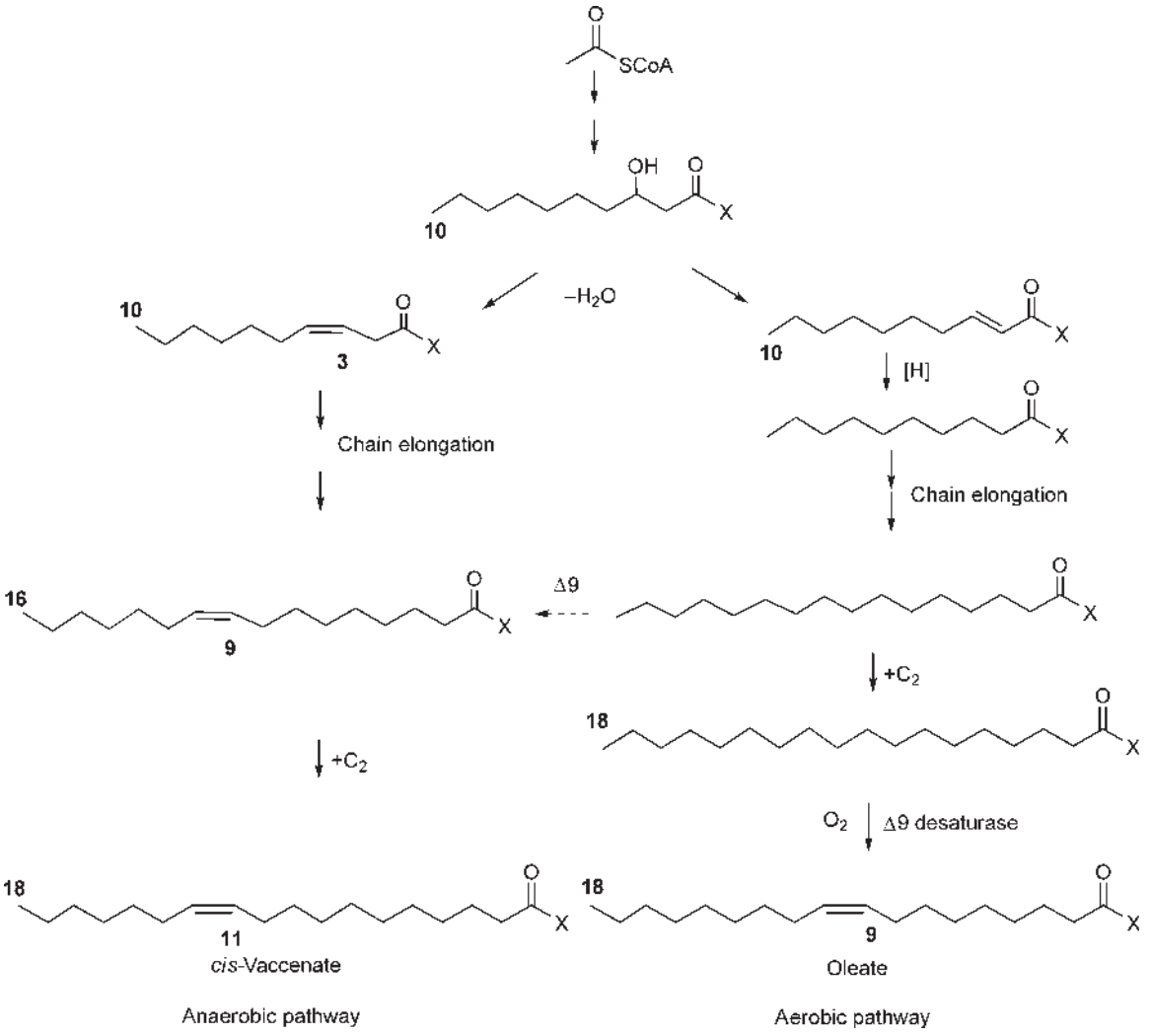
This chapter focuses on the mechanistic aspects of unsaturated fatty acid biosynthesis and related compounds; as such it expands on the previous review of this topic found in the first edition of *CONAP*[14] and contains a more focused treatment on desaturases.[15] The reader is directed to excellent reviews on other aspects relevant to this topic including essays on the chemistry of fatty acids,[16] molecular biology of desaturases,[17] structural aspects of desaturase and desaturase-like enzymes,[18] soluble plant desaturases,[19] plant biotechnology,[20–22] cold acclimation,[23,24] metabolic syndrome,[25] insect pheromones,[26] and sphingolipids.[27] Updated comprehensive monographs on lipid biochemistry are also available.[28,29]

1.02.2 Monounsaturated Fatty Acids

Most naturally occurring monounsaturated fatty acids are biosynthesized by regioselective, O_2-dependent, dehydrogenation (desaturation) of a saturated substrate catalyzed by a family of enzymes known as desaturases (**Scheme 1**).[30] The term 'desaturation'[31] was first used in recognition of the fact that the O_2-dependent dehydrogenation of adjacent unactivated methylene groups is mechanistically distinct from dehydrogenation alpha to the thioester group in fatty acid degradation. A relatively minor, anaerobic pathway to olefinic fatty acids is found only in certain bacteria (e.g., *Escherichia coli, Lactobacilli*) and involves a unique enzyme activity that catalyzes formation of a (3*Z*)-3-decenoyl ACP (acyl carrier protein) intermediate that is then elongated. (**Scheme 1**).[32] There is potential for pathway crossover in organisms where the Δ9 desaturase acts on a palmitoyl CoA (coenzyme A) to give palmitoleyl CoA (9*Z*-C16:1) that is subsequently elongated to *cis*-vaccenate. Until recently, it was thought that the purely anaerobic and aerobic routes were mutually exclusive in any given organism. However, Zhu *et al.* have shown using specific gene-knock out studies that *Pseudomonas aeruginosa* employs both pathways to generate the monounsaturated fatty acids required to regulate membrane lipid fluidity.[33]

1.02.2.1 Oleic Acid

The most common monounsaturated fatty acid is oleic acid (**Scheme 1**). The glycerol ester of oleate is a major component of industrially important plant seed oils. The olive tree has been cultivated for its oleate-rich oil in the Mediterranean basin for thousands of years; indeed, the commercial value of this commodity was so high, that trade in olive oil is considered to be a major contributor to the prosperity of Graeco-Roman civilization.[34] Nowadays, oleic acid is one of the cornerstone molecules of a 50 billion dollar per annum (US, 2006) global enterprise that bears its name: the oleochemical industry.[35] For example, oxidative cleavage at the 9,10-double bond gives ready access to useful synthons – nonanoic acid and nonan-1,9-dioic (azelaic acid). Oleic acid is formed by the aerobic pathway in all life forms examined to date, including bacteria, fungi, yeasts, algae, plants, insects, and mammals. The introduction of the double bond at the 9,10 position of a stearoyl substrate is catalyzed by two structurally distinct enzymes: a soluble protein found only in plants and a ubiquitous,



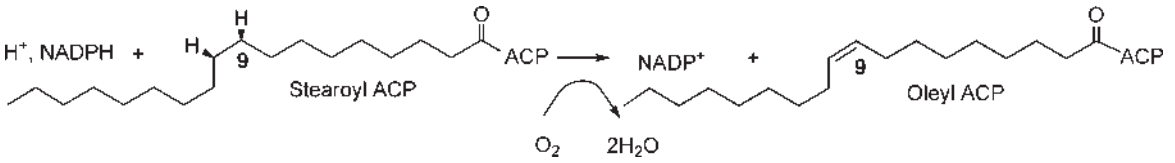
Scheme 1 Anaerobic and aerobic synthesis of unsaturated fatty acids.

membrane-bound enzyme. Oleyl phospholipids are major components of cell membrane bilayers and it should be noted that introduction of the *cis*-double bond at the middle of a saturated chain leads to the greatest melting point depression.[36]

1.02.2.1.1 Soluble (plant) stearoyl ACP Δ9 desaturase

Plant seed oils represent the major source of oleic acid where it is stored as triacylglyceride (TAG). The route to oleyl TAG originates with palmitoyl ACP exiting the chloroplastidial fatty acid synthase assembly line (aerobic pathway, **Scheme 1**) and undergoing chain elongation to give stearoyl ACP; this species is then desaturated *in situ* by a soluble, stearoyl ACP specific, Δ9 desaturase; the product of this reaction – oleyl ACP – is then released to the extraplastidial sector where it enters the fatty acyl CoA pool en route to oleyl TAG biosynthesis.[37]

The overall chemical equation for soluble (plant) stearoyl ACP Δ9 desaturase-mediated dehydrogenation is typical of all desaturases and involves a four-electron reduction of molecular oxygen to two molecules of water. Two electrons are provided by NAD(P)H via a flavin-dependent NADPH reductase and the electron transfer protein, ferredoxin; the remaining two electrons are donated by substrate C–H bonds (**Scheme 2**).



Scheme 2 Overall chemical equation for a prototypical soluble Δ9 desaturase-mediated dehydrogenation.

Significant progress in the biochemical characterization of soluble plant desaturases has been achieved in the last two decades. The critical breakthrough occurred in the 1990s when castor stearoyl-ACP Δ9 desaturase was purified and overexpressed[38,39] to permit examination by various physical methods including Mössbauer spectroscopy,[40] resonance Raman spectroscopy,[41] X-ray crystallography,[42] and magnetic circular dichroism (MCD).[43] This pioneering work revealed that the active site of this enzyme contained a nonheme, di-iron cluster coordinated to several bridging and nonbridging carboxylate ligands as well as two histidines (**Figure 2**). The similarity of this catalytic core to that found in methane monooxygenase (MMO)[44] was striking and provided an immediate structural and mechanistic link between the activation of the primary C–H bond of methane and the secondary C9-H, C10-H bonds of stearoyl ACP. Hydrophobic patch analysis of the crystal structure revealed a potential substrate binding region in the middle of a globular protein (**Figure 2**).[42] This picture provided the first clue as to how active site topology might influence the conformational preference of a mobile hydrocarbon moiety such that a thermodynamically less stable (*Z*)-olefinic bond is generated upon dehydrogenation. That is, a stearoyl substrate can be docked into the active site if it adopts a quasi-*gauche* conformation at the C9–C10 position in the region situated opposite to the two iron atoms. Such an arrangement would facilitate regio- and stereoselective oxidation leading to the desired oleyl product (**Figure 2**). Unfortunately, it has not been possible, to date, to obtain crystals of substrate/enzyme of sufficient quality in order to gain more insight into this scenario. However, support for the model displayed in **Figure 2** was obtained by the results of a stereochemical investigation using stereospecifically monodeuterated substrates; it was shown unambiguously that the substrate hydrogens are removed in *syn*-fashion with a pro *R* enantioselectivity at C9 and C10[45] (compare positions of C9, C10 hydrogens in **Scheme 2** and **Figure 2**).

Two additional X-ray structures of the castor protein[46,47] and that of a related protein found in mycobacteria[48] confirm the relative location and coordination of the diiron core within the desaturase protein. The overall kinetics and bioinorganic aspects of stearoyl ACP oxidation have been studied in some detail.[49] The

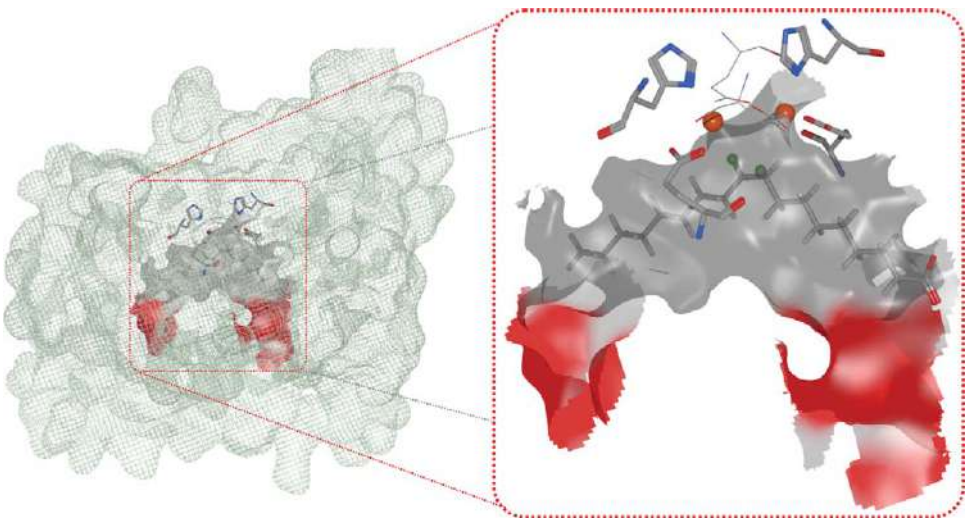


Figure 2 X-ray structure of a soluble Δ9 desaturase showing how the topology of the active site dictates *syn*-removal of the two pro *R* hydrogens at C9 and C10 of a stearoyl ACP substrate. The reaction produces a (*Z*)-olefin (oleyl ACP) with >99.9% stereochemical purity.

situation is complicated by the fact that the castor desaturase is a functional homodimer but the picture that is emerging is one of substrate binding promoting electron transfer to the diiron core and subsequent oxygen binding en route to a compound Q-like oxidant that is capable of attacking unactivated C–H bonds.[50,51] The modulating effect of substrate binding is reminiscent of the role of the effector proteins in related enzymes such as toluene-4-monoxygenase (TOM)[52] and MMO.[53]

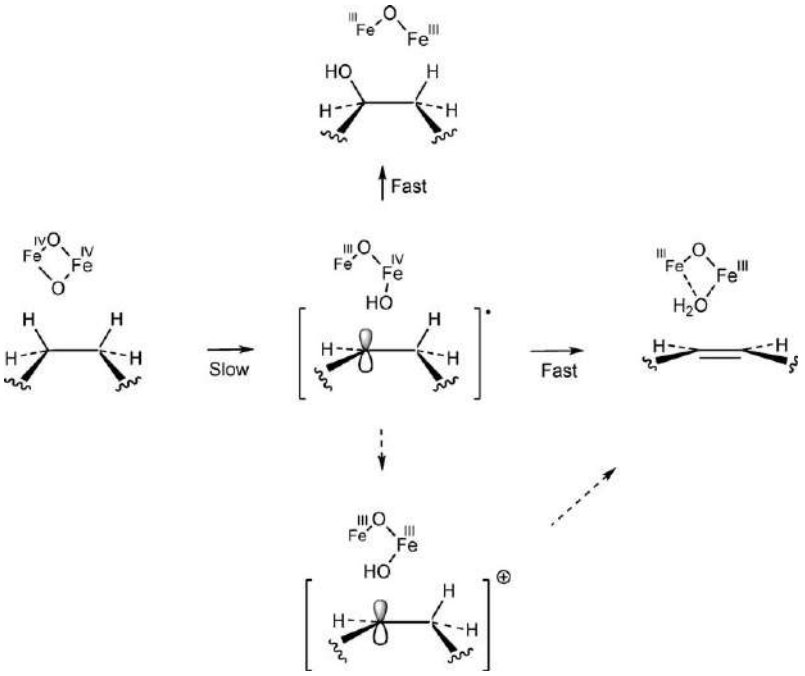
A topic of long-standing interest to bioorganic chemists is the mechanism of C–H activation. There are essentially three fundamental questions that are pertinent to this topic:

1. What is the stereochemistry of hydrogen removal as this relates to enantioselectivity of hydrogen removal (pro *R* or pro *S*) at the prochiral methylene centers and the relative stereochemistry of C–H bond cleavage at the adjacent carbons (*syn* vs *anti*)?
2. What is the site of initial oxidation (cryptoregiochemistry), assuming the hydrogens are removed in a nonsynchronous fashion?
3. Which factor(s) prevent a competing hydroxylation reaction?

The answer to the first question has already been provided for the soluble plant Δ9 desaturase (*vide supra*) via the appropriate deuterium labeling experiment but attempts to tackle questions (2) and (3) are still very much works in progress. Indeed, the current X-ray crystallographic data offers no readily interpretable information in this regard.[42]

The availability of a stable, soluble desaturase such as the castor protein affords a unique opportunity to apply suitable mechanistic probes that had previously been developed using *in vivo* desaturating systems.[15] The choice of probe is limited by the requirement to activate substrate as the ACP thioester via an enzymatic coupling reaction. This places some limitation on the types of substrate analogues one can use to probe the active site. However, in general it has been found that sterically unobtrusive substrates bearing deuterium, oxygen, sulfur (methylene isostere), and monofluorine substituents function well in ACP activation and desaturation. Selected (*E*)-olefinic substrates also bind to the active site.[54]

In order to develop approaches to answering questions (2) and (3), a working hypothesis for the mechanism of C–H activation in desaturation was required (**Scheme 3**). A similar scheme was first advanced for a cytochrome



Scheme 3 Generic mechanism for C–H activation during fatty acid desaturation showing its relationship to hydroxylation. The structure of the di-iron oxidant and the reactive intermediates are speculative.

P450 system by Baillie upon the discovery that hepatic cytochrome P450 could both hydroxylate and dehydrogenate the epileptic drug, valproic acid.[55] The mechanistic proposal presented in the latter work was adopted as a generic model for fatty acid desaturation[56,57] under the assumption that both heme iron and nonheme iron-based oxidations of unactivated methylene groups were likely to operate by similar mechanisms. For cytochrome P450, the active oxidant is thought to be a hypervalent iron oxo species; for nonheme diiron systems, the equivalent oxidant is thought to be compound Q[58] as determined by parallel mechanistic work addressing the mode of C–H activation for MMO.[59]

As shown in **Scheme 3**, desaturation is thought to be initiated by an energetically difficult C–H activation step to produce a carbon-centered radical/FeOH pair which collapses rapidly by a second hydrogen abstraction step to give an olefinic product and iron-bound water either directly or via a one electron oxidation/deprotonation sequence. Many membrane-bound (but not soluble) desaturases apparently allow a competing 'hydroxyl rebound'[60] (better termed as SH_2[61]) pathway to give variable amounts of a secondary alcohol byproduct.[62] Interestingly, this pathway can be induced for the soluble enzyme through the use of substrate analogues (*vide infra*). There is no evidence to support the notion that hydroxylated intermediates are on the pathway to olefin.[63]

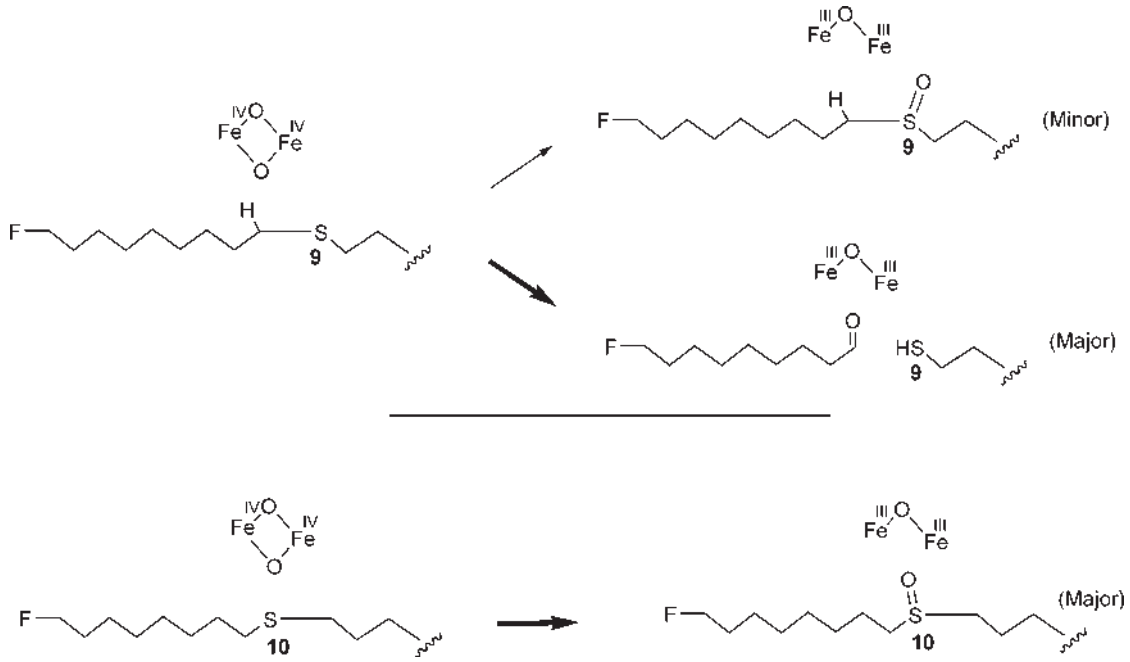
To deduce the site of initial oxidative attack (cryptoregiochemistry[64]) involved in double bond formation, the use of kinetic isotope effect (KIE) techniques is the most straightforward approach. This methodology is based on the reasonable premise that the initial C–H bond cleavage step should be energetically more difficult and therefore more sensitive to isotopic substitution than the second C–H bond breaking step (**Scheme 3**). However, the KIE approach cannot be used in the case of the soluble plant Δ9 desaturase since other enzyme events such as substrate binding are kinetically more important than the chemical events. Both noncompetitive and competitive intermolecular primary deuterium KIE on C–H cleavage at either the C9 or the C10 positions were estimated to be ∼1.[65,66] This fact has prevented the direct determination of the site of initial oxidation for this enzyme, a parameter which is so readily accessible for membrane-bound desaturases (*vide infra*).

Other mechanistic probes have yielded valuable information as to which methylene group undergoes initial oxidative attack in the active site of stearoyl ACP Δ9 desaturase.

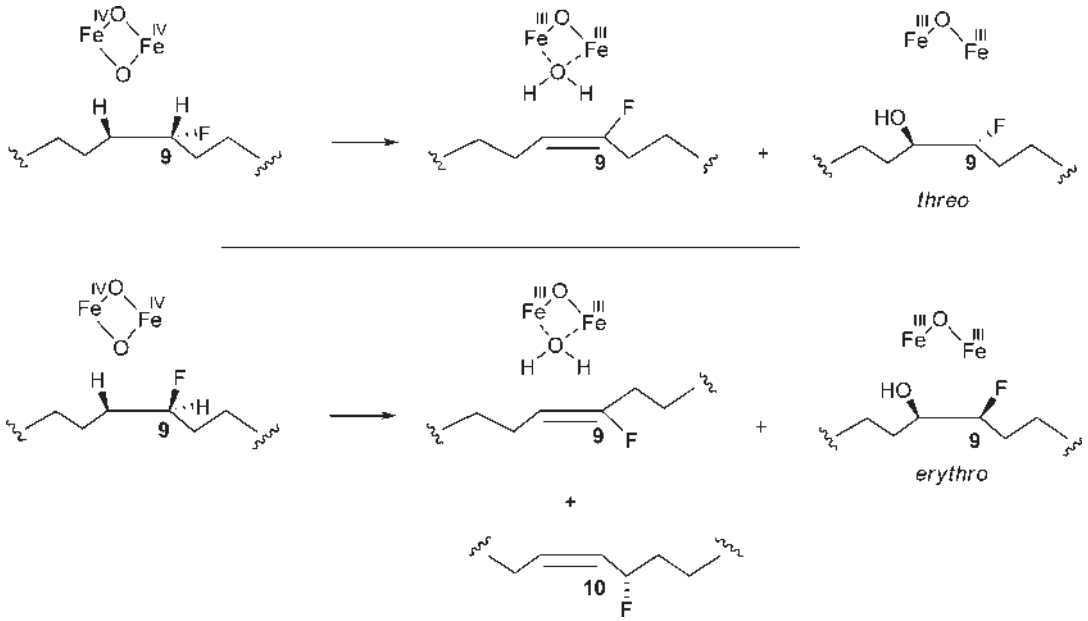
This was accomplished in the first instance by measuring the efficiency of desaturase-mediated oxo transfer to a series of thia-substrate analogues.[65] The assumption is that the relative amounts of sulfoxide obtained by oxo transfer from each thia isomer reflects the positioning of the putative di-iron oxidant relative to substrate. It was found that 10-thiastearoyl ACP was converted cleanly to the corresponding sulfoxide while oxidation of the 9-thia isomer gave the corresponding sulfoxide in much lower yield (∼10%). It should be noted that when substrates bearing sulfur at other carbon sites were incubated with the desaturase, normal dehydrogenation products were detected. This set of experiments was repeated using ω-fluorinated thia probes, followed by[19] F NMR analysis and yielded similar results (**Scheme 4**).[67] In the case of the 9-thia experiment, the major product obtained was 9-fluorononanol[68] – a compound thought to arise by a chain cleavage process in the enzymatic event (the intermediate aldehyde is reduced to alcohol by the normal reductive workup of the enzymatic mixture). These results have been corroborated in every detail by Fox and coworkers and, in addition, it has been shown that oxygenation at sulfur is an aerobic process by[18]O-labeling.[69]

Thus one can conclude that the soluble enzyme may in fact operate by initial attack at the C10 position. Some support for this hypothesis is available from work using probes containing other heteroatoms. For example, incubation of a series of oxo-substrates generated the expected chain cleavage product (nonanal) from 9-oxostearoyl ACP, while a 10-oxo substituent blocked normal oxidative attack and was desaturated at the C11,12 position.[70] This regiochemical error was induced in the latter case because hydroxylation at C9 was not possible presumably due to geometric constraints.

A similar loss of regiochemical precision has been observed upon incubation of racemic 10-monofluorostearate with the castor enzyme; no 9-hydroxylated products were detected with these substrates as well. Interestingly, 10-hydroxylated products were obtained upon desaturase-mediated oxidation of (*R*)- and (*S*)-9-fluorostearoyl ACP (**Scheme 5**).[45,71] When fluorine is a spectator atom at the pro *S* position, dehydrogenation can proceed along to give fluoroolefin along with a *threo*–fluorohydrin byproduct. The latter was a product of hydroxylation as shown by [18]O-labeling. In contrast, the (9*R*)-fluorostearate was processed in an unexpected fashion to give mainly fluoroolefinic products of the 'wrong' stereochemistry and regiochemistry. The 9, 10- fluorohydrin obtained in this experiment was '*erythro*' as determined by [19]F NMR in



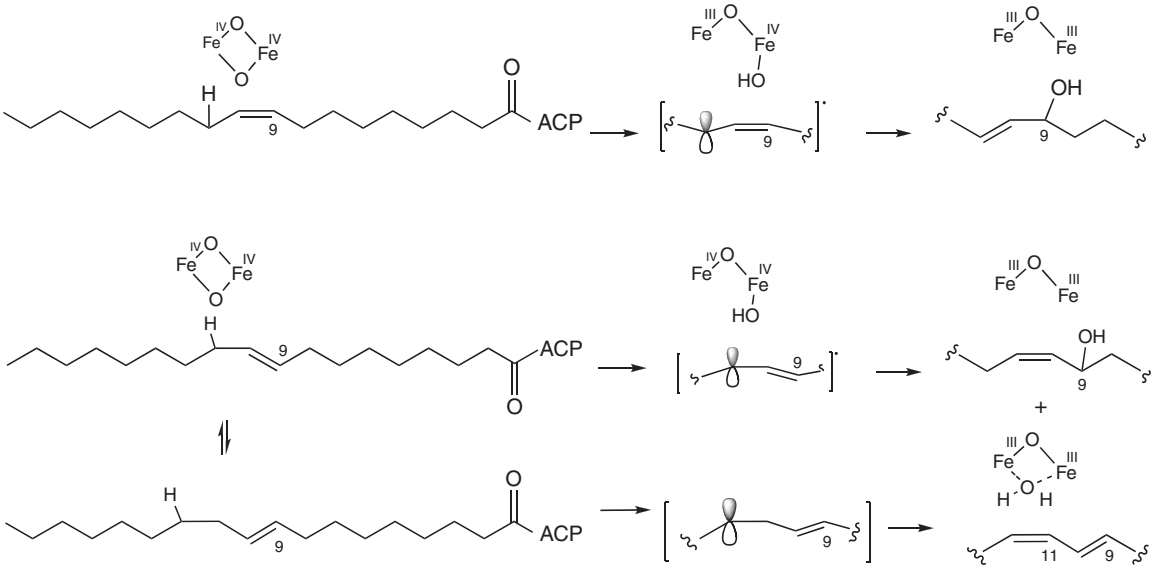
Scheme 4 Thia test for cryptoregiochemistry of a soluble plant Δ9 desaturase: regioselective oxidative cleavage and sulfoxidation chemistry.



Scheme 5 Comparison of (*R*)- and (*S*)-9-fluorostearate oxidation by a soluble plant Δ9 desaturase leading to novel hydroxylated and olefinic products.

combination with synthetic reference standards. The commitment to dehydrogenation versus hydroxylation when presented with a substrate designed to block the former pathway is remarkable.

More light has been shed on the mechanism of desaturation in this system through the use of a triple mutant of the castor acyl-ACP Δ9 desaturase (T117R/G188L/D280K).[72] This mutant enzyme was originally designed to switch the regiochemistry of desaturation from Δ9 to Δ4. Instead, stearoyl-ACP was converted to the oleyl



Scheme 6 Products obtained by the action of a triple mutant of a soluble plant Δ9 desaturase on oleyl and elaidyl ACP.

product that was then further oxidized to give the allylic alcohol (*E*)-10-18:1-9-OH via a (*Z*)-9-18:1 allylic intermediate (**Scheme 6**). The use of regiospecifically deuterated substrates showed that the conversion of (*Z*)-9-18:1 substrate to (*E*)-10-18:1-9-OH product proceeds via hydrogen abstraction at C11 and highly regioselective hydroxylation (>97%) at C9. ^{18}O-labeling studies show that the hydroxyl oxygen in the reaction product is exclusively derived from molecular oxygen – a result that makes a carbocationic intermediate less likely since such a species would tend to be quenched by water in the active site. The mutant enzyme also converts (*E*)-9-18:1-ACP into two products, (*Z*)-10-18:1-9-OH, and the conjugated linolenic acid (CLA) isomer, (*E*)-9-(*Z*)-11-18:2. The observed product profiles can be rationalized by differences in substrate binding as dictated by the curvature of substrate channel at the active site. In both cases, the course of reaction is consistent with the location of the oxidant close to C10 rather than C9 as noted above. It is interesting to note the similarity of the allylic hydroxylation reaction of the triple mutant and the pathway postulated for dimorphecolate formation in *Dimorpheteca sp.*[73] (*vide infra*, **Scheme 10**(**c**)).

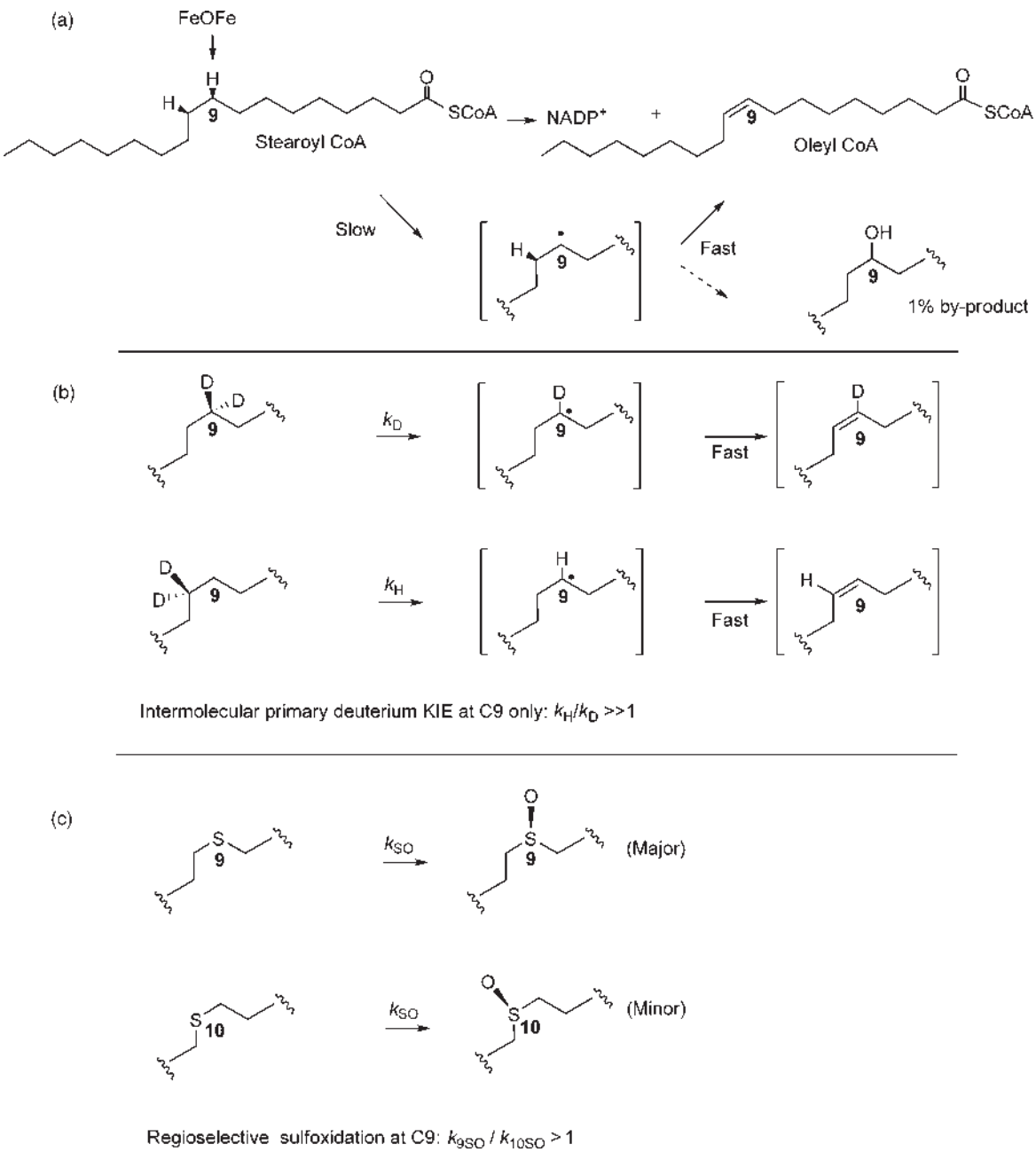
The nature of the switch-controlling reaction outcome (dehydrogenation vs oxygenation) is a critically important issue which has not been fully resolved to date due to the paucity of structural data. This mechanistic dichotomy has been documented for nonheme monoiron,[74] nonheme diiron,[75] and heme iron (cytochrome P450) biooxidants.[76] Certainly, the holy grail in the desaturase area is to be able to convert a pure desaturase to a pure (regio- and enantioselective) hydroxylase by protein engineering. In one sense, the experiment has already been performed as in the case of the castor plant where the 12-hydroxylase responsible for ricinoleic acid production was found to be closely related in sequence space to the corresponding Δ12 desaturase – an enzyme shown to initiate oxidation at C12.[77] There are at least three possible factors that could be involved in controlling the relative energetics of the two oxidative pathways. (1) Formation of carbocationic intermediates favor desaturation while radical intermediates favor hydroxyl rebound. The former pathway could be favored by a stabilizing factor in the substrate or via pi-cation stabilization by proximal aromatic residues. (2) Stereoelectronic effects relating to the alignment of the C–H bond beta to the singly occupied molecular orbital SOMO of the putative radical intermediate. Misalignment promotes a default hydroxyl rebound reaction; proper periplanar alignment of orbitals promotes olefin formation. The precise positioning of the iron-hydroxyl group relative to the SOMO may also influence the rate of hydroxyl rebound versus abstraction of the second hydrogen. (3) The driving force of the reaction is determined by the relative energetics associated with the final coordination of the metal site by water (desaturation) versus no ligand (hydroxylation).[74] Any of these considerations either singly or in combination have not been ruled out conclusively.[78]

1.02.2.1.2 Membrane-bound (microbial and animal) Δ9 desaturases

As mentioned previously, 9,10-dehydrogenation of a stearoyl substrate is executed by two structurally distinct enzymes: a soluble plant Δ9 desaturase that is specific for the ACP derivatives and a set of membrane-bound Δ9 desaturases that accept stearoyl CoA or stearate linked to PL (phospholipid). Historically, stearoyl CoA Δ9 desaturase played an important role in the development of desaturase research. An account of the circumstances surrounding the discovery of desaturases by Konrad Bloch[79] some 50 years ago has been highlighted in a recent review.[15] The first publication describing fatty acid desaturation – the aerobic formation of oleic acid by yeast – appeared in *Biochemia et Biophysica Acta* in 1958.[80] In this publication and in an accompanying paper, Bloomfield and Bloch[80,81] showed that a particulate (membranous) fraction from yeast could 'desaturate' palmitoyl CoA (and by implication stearoyl CoA) in the presence of molecular oxygen and NADPH. Thus the overall stoichiometry of desaturation was established and is identical to that shown in **Scheme 2**. Similar Δ9 desaturase activity was discovered in the microsomes of several other organisms but it was not until the early 1970s that the fractionation of rat liver microsomes by Strittmatter's group yielded purified integral membrane-bound Δ9 desaturase.[82] Reconstitution of this enzyme in detergents along with the electron-transfer components, membrane-associated NADH (nicotinamide adenine dinucleotide) reductase and cytochrome b5, represented one of the early triumphs in membrane enzymology. At this time, the presence of functional nonheme rather than heme-iron in the desaturase protein was confirmed by spectroscopy.[83] Hydropathy analysis of the rat liner Δ9 desaturase amino acid sequence[84] together with that of a homologous yeast Δ9 desaturase generated the first 2-D model[85] that featured two membrane-spanning regions with the bulk of the protein facing the cytosolic side of the endoplasmic reticulum (ER) membrane. A more detailed analysis based on a larger sequence set, coupled with site-directed mutagenesis experiments, revealed the presence of eight essential histidines residing in three highly conserved $HX_{(2-4)}H$ boxes.[86] It is now believed that these histidines coordinate to a di-iron center as documented by Mössbauer data obtained on a closely related nonheme iron-containing ω-hydroxylase alkB.[87] It should be noted that a number of membranous desaturases, including the aforementioned yeast Δ9 desaturase, have been shown to contain a fused cytochrome b5 component that presumably facilitates electron transfer from NADH reductase to the catalytic di-iron core.[88] All attempts to obtain more detailed structural data on membranous Δ9 desaturases and related desaturase enzymes have been uniformly unsuccessful. Sperling and Heinz have carried out a sequence comparison between various subgroups found within the Δ9 desaturase envelope.[17]

Much of the early mechanistic work on desaturation was carried out using intact microorganisms containing membrane-bound Δ9 desaturases, including the classic study by Schroepfer and Bloch using *Corynebacterium diphtheriae*.[89] In this work, mass spectrometric analysis of biosynthetic oleate derived from *erythro*- and *threo*-dideuterated substrates established the '*syn*' removal of hydrogens from C9 and C10. That it is the pro *R* hydrogens that are abstracted was determined by monitoring the loss of tritium label from the four possible, stereospecifically tritiated substrates as compared to a reference (^{14}C-labeled) substrate. The stepwise nature of Δ9 desaturation as subsequently portrayed in the mechanistic model outlined in **Scheme 7(a)** was first postulated by Bloch as a spinoff result from this experiment. That is, the $^{3}H/^{14}C$ ratio of each residual, stereospecifically tritiated substrate obtained after incubation was compared with that of the starting ratio. Significant enrichment of tritium in only one of the starting materials – (9*R*)-^{3}H-stearate signaled a large primary ^{3}H isotope effect associated with the loss of that tritium atom in the first step of desaturation. This result implied that Δ9 desaturation was initiated by abstraction of the C9 (H_R) hydrogen in a relatively slow, isotopically sensitive step.

Similar stereochemical results for Δ9 desaturation were obtained by Morris and James using a *Chlorella vulgaris* system.[90] However, their analysis of the labeling data led to the inference that C–H cleavage reactions at C9 and C10 were both sensitive to deuterium or tritium substitution. A subsequent set of *in vitro* experiments using a noncompetitive KIE design led to similar conclusions.[91] This apparent contradiction between stepwise versus concerted mechanisms for desaturations persisted in the review literature for a number of years. That Δ9 desaturation as mediated by membrane-bound Δ9 desaturases is clearly stepwise and initiated at C9 was finally demonstrated unambiguously by a series of incubation experiments in yeast (*Saccharomyces cerevisiae*),[64] cyanobacteria,[92] and green algae *Chlorella*[93] using competitive KIE measurements (**Scheme 7(b)**). This approach involved incubation of a ~1:1 mixture of regiospecifically dideuterated substrate and its nondeuterated parent with a convenient source of the desaturase; the d_1/d_0 ratio of the olefinic product was



Scheme 7 (a) Postulated mechanistic pathway for membrane-bound Δ9 desaturase-mediated oxidation of stearoyl CoA. (b) The KIE test for cryptoregiochemistry. (c) The thia test for cryptoregiochemistry.

then compared to the starting d_2/d_0 ratio of the substrate by mass spectrometric analysis. In both methods, secondary deuterium KIEs are embedded in the value for the primary KIEs but this perturbation is expected to be small.[94] In each case,[64,92,93] a large deuterium KIE was observed at C9 and a negligible KIE at C10.

Corroborating evidence for location of the initial oxidation at C9 in membranous Δ9 desaturases was obtained using a 'thia test' – an experiment that measures the efficiency of oxo transfer to thia substrate analogues as a function of sulfur position (*vide supra*) (**Scheme 7**(**c**)). A series of sulfur-substituted stearates and ω-phenyl thia fatty acid analogues were incubated with a highly active stearoyl CoA Δ9 desaturase in growing *S. cerevisiae* cultures.[56] The polar sulfoxy products were excreted into the culture media from which they could

be conveniently extracted as the carboxylic acids. Sulfoxide levels could be quantitated through mass spectrometric or NMR analysis of product mixtures. The sensitivity of the latter method was greatly enhanced through the use of ω-fluoro-tagged thiasubstrates followed by ^{19}F NMR analysis.[95] It was found in these series of experiments that 9-sulfoxidation was consistently more efficient than 10-sulfoxidation, as one might expect if the oxidant hovered over C9 of the parent substrate. Furthermore, it was found the sulfoxidation reaction was highly enantioselective and produced the expected sulfoxide enantiomer in high % ee as determined by Pirkle-type chiral shift experiments in combination with chiral synthetic standards.[56] The stereochemistry of sulfoxidation matched the known enantioselectivity of hydrogen removal for the parent reaction. Thus thia probes were used to derive stereochemical and cryptoregiochemical information in a single set of experiments.

Further evidence in support of the cryptoregiochemical assignment comes from the observation that ~1% of a 9-hydroxystearate byproduct is formed in yeast cultures containing active Δ9 desaturase from a variety of sources.[62] The quantitation has been confirmed using a 16-fluoropalmitate substrate and GC-MS detection of a 16-fluoro-9-hydroxypalmitate.[96] It is thought that these byproducts are formed by hydroxyl quenching of a C9 carbon centered radical (**Scheme 7**(**a**)). It is interesting to note that the formation of hydroxylated byproduct had been overlooked for many years due to the fact that hydroxyfatty acids do not chromatograph well on GC columns unless they are derivatized as silyl ethers.

The results of the thia test and hydroxyl byproduct analysis are opposite to that observed for the case of the soluble Δ9 desaturase as discussed above (desaturation is believed to be initiated at C10 of the parent substrate). This contrast in cryptoregiochemical preference reflects the fact that the two enzymes in question belong to two structurally distinct protein groups and have different substrate specificity profiles relating to different active site architectures: the soluble desaturase has a strict preference for C18 substrates while the membranous desaturase accepts substrate chain lengths ranging from C15 to C19.[97] Despite these differences, the overall mechanism of C–H activation is probably very similar in both cases.

The subject of whether carbon-centered radicals or carbocationic intermediates are formed by membranous Δ9 desaturases is difficult to address experimentally in an unambiguous manner. To date, explicit attempts to trap out the putative radical intermediates in desaturases through the use of cyclopropyl radical clocks for desaturases have not been successful.[54,98] However, it should be noted that the membrane-bound, nonheme monooxygenase of *Pseudomonas oleovorans* (*alk B*), which serves as a structural model for this set of desaturases,[99] has been probed in this regard. Two mechanistic studies involving *alk B* operating on cyclic substrates point to the intermediacy of a secondary radical but not a secondary carbocationic species along the reaction pathway.[100,101]

The importance of gaining more fundamental knowledge on membranous Δ9 desaturases has become evident with the recent discovery that overexpression of stearoyl CoA Δ9 desaturase (SCD) in animals may correlate with the symptoms of a variety of metabolic diseases such as obesity and diabetes.[25] Consequently, the search for therapeutically useful SCD inhibitors is an active research area and some progress has been made.[102,103] One of the most potent SCD inhibitors is the naturally occurring cyclopropenyl fatty acid, sterculic acid (**Figure 1**) – an exotic fatty acid that presumably binds to the SCD active site and interacts in some way with the diiron catalytic core via the strained cyclopropenyl ring. The mechanism of inhibition is difficult to study due to the lack of highly active, stable SCD enzyme preparations and the lability of the sterculyl cyclopropene ring.

1.02.2.2 Regiochemical Variants of Δ9 Desaturation

The ability to install functionality at various unactivated centers along a conformationally mobile hydrocarbon chain in a highly regioselective manner is the primary hallmark of the desaturase family enzymes (**Figure 3**). This display of bioselectivity has enticed researchers to search for clues among the many sequences of soluble and membrane desaturases in order to discover the amino acid hotspots controlling regiochemical outcome.[104,105] Sperling and Heinz have provided a comprehensive overview of primary sequences for all known desaturases (as of 2003); this work is a valuable resource for all future efforts to decipher the mysteries of regioselectivity contained in this information.[17]

As noted previously, oleic acid is by far the most common monounsaturated fatty acid. Its C16 homologue, palmitoleate, can be formed by the anaerobic pathway or by the aerobic pathway using membrane-bound

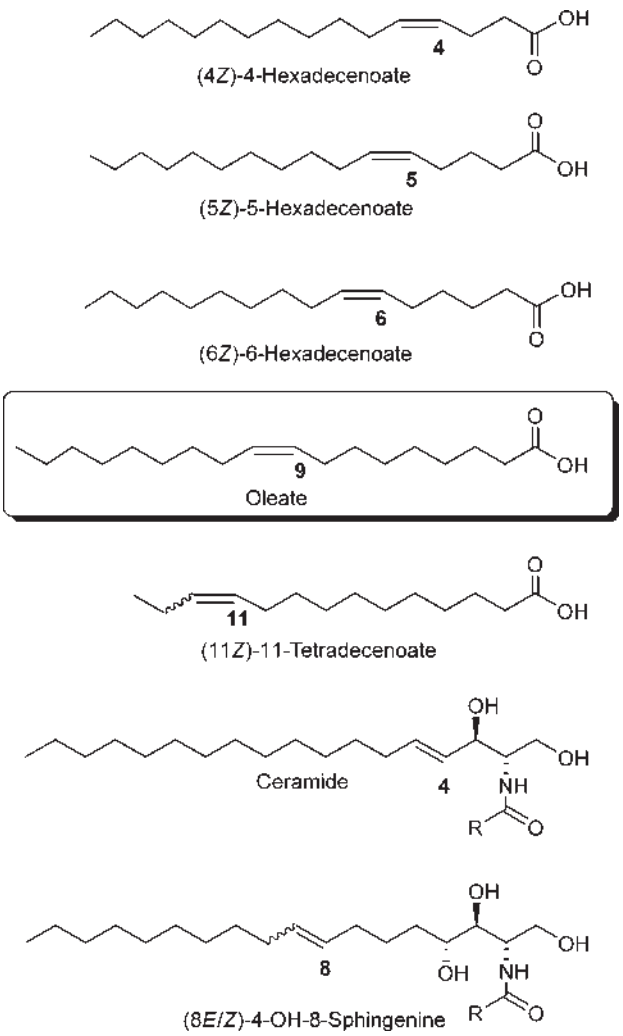


Figure 3 Examples of regiochemical variants of Δ9 desaturation reactions which lead to monounsaturated products.

Δ9 desaturases only (**Scheme 1**). The recognition element determining the position of the incipient double bond in the latter case is the C1 acyl head group since the double bond is introduced at the C9,10 position regardless of overall chain length. This type of recognition is denoted by the use of the symbol Δ. Regiochemical variants are found in the soluble, fatty acyl ACP-recognizing, desaturase series (Δ4 and Δ6) and in the membrane desaturase series recognizing the CoA, PL, or sphingolipid head group (Δ4, Δ5, Δ8, Δ11). The variability in polar head group recognition and substrate chain length optimum complicates the comparison between enzymes.

1.02.2.2.1 (Z)-4-Hexadecenoic acid

The title compound (**Figure 3**) is produced by the action of a Δ4 palmitoyl ACP desaturase found in some species of *Umbelliferae*, *Araliaceae*, and *Garryaceae* and is chain elongated to produce petroselinic (6*Z*)-(6--octadecenoic) acid.[106] This acid accounts for as much as 85% of the total fatty acid content in these plants and is of interest to plant scientists because the shift of the double bond position from the more common Δ9 position raises the melting point by some 20 °C. A significant opportunity to place desaturase regioselectivity on a firm structural basis presented itself with the discovery that Δ4 palmitoyl ACP desaturase from *Hedera helix* (English ivy) was amenable to analysis by X-ray crystallography.[107] The outcome of this work was

surprising. The structures of the Δ4 palmitoyl ACP desaturase and the original Δ9 stearoyl ACP desaturase are startlingly similar. Subtle differences were noted in residue conformation at the methyl terminus of the putative hydrophobic substrate binding site and at the ACP binding area near the opening of the active site. The tentative conclusion was that these modifications allowed a hexadecanoyl ACP substrate to penetrate the ivy Δ4 protein more deeply than is the case for the corresponding stearoyl ACP substrate and the castor Δ9 protein. Due to the lack of definitive data on the mode of substrate binding, it was important to examine the stereochemistry of Δ4 desaturase-mediated dehydrogenation.[108] This was accomplished by tracking the fate of deuterium atoms located on stereospecifically monodeuterated substrates, (4*S*)- and (4*R*)-[4-$^{2}H_{1}$]-palmitoyl-ACP and (5*S*)- and (5*R*)-[5-$^{2}H_{1}$]-palmitate-ACP. These compounds were prepared by a novel, general route involving nucleophilic opening of chiral terminal epoxides that are available in very high % ee. It was found that the introduction of the (*Z*)-double bond between C4 and C5 of a palmitoyl substrate occurs with pro *R* enantioselectivity – a result which matches that obtained for a closely related homologue castor stearoyl-ACP Δ9 desaturase. These data suggest that despite the marked difference in the regioselectivity between the two enzymes, the stereochemistry of hydrogen removal is conserved. Of great interest would be the elucidation of the cryptoregiochemistry of the Δ4 desaturation.

Analysis of the X-ray structure of the ivy enzyme also yielded unexpected information on the diiron catalytic core in that the iron atoms were observed in the diferric oxidation for the first time. This allowed a direct comparison between the ligand coordination for the reduced and oxidized state and exposed carboxylate shifts that are characteristic of this class of enzyme.[109] While the mechanistic significance of these differences cannot be ascertained at this time, it is hoped that these data can be used to build up a consistent picture of the entire catalytic cycle from substrate binding to exit of product.

1.02.2.2.2 (Z)-5-Hexadecenoic acid

(*Z*)-Δ5 desaturation of cell membrane phospholipids in various species of *Bacillus* is inducible by a 15 °C reduction in the ambient temperature (37–20 °C).[110] This constitutes an ideal system to study the mechanisms controlling the degree of fluidity of biological membranes through the (*Z*)-dehydrogenation of their fatty acyl hydrocarbon chains. (*Z*)-Δ5 unsaturation is atypical and is presumably optimized for the ecological niche of this organism. The presence of branched chain fatty acids (iso and anteiso) in the cell membranes may play a role in accommodating a 5*Z*-C16:1 component. The control mechanisms involved in the signal transduction process have been worked out by DeMendoza and coworkers.[111] The response to a decrease in growth temperature involves hyper-expression of the gene, coding for the acyl-lipid Δ5 desaturase. Regulation is accomplished by a two-component system composed of a membrane-associated kinase, DesK, and a soluble transcriptional activator, DesR. The temperature-sensing ability of the DesK protein is regulated by the extent of disorder within the membrane lipid bilayer. The sensor protein DesK controls the signal decay of its cognate partner, DesR, and this response regulator activates transcription of its target promoter. Analysis of the Δ5 desaturase gene[112] found in *Bacillus subtilis* revealed that this protein occupied its own unique position in the sequence space of membrane-bound desaturases. Protein fusion expression using alkaline phosphatase, combined with site-directed mutagenesis experiments and hydropathy analysis, have yielded a novel 2-D model of the *Bacillus* Δ5 desaturase.[113] The generality of this result remains to be determined.

The novel regiochemistry of this bacterial desaturase prompted a comparison of the site of initial oxidation for this enzyme with that of the prototypical Δ9 desaturase. The intermolecular, competitive, primary deuterium KIE on each C–H cleavage step of Δ5 desaturation was determined: 3.9 ± 0.4 at C5 while the C_6–H bond breaking step was shown to be insensitive to deuterium substitution (KIE $= 1.17 \pm 0.02$).[114] Based on the mechanistic model for desaturation (**Scheme 3**), these results suggest that the site of initial oxidation for Δ5 desaturation is at C5. In accordance with this result, Shanklin and coworkers have recently detected low levels of 5-hydroxypalmitate derived from the *Bacillus* Δ5 desaturase expressed in *E. coli*.[62] Surprisingly, the enantioselectiviy of Δ5 desaturation has not been elucidated to date.

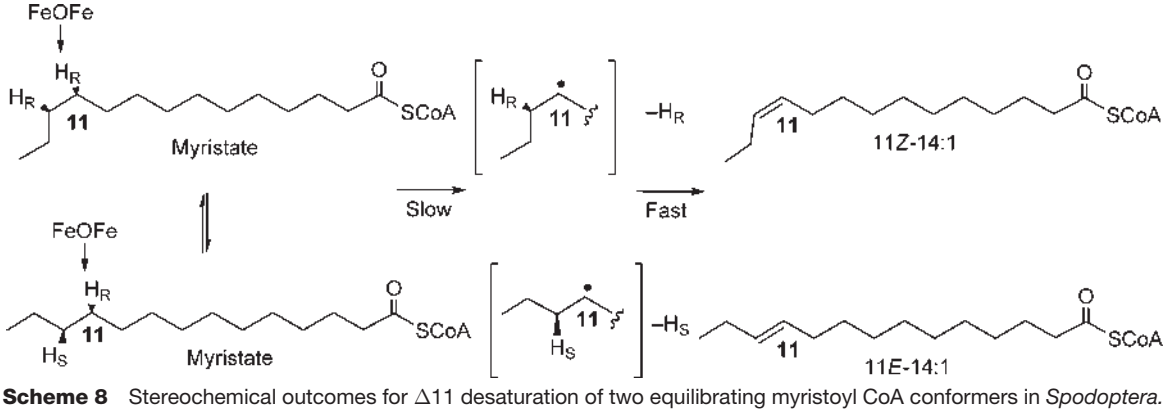
1.02.2.2.3 (Z)-6-Hexadecenoic acid

(*Z*)-6-Hexadecenoic acid (16:1 Δ6) comprises more than 80% of the seed oil of *Thunbergia alata* (Black-eyed Susan vine) and it was determined that this unusual acid was the product of a soluble Δ6 16:0-ACP desaturase that shares 66% sequence identity with the castor Δ9 18:0-ACP desaturase and 57% identity with the

Δ4 16:0-ACP desaturase.[115] The biological purpose for accumulating a hexadecenoate with a double bond at C6,7 rather than at the more usual C9,10 position is unknown. The X-ray structure of the castor Δ9 enzyme was used as the basis for site-directed mutagenesis studies to determine the structural determinants for the substrate and double bond positional specificities displayed by acyl-ACP desaturases.[116] By replacement of specific amino acid residues in a Δ6-palmitoyl (16:0)-ACP desaturase with their equivalents from a Δ9-stearoyl (18:0)-ACP desaturase, mutant enzymes were identified that have altered fatty acid chain-length specificities or that can insert double bonds into either the Δ6 or the Δ9 positions of 16:0- and 18:0-ACP. Most notably, by replacement of five amino acids (A181T/A200F/S205N/L206T/G207A), the Δ6-16:0-ACP desaturase was converted into an enzyme that functions principally as a Δ9-18:0-ACP desaturase. Many of the determinants of fatty acid chain-length specificity in these mutants are found in residues that line the substrate-binding channel as revealed by X-ray crystallography of the Δ9-18:0-ACP desaturase. The crystallographic model of the active site is also consistent with the diverged activities associated with naturally occurring variant acyl-ACP desaturases. In addition, on the basis of the active-site model, a Δ9-18:0-ACP desaturase was converted into an enzyme with substrate preference for 16:0-ACP by replacement of two residues (L118F/P179I). These results demonstrated the ability to rationally modify acyl-ACP desaturase activities through site-directed mutagenesis and represented an early success story in protein engineering of desaturases. No stereochemical or cryptoregiochemical studies have been carried out on the Δ6 desaturase system although it would be of interest to compare this enzyme with the Δ4 ivy desaturase (*vide supra*).

1.02.2.2.4 (Z/E)-11-Tetradecenoic acid

The rich chemical biology of the carbon–carbon double bond is fully evident in the fascinating ability of insects to generate species-specific, olefinic mating pheromones. The role of desaturases in producing volatile compound blends with distinctive stereochemistry and regiochemistry that can be distinguished from the normal fatty acid profile was first elucidated by Roloefs and coworkers.[117] The bioorganic work in this area was initiated by Boland *et al.* who showed that a Δ11 insect desaturase found in *Mamestra brassicae* effected the *syn* removal of pro *R* hydrogens.[118] *In vivo* work in this area is technically very challenging and requires mass labeling of substrates with multiple deuterium atoms to prevent mass spectral interference by endogenous unlabeled substrate. Similar results were obtained for pheromone biosynthesis found in *Manduca sexta* and *Bombyx mori*.[119] This work was continued by the Fabrias bioorganic group and a series of stereochemical and cryptoregiochemical studies have appeared in the last decade with a special focus on *Spodoptera littoralis*. This insect species produces a unique blend of monoenoic and dienoic tetradecanoates. It is now known that the monoene stereoisomers are formed by the action of a single Δ11 desaturase[120] which can accommodate two equilibrating substrate conformers (**Scheme 8**). Interestingly, the ratio of *E*/*Z* changes as a function of chain length. The initial abstraction of the pro *R* hydrogen of the stationary C11 methylene group generates a short-lived radical intermediate which undergoes a second hydrogen abstraction to generate a mixture of (*E*/*Z*)-11-tetradecenoates.[121] The relative stereochemistry of hydrogen removal is '*syn*' in both cases.[122]



Scheme 8 Stereochemical outcomes for Δ11 desaturation of two equilibrating myristoyl CoA conformers in *Spodoptera*.

The ability to functionally express these enzymes in yeast has allowed one to study the transformation of long-chain fatty acyl substrates to their 11-ene products in greater mechanistic detail.[123] This development allowed further mechanistic studies and led to the observation of a minor 11-hydroxylated byproduct in the fatty acid profile.[124] This compound presumably arises by hydroxyl trapping of a carbon-centered radical at C11 in the same way that 9-hydroxystearate is a byproduct of a C9-initiated Δ9 desaturase (*vide supra*).

Use of a 11-cyclopropylundecanoyl probe failed to intercept the putative radical intermediate through a ring opening rearrangement or hydroxyl quenching;[98] instead a highly strained cyclopropylidene product was formed indicating the very high commitment to the dehydrogenation pathway displayed by these systems. This result may serve as a model for biological allene formation.[125]

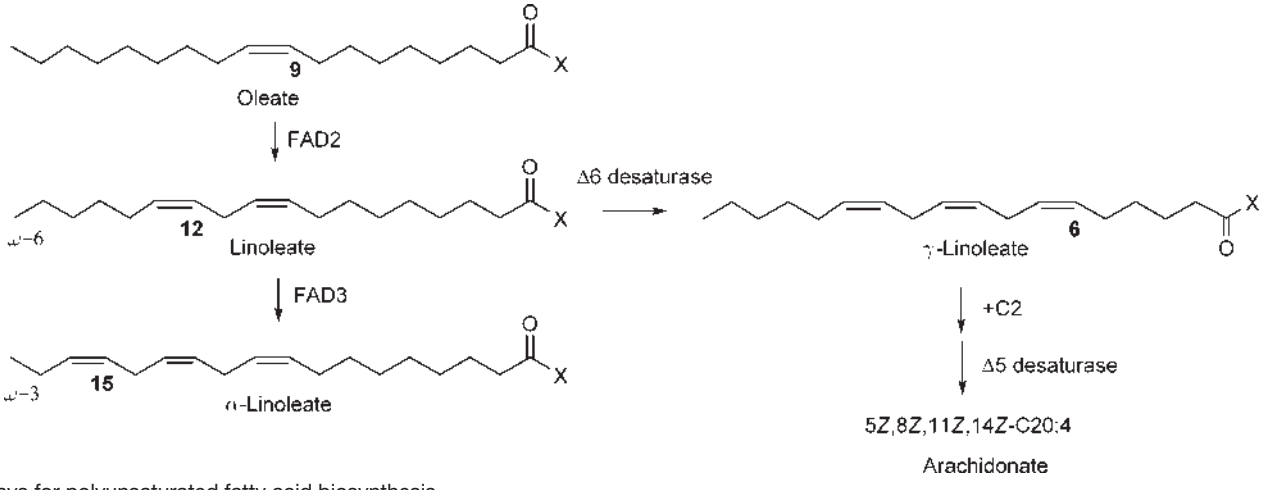
1.02.2.2.5 Sphingolipids

Sphingolipid derivatives containing an (*E*)-4-ene or 4-hydroxy functionality are thought to play a critical role as signaling agents in many biological processes such as cell proliferation and apoptosis in mammals,[126] gamete-specific cell-cycle progression in *Drosphila*,[127] and stress responses in plants[128] and microorganisms.[129] A large Δ4-sphingolipid desaturase family has been identified using a bioinformatic approach and the previously suspected close relationship between dihydroceramide 4(*E*)-dehydrogenation and 4(*R*)-hydroxylation was firmly established.[130] This conclusion is also supported at the mechanistic level. A pioneering *in vivo* rat study by Stoffel *et al.*[131] showed that Δ4 desaturation involves *syn*-removal of the C(4)-H_R and C(5)-H_S hydrogens and that this process was likely initiated at C4 on the basis of a large primary ^3H isotope effect at C4 and not at C5. These results were corroborated by the stereochemical analysis of a bifunctional Δ4-desaturase/4-hydroxylase from *Candida albicans*[132] and an *in vitro* cryptoregiochemical study of a rat liver microsomal Δ4-desaturase.[133] Clearly, 4-hydroxylation and Δ4 desaturation are both initiated by removal of the pro *R* C4 hydrogen. The bifunctional *Candida* system would appear to offer a particularly convenient opportunity to determine more precisely the point at which desaturation and hydroxylation pathways bifurcate. Another important development in the sphingolipid area is the discovery of potent synthetic dihydroceramide Δ4 desaturase inhibitors which features a bioinspired cyclopropene moiety at the C4,5 position[134] and a sulfur atom at the C5 position.[135] A detailed study of the mechanism of inhibition has been carried out for the cyclopropenyl compound.[136] Interestingly, it appears that irreversible binding to the protein does not seem to be involved. Recently, the Fabrias compound has found application in cancer research.[137]

In 1998, a novel sphingolipid Δ8 desaturase[138,139] was characterized which, at the level of its primary sequence, bears a strong resemblance to the membranous plant Δ6 desaturases discussed later in this chapter. A unique feature of the Δ8 desaturase is that it catalyzes the introduction of a double bond at the 8,9-position of phytosphinganine with less than 100% stereoselectivity (7:1 (*E*):(*Z*)). While the production of a stereoisomeric mixture of olefins has good precedent in insect pheromone biosynthesis (*vide supra*), the formation of such isomers is rare at methylene positions closer to C1. Apart from the broader biological significance of such apparently 'sloppy' biochemistry, it was of interest to probe this lack of precision in the dehydrogenation event more closely. The sphingolipid Δ8 desaturase gene cloned from sunflower (*Helianthus annuus*) was expressed in yeast and the stereochemistry of hydrogen removal assessed by incubation with an appropriate enantiomerically enriched precursor bearing deuterium at the vicinal pro *R* positions.[140] ESI-MS analysis of the sphingolipid olefinic products showed that desaturation had proceeded with *syn*-stereochemistry for both isomers. The cryptoregiochemical analysis of the same system carried out using a pair of mass-labeled racemic monodeuterated substrates yielded unexpected results. While the site of initial oxidation for the production of the major (*E*)-isomer appears to be at C8 (primary KIE at C8 ~2; at C9 ~1), the KIE signature found for the (*Z*)-isomer appeared to be reversed (KIE at C8 ~2, C9 ~4).[140] In the latter case, neither of the KIE values is close to unity; this may be interpreted in terms of a less than absolute preference for initial attack at C9.

1.02.3 Methylene-Interrupted Polyunsaturated Fatty Acids

The biological formation of fatty acids bearing multiple nonconjugated double bonds (PUFA) relies on a number of enzymes that are asymmetrically distributed among various biological species. A highly condensed version of the interrelationships between some of the more prominent members of the PUFA family is given in **Scheme 9**. The most important aspect of this picture is the fact that mammals do not have the enzymes to



Scheme 9 Pathways for polyunsaturated fatty acid biosynthesis.

biosynthesize two nutritionally essential fatty acids, namely linoleate and α-linolenate. Fortunately, these compounds are found in abundance in the seed oils of various commercially viable crops. Two more mammalian desaturases, a Δ6 and a Δ5, are required to produce arachidonic acid – a critically important eicosanoid precursor.

The production of polyunsaturated fatty acids in plants is part of the cold acclimation strategy to preserve the integrity of various membrane assemblies over a temperature range of some 100 °C. Numerous studies that have been carried out in model systems examine the control mechanisms involved in cold tolerance using *Arabidopsis mutants*,[24] cyanobacteria,[23] and animals.[141] An additional role for PUFA is the generation of various plant hormones such as jasmonic acid in response to various stressors.[7]

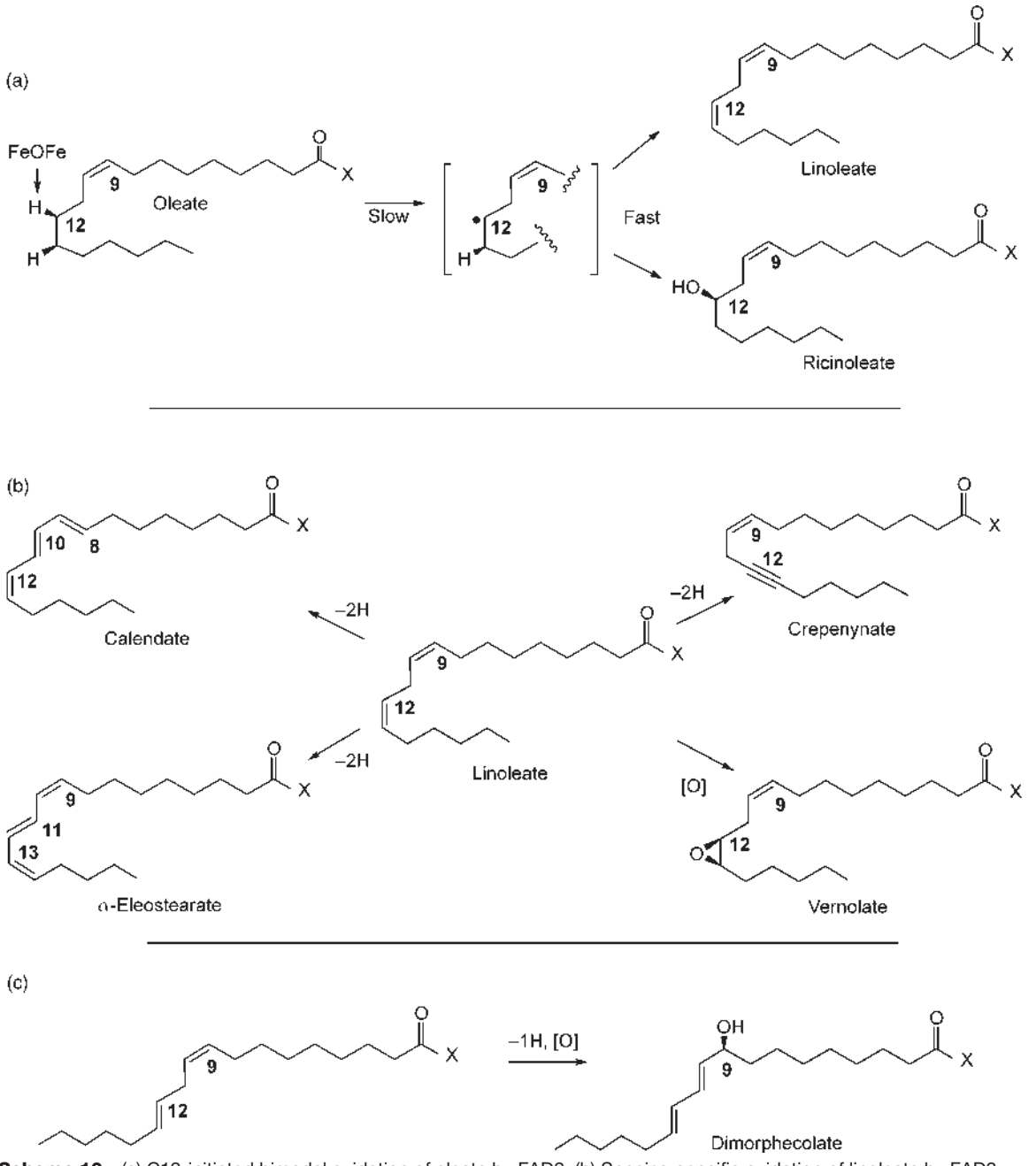
1.02.3.1 Linoleic Acid

The 'essential fatty acid,' linoleic acid, is biosynthesized in the plant by one of two enzyme systems: (1) a plastidial enzyme (FAD6) which uses the methyl terminus as a reference point and which has been classified as an ω-6 desaturase because it introduces the double bond six carbons from the ω-carbon; (2) an extraplastidial system known as oleate Δ12 desaturase (FAD2) which is selective for C12,13 oxidation independent of chain length and acts on 9(*Z*)-monoenoic substrates.[142] When presented with a 10(*Z*)-C19:1 substrate, this enzyme will use the existing double bond as a reference point and produces a methylene interrupted diene: 10(*Z*)13(*Z*)C19:2.[143] FAD2 desaturase is an interesting enzyme because closely related FAD2 variants generate a number of unusual fatty acids (see next section) such as ricinoleate ((*R*)-12-hydroxyoleate). The molecular biological and stereochemical data implied that both linoleate and ricinoleate formation were initiated by abstraction of the pro *R* C12 hydrogen (**Scheme 10(a)**).[77,90] Heterologous expression of FAD2 in yeast by Covello[144] represented a breakthrough in that it made possible mechanistic studies that would have been extremely difficult to carry out *in planta*. Support for the hypothesis that linoleate formation was initiated at C12 was obtained by comparing the primary deuterium KIE on C–H cleavage at C12 and C13 using a strain of *S. cerevisiae* containing a functionally expressed plant oleate Δ12 desaturase from *Arabidopsis thaliana*.[144] As anticipated, a large primary deuterium KIE at C12 ($k_H/k_D = 7.3 \pm 0.4$) and essentially no effect at C13 ($k_H/k_D = 1.05 \pm 0.04$) was observed.[75] Similar results have been obtained for the oleate Δ12 desaturase in *Chlorella*.[93] There is real potential to carry out further mechanistic studies on Δ12 desaturases given that a cyanobacterial enzyme can be partially purified and will accept oleic acid salts as substrates.[145,146]

1.02.3.2 α-Linolenic Acid

α-Linolenic acid (ALA) is also considered an essential fatty acid but the precise biological role of this compound or its derivatives has been difficult to decipher.[147] The role of ALA in plants, as an essential part of the cold acclimation process, has been studied extensively and with clear-cut results through the use of various mutants of plants that lack this fatty acid.[24] The C16 equivalent of ALA – 9Z, 12Z, 15Z-hexadecenoate – features an unusual terminal double bond and has been recently found in a soil protozoan[148] and in a precursor role for the production of sorgoleone – a highly bioactive natural product found in *Sorghum bicolor* root hairs.[149]

Cryptoregiochemical analyses of the type described in the previous section have also been carried out in mechanistic studies on the formation of α-linolenic acid in the nematode worm *Caenorhabditis elegans* (FAT1)[150] and plants (FAD3).[151] The nematode system normally produces eicosapentaenoic acid (EPA) from arachidonic acid.[152] These enzymes are more accurately termed ω-3 desaturases rather than Δ15 since, when presented with a series of fatty acyl substrates of varying chain length, a double bond is always introduced by three carbons from the methyl terminus. Using strains of *S. cerevisiae* containing *fat-1*[152] genes or *fad-3*[153] genes and the appropriate, regiospecifically deuterated substrates, it was found that ω-3 desaturation proceeds with large primary deuterium KIE only at the ω-3 position in both cases. A similar trend in KIEs (one large, one negligible) was obtained for this reaction as it occurs in *Chlorella vulgaris*.[154] These results clearly point to ω-3 ($18 - 3 = 15$) as the site of initial H-abstraction for ω-3 desaturases; importantly, small amounts of



Scheme 10 (a) C12-initiated bimodal oxidation of oleate by FAD2. (b) Species-specific oxidation of linoleate by FAD2 variants. (c) Biosynthesis of dimorphecolate by a FAD2 variant.

15-hydroxylinoleate have been found in flax seed oil – a rich source of α-linolenate.[62] The stereochemistry of hydrogen removal for either FAD3 or FAT1 catalyzed transformations has not been determined to date.

In a quest for locating the structural determinants for regioselectivity in FAD2 and FAD3 type desaturases, the groups of Feussner[104] and Covello[105] have studied bifunctional oleoyl-Δ12/linoleoyl-ω-3 desaturases in *Aspergillus nidulans* and *Claviceps purpurea*. A number of domains critical to the selectivity of these enzymes have been identified via a series of mutagenesis experiments. 3-D structural data will be required to fully interpret these data.

1.02.3.3 γ-Linolenic Acid

The Δ6 desaturation of a linoleyl substrate to produce γ-linolenic acid (GLA) is an important step in the biosynthetic pathway leading to arachidonic acid and other polyunsaturated fatty acids (PUFAs) in animals.[155] A mouse model[156] has recently become available that will allow a more precise definition of the role of this pathway in human health. Surprisingly, Δ6 desaturation also occurs in some plants and the value of GLA-containing dietary supplements such as evening primrose oil in managing various disease states for example has been well-documented.[157,158] Thus considerable effort has been expended by plant biotechnologists in the characterization of Δ6 desaturase genes from various sources. It was a consequence of this activity that the first example of a desaturase fused to an N-terminal cytochrome b5 domain – the Δ6 desaturase from borage – was discovered.[159] Domain-switching experiments have been carried out using chimeras of Δ6 fatty acid and Δ8 sphingolipid desaturases in order to elucidate the connection between these two enzyme activities that are so closely related at the sequence level.[160] A bifunctional Δ6 desaturase/acetylenase has been discovered in *Ceratodon purpureus*.[161] In addition, a bifunctional Δ5, Δ6 desaturase has been identified in zebrafish.[162]

Detailed mechanistic studies have not been carried out on any of the systems mentioned above. However, a report of differential levels of incorporation of 6,6-d_2- versus 7,7-d_2-oleates into Δ6-desaturated lipids of *Tetrahymena*[163] prompted a two-pronged approach to the cryptoregiochemical analysis of the Δ6 desaturase found in this protozoan.[164] Application of both the competitive KIE test and the thia test (*vide supra*) strongly suggested that Δ6 desaturation in this system is initiated at C6: a large primary deuterium KIE at C6 ($k_H/k_D = 7.1 \pm 0.5$) versus negligible KIE at C7 ($k_H/k_D = 1.04 \pm 0.05$) and preferential (>10-fold) sulfoxidation at S-6 versus S-7 of thia analogues were observed. In addition, the 6-sulfoxide produced by Δ6 desaturase-mediated oxo transfer was shown to be formed in high enantiomeric excess (>95% ee) and its absolute configuration (*S*) suggested that initial H-abstraction at C6 of the parent substrate occurs with the same stereochemical preference as that previously determined for oleate and linoleate biosynthesis (*vide supra*).

1.02.4 Unusual Unsaturated Fatty Acids

Fatty acids are considered unusual if they bear functionality other than methylene-interrupted double bonds or have uncommon chain lengths. A systematic search of hundreds of different plant species has revealed well over 1000 novel structures[16] and a searchable database for seed oil fatty acids (SOFA) has been constructed.[165] It is thought that the presence of these compounds confers an antifeedant property to the plant in question that is, or was, specific to its ecological role. Currently only a few of these compounds are available in sufficiently high volume to have industrial value. The most important of these is ricinoleate (**Figure 1**) – a highly valuable compound found in the seed oil of the castor plant. Expanding the range of compounds available as industrial feedstocks remains one of the goals of plant biotechnology as originally envisioned by Somerville and coworkers nearly two decades ago.[166]

For the bioorganic chemist, elucidating the catalytic mechanisms underlying the production of unusual fatty acids is an exciting and rewarding task; in some cases, new chemical pathways are uncovered that cannot be duplicated in the laboratory since conventional reagents lack the necessary selectivity. A unifying theme connecting the various mechanistic pathways is that they are mediated by desaturases or desaturase-like enzymes.

Indeed, a number of enzymes belonging to the plant oleate Δ12 desaturase (FAD2) subfamily generate a collection of lipids bearing a variety of unusual lipidic functional groups including allylic alcohols,[73] homoallylic alcohols,[77] conjugated dienes,[167–173] alkynes,[174] and epoxides[174] (**Scheme 10(b)**). One gets the sense that these compounds represent only a small fraction of the compounds waiting to be discovered.

1.02.4.1 Ricinoleic Acid

As alluded to earlier, ricinoleate (**Scheme 10(a)**) is an important natural raw material with great value as a petrochemical replacement in some important industrial processes. The seeds of castor plant

(*Ricinus communis L.*) are the major source of ricinoleate, which constitutes about 90% of the total fatty acids of the seed oil. However, oilseed castor cultivation is difficult to adapt to large-scale production owing to toxicity issues and a great deal of effort is being expended to produce ricinoleate in alternative oilseed crops through genetic engineering.[175] The >99.9% enantiomeric purity of this compound has not been exploited to our knowledge.

The mechanism of formation of ricinoleate can be understood in terms of funneling substrate through the hydroxyl rebound pathway as suggested by the mechanistic model displayed in **Scheme 3**. The hydroxylation/dehydrogenation connection became more obvious with the discovery of a *Lesquerella* bifunctional system that produces a mixture of 12-hydroxylated and 12-monoenoic product (**Scheme 10**(**a**)).[77] Indeed, wild type FAD2 found in the model plant *Arabidopsis thaliana* possesses inherent, low-level hydroxylation activity as do a number of other membrane-bound enzymes (*vide supra*).[62] It has also been shown that the ratio of 12-ol to 12-ene can be adjusted through site-directed mutagenesis experiments involving as few as four amino acid residues.[176] More recently, it has been shown that changes in amino acid at positions 148 and 324 of *Arabidopsis* FAD2 play an important role in determining the ratio of dehydrogenation to hydroxylation.[62] In addition, the stereochemistry of hydroxyl byproduct formation during desaturation (**Scheme 3**) matches the enantioselectivity of hydrogen removal at that position. As mentioned earlier, the two processes also share identical sites of initial hydrogen abstraction as determined by a cryptoregiochemical study of linoleate biosynthesis.[75]

1.02.4.2 Vernolic Acid

Epoxy fatty acids have high-value uses in glues, resins, and surface coatings and so there is some interest in seed oils with high epoxide content. These include oils found in *Vernonia galamensis*,[177] *Euphorbia lagascae*,[178] and *Crepis palaestina*.[174] The most prominent member of this family of compounds is vernolic acid (**Scheme 10**(**b**)) derived by epoxidation of linoleate. The absolute configuration at the epoxy center from *Vernonia* oil matches that expected of a FAD2 variant. Mechanistic investigations on epoxide formation are complicated by the lack of a good working hypothesis. To determine whether the reaction is stepwise, one would have to look for relatively small secondary isotope effects due to a change in coordination from sp2 to a strained sp3 center. In the case of vernolate formed in *C. palaestina*, it is now known that an acetylenase homologue is involved (*vide infra*). Vernolate found in *Euphorbia* is cytochrome P450-derived.[178]

1.02.4.3 Crepenynoic Acid

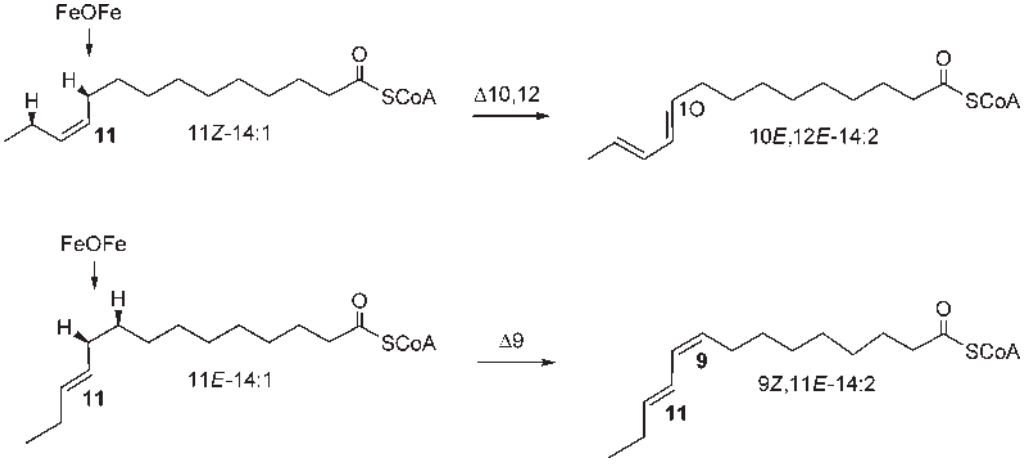
The dehydrogenation of linoleate to give crepenynate (**Scheme 10**(**b**)) is the first step along a pathway yielding a very large family of bioactive compounds known as polyacetylenes.[179,180] This remarkable transformation represents an excellent example of 'extreme' enzymatic behavior given that the strong vinyl C–H bond in the substrate is removed in preference to the energetically far more favorable allylic hydrogen abstraction or epoxidation processes. The latter would lead to the products shown in **Scheme 10**(**b**) via other reaction channels that are presumably discouraged by subtle details of acetylenase active site architecture. A major methodological advance in this area was achieved by the cloning of a cDNA from *Crepis alpina* (Asteraceae), which encodes the enzyme responsible for the conversion of linoleate to crepenynate.[174] Functional expression of this enzyme in yeast permitted a mechanistic study that looked at whether the site of initial oxidation for crepenynate formation was identical to that found for other FAD2-catalyzed processes, namely linoleate and ricinoleate formation.[181] This was precisely what was found. The operation of a large primary deuterium isotope effect (KIE = 14.6 ± 3.0) was demonstrated for the C–H bond cleavage at C12 while the C13–H bond breaking step was found to be relatively insensitive to deuterium substitution (KIE = 1.25 ±0.08). Interestingly, the 12-acetylenase also functions as a desaturase with oleate as substrate but produces a mixture of (*E*/*Z*) 12,13-olefinic isomers – an indication perhaps of a more capacious active site required to accommodate the linear acetylenic product in the parent reaction.[181] Other examples of acetylenase-catalyzed chemistry have been found to accompany Δ6 and Δ11 desaturation of fatty acids in the moss *Ceratodon purpureus*[161] and in the insect *Thaumetopoea* (Lepidoptera) respectively.[182] The cryptoregiochemistry of the Δ11 acetylenase has been examined using KIE techniques and found to be initiated at C11 as anticipated.[183]

1.02.4.4 Conjugated Fatty Acids

Conjugated linoleates (**Figure 1**) are thought to be highly beneficial for human health and are found in dairy products as a result of the action of rumen bacteria on a linoleate precursor.[184] The search for a means of producing these compounds in seed oils prompted a closer analysis of the occurrence of conjugated olefins in various plant species. Sequence analysis of plant genes responsible for the production of conjugated trienoic acids of the type displayed in **Scheme 10**(**b**) revealed the involvement of FAD2 variants.[167–173] The enzymes involved were called 'conjugases.' To avoid confusion with unrelated enzymes bearing this name, it would be more advisable to use the term '1,4-desaturase' or $\Delta^{x,y}$-desaturase. By inspection, it appears that these materials are biosynthesized by an apparent (1,4)-dehydrogenation process analogous to that proposed for the more common (1,2)-dehydrogenation reactions of fatty acid desaturases.

Support for this hypothesis had already been obtained by Crombie and Holloway[185] some years ago: the results of labeling experiments using marigold seed homogenates and labeled linoleate precursors demonstrated that calendic acid (**Scheme 10**(**b**)) is produced by loss of hydrogen from C8 and C11 and that no oxygenated intermediates could be detected.[185] The availability of a convenient yeast expression system for *Fac2* – a *Calendula officinalis* gene encoding the (1,4)-desaturase involved in calendic acid production – allowed a cryptoregiochemical study using KIE methodology to be carried out in convenient fashion.[186] It was found that the site of initial hydrogen abstraction for calendate formation is at C11 as would be expected for a FAD2 variant; recall that the parent enzyme initiates oxidation at C12 (**Scheme 10(a)**). Thus only a slight shift in the position of the iron oxidant relative to substrate is required to perform the 1,4-dehydrogenation reaction. An interesting mechanistic variant of 1,4-dehydrogenation has been discovered for dimorphecolate formation (**Scheme 10(c)**) – a pathway that would also be initiated at C11 but followed by a quench of the intermediate allylic radical by regioselective and enantioselective hydroxyl rebound.[73] A very good enzyme model for this process has recently been uncovered in serendipitous fashion (**Scheme 6**).[72]

Further evidence for the close relationship between 1,2 and 1,4-dehydrogenation was obtained for a *Spodoptera littoralis* Δ9 desaturating system which processes the (*E*)-11-tetradecenoate via C9-initiated dehydrogenation to give a 9(*Z*), 11(*E*)-diene. However, the 11(*Z*)-tetradecenoate is oxidized by the same enzyme to give (*E*,*E*)-10,12-tetradecadienoate by 1,4-desaturation (initial site of oxidation at C10). The stereochemistry and cryptoregiochemistry of both transformations is shown in **Scheme 11**.[187–190] A remarkable example of how stereochemical analysis can be used to gain indirect information on the topology of active sites was provided by the Fabrias group in their mechanistic study of an unusual Δ13 desaturation involved in the biosynthesis of *Thaumetopoea pityocampa* sex pheromone: the removal of the C13 and C14 hydrogens in 11-hexadecynoate and (11*Z*)-11-hexadecenoate are pro *R*- and pro *S*-specific *syn*-dehydrogenation processes, respectively. The results could be rationalized by molecular modeling of the substrate into a simulated active site with appropriate geometric restraints.[191]



Scheme 11 1,2- and 1,4-Dehydrogenation of myristoleyl CoA stereoisomers by *Spodoptera.*

1.02.4.5 Allenic Fatty Acids

Fatty acids containing an allenic moiety are rare although a recent review of the allenic natural products revealed about 150 examples.[125] The seed oil of the subtropical plant *Leonotis neptaefolia* (Lion's ear), a member of the mint family (Labiatae), was found to contain a 5,6 monoallenic octadecadienoic acid – laballenic acid (**Scheme 12**).[192] The absolute configuration of laballenic acid was determined by stereoselective synthesis.[193] A possible mechanism for formation of allenes would involve desaturation of a monoolefin via stepwise hydrogen abstraction similar to that proposed for the alkene–alkyne conversion (**Scheme 12**).

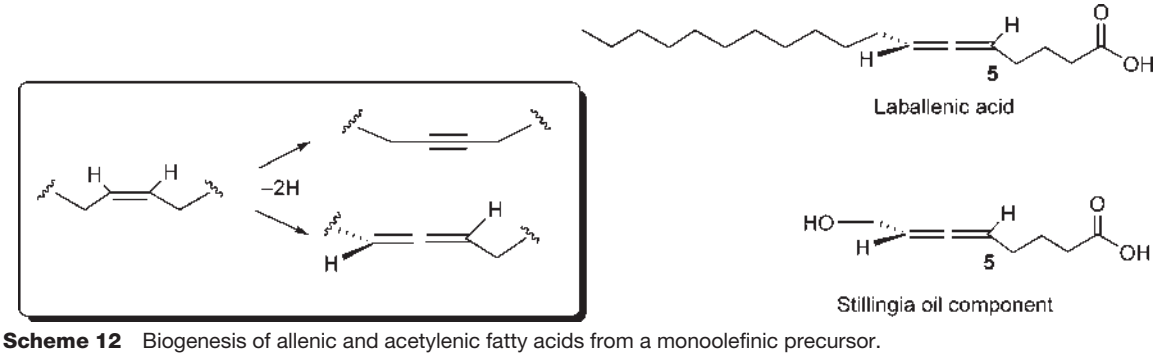
The biological role of allenic fatty acids is not obvious but they may serve as precursors to bioactive conjugated polyacetylenic/allenic compounds elaborated by fungi similar to phomoallenic acid, an antimicrobial compound discovered recently by Merck Research Laboratories.[194] A chain shortened allenic fatty acid has been found as a component of stillingia oil found in the seeds of *Sapium sebiferum* (Chinese tallow tree).[195]

1.02.4.6 Sterculic Acid

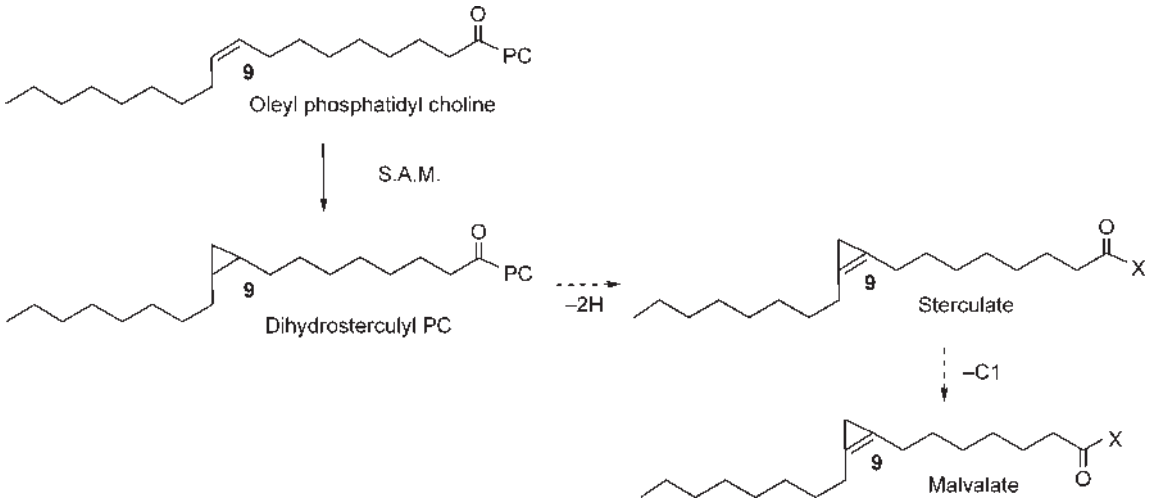
In 1952, Nunn correctly deduced the structure of sterculic acid (**Figure 1**), an unusually labile compound isolated from the seeds of the tropical tree *Sterculia foetida*.[196] This discovery initiated a systematic study of the natural occurrence of cyclopropenyl fatty acids and sterculate was found in the seed oils of many plant species belonging to the Sterculiaceae, Malvaceae, Tiliaceae, and Bombacaceae families.[197] The seeds of *S. foetida* (skunk tree) are unique in that they contain up to 78% of this highly strained compound. Cotton oil has far less cyclopropene fatty acid content (~1%), but the economic impact is considerable given the volume of its commercial production and that many physiological disorders arise in animals fed cottonseed. The latter phenomenon is due to the strong inhibition of mammalian stearoyl CoA desaturase by sterculate as alluded to in an earlier section. In fact, sterculia oil has been used in metabolic studies as a convenient means to assess the impact of an inactive SCD on other parameters.[198] The strong bioactivity of cyclopropenyl fatty acids strongly suggests that production of these compounds is part of an antifeedant strategy. The concept of using a cyclopropene ring to inhibit desaturases has inspired the Fabrias[134] and Baird groups[199] to synthesize appropriate substrate analogues bearing this unique functionality at appropriate positions along the fatty acyl chain.

The current hypothesis regarding the formation of this cyclopropenyl fatty acids features a pathway beginning with cyclopropanation of oleic acid followed by a desaturation at carbons 9 and 10 (**Scheme 13**). An apparent alpha oxidation produces malvalate in some species; 2-hydroxysterculate has been identified in some seed oils of the Bombacaceae family.[200] The cyclopropane synthase found in *S. foetida* has been characterized biochemically and functionally expressed by Ohlrogge and Pollard.[201,202]

The absolute configuration of the putative cyclopropyl intermediate is not known but the methodology is in place to make such a determination.[203] Desaturation of dihydrosterculate is unique in that it is the only known fatty acid desaturase to operate directly on two chiral methine rather than prochiral methylene centers. The gene responsible for the putative desaturase-mediated process has not been positively identified.



Scheme 12 Biogenesis of allenic and acetylenic fatty acids from a monoolefinic precursor.



Scheme 13 Putatative biosynthetic pathway of cyclopropenyl fatty acids.

1.02.5 Summary and Future Prospects

The invention and application of novel mechanistic probes (deutero-, fluoro-, thia-containing substrate analogues) has led to a more sophisticated understanding of how fatty acid desaturases and related enzymes carry out their ultra-selective oxidation chemistry.

Three mechanistic issues with respect to desaturase mechanism have been addressed with varying levels of success. These are (1) stereochemistry of hydrogen removal, (2) the site of initial oxidative attack (cryptoregiochemistry), and (3) the mechanistic switch governing dehydrogenation versus oxygenation pathways. Items (1) and (2) are easily answered in the case of membranous desaturases because of the convenient microbial expressions that are now available. All desaturases of this class initiate oxidation at the carbon closest to the C1 terminus. For the one soluble desaturase that has been examined so far, the cryptoregiochemical preference appears to be in the opposite sense. For both classes of desaturases *syn*, pro *R* selective removal of proximal hydrogens is observed.

This mechanistic information can now serve as a useful platform for the advancement of more applied projects including the design of medically relevant desaturase inhibitors and engineering new catalytic activities for the seed oil commodity market. With respect to the former, there is great scope for additional research activity in the lipid biochemistry of cell signaling (sphingolipids), metabolic syndrome (SCD), and tuberculosis (desaturases). In the area of plant biotechnology, more molecular biological work is required to provide bioorganic chemists with functionally expressed enzymes catalyzing the exotic biochemistry of desaturase variants. Common to both types of projects is the urgent need for more direct information on active site architecture. The structural elucidation of membrane-bound desaturases constitutes the next frontier of research in this area.

Abbreviations

ACP	acyl carrier protein
ALA	α-linolenic acid
CLA	conjugated linoleic acid
CoA	coenzyme A
EPA	eicosapentaenoic acid
GLA	γ-linolenic acid

KIE	kinetic isotope effect
MCD	magnetic circular dichroism
MMO	methane monooxygenase
NADH	nicotinamide adenine dinucleotide
PL	phospholipid
PUFA	polyunsaturated fatty acids
SCD	stearoyl CoA desaturase
SOFA	seed oil fatty acids
TAG	triacylglyceride
TOM	toluene-ortho-monooxygenase

References

1. M. I. Gurr; J. L. Harwood; K. N. Frayn, *Lipid Biochemistry: An Introduction*, 5th ed.; Blackwell Science: Oxford, 2002; pp 242–244.
2. K. M. Schmid; J. B. Ohlrogge, Lipid Metabolism in Plants. In *Biochemistry of Lipids, Lipoproteins and Membranes,* 4th ed.; D. E. Vance, J. E. Vance, Eds.; Elsevier: Amsterdam, 2002; pp 105–109.
3. F. J. van de Loo; B. G. Fox; C. Somerville, Unusual Fatty Acids. In *Lipid Metabolism in Plants*; T. S. Moore, Jr,Ed.; CRC Press: Boca Raton, 1993; pp 91–126.
4. N. J. Oldham; W. Boland, *Naturwissenschaften* **1996**, *83*, 248.
5. R. W. Howard; G. J. Blomquist, *Ann. Rev. Entomol.* **2005**, *50*, 371.
6. L. He; H.-S. Byun; J. Smit; J. Wilschut; R. J. Bittman, *J. Am. Chem. Soc.* **1999**, *121*, 3897.
7. I. Feussner; C. Wasternack, *Annu. Rev. Plant Biol.* **2002**, *53*, 275.
8. W. L. Smith; R. C. Murphy; D. E. Vance; J. E. Vance, Eds., *Biochemistry of Lipids, Lipoproteins and Membranes,* 4th ed; Elsevier: Amsterdam, 2002; pp 341–371.
9. J. M. Dyer; S. Stymne; A. G. Green; A. S. Carlsson, *Plant J.* **2008**, *54*, 640.
10. J. A. Napier, *Annu Rev. Plant. Biol.* **2007**, *58*, 295.
11. A. Dobrzyn; J. M. Ntambi, *Prostaglandins Leukot. Essent. Fatty Acids* **2005**, *73*, 35.
12. B. Phetsuksiri; M. Jackson; H. Scherman; M. McNeil; G. S. Besra; A. R. Baulard; R. A. Slayden; A. E. DeBarber; C. E. Barry; M. S. Baird; D. C. Crick; P. J. Brennan, *J. Biol. Chem.* **2003**, *278*, 53123.
13. W. Zheng; J. Kollmeyer; H. Symolon; A. Momin; E. Munter; E. Wang; S. Kelly; J. C. Allegood; Y. Liu; Q. Peng; H. Ramaraju; M. C. Sullards; M. Cabot; A. H. Merrill, Jr., *Biochim. Biophys. Acta* **2006**, *1758*, 1864.
14. A. Kawaguchi; A. Iwamoto-Kihara; N. Sato, Biosynthesis and Degradation of Fatty Acids. In *Comprehensive Natural Products Chemistry, Volume1: Polyketides and Other Secondary Metabolites Including Fatty Acids and Their Derivatives*; D. H. R. Barton, O. Meth-Cohn, K. Nakanishi, Y. Sankawa (volume editors); Elsevier: Oxford, 1999; pp 23–59.
15. P. H. Buist, *Nat. Prod. Rep.* **2004**, *21*, 249.
16. F. D. Gunstone, *Biochim. Biophys. Acta* **2003**, *1631*, 207.
17. P. Sperling; P. Ternes; T. K. Zank; E. Heinz, *Prostag. Leukotr. Ess.* **2003**, *68*, 73.
18. J. Shanklin; E. B. Cahoon, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1998**, *49*, 611.
19. B. G. Fox; K. S. Lyle; C. E. Rogge, *Acc. Chem. Res.* **2004**, *37*, 421.
20. J. M. Dyer; D. C. Chapital; J. W. Kuan; R. T. Mullen; A. B. Pepperman, *Appl. Microbiol. Biotechnol.* **2002**, *59*, 224.
21. C. R. Somerville; D. Bonetta, *Plant Physiol.* **2001**, *125*, 168.
22. J. Jaworski; E. B. Cahoon, *Curr. Opin. Plant Biotech.* **2003**, *2*, 178.
23. M. Inaba; I. Suzuki; B. Szalontai; Y. Kanesaki; D. A. Los; H. Hayashi; N. Murata, *J. Biol. Chem.* **2003**, *278*, 12191.
24. J. G. Wallis; J. Browse, *Prog. Lip. Res.* **2002**, *41*, 254.
25. J. M. Ntambi; M. Miyazaki, *Curr. Opin. Lipid.* **2003**, *14*, 255.
26. W. L. Roelofs, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 44.
27. Y. A. Hannun; L. M. Obeid, *Nat. Rev. Mol. Cell. Biol.* **2008**, *9*, 139.
28. M. I. Gurr; J. L. Harwood; K. N. Frayn, *Lipid Biochemistry: An Introduction*, 5th ed.; Blackwell Science: Oxford, 2002.
29. D. E. Vance; J. E. Vance, Eds., *Biochemistry of Lipids, Lipoproteins and Membranes,* 4th ed.; Elsevier: Amsterdam, 2002.
30. K. Bloch, *Acc. Chem. Res.* **1969**, *2*, 193.
31. R. R. Brenner, *Prostag. Leukotr. Ess.* **2003**, *68*, 71.
32. R. J. Heath, S. Jackowski, C. O. Rock, Lipid Metabolism in Plants. In *Biochemistry of Lipids, Lipoproteins and Membranes,* 4th ed.; D. E. Vance, J. E. Vance, Eds.; Elsevier: Amsterdam, 2002; pp 62–63.
33. K. Zhu; K. H. Choi; H. P. Schweizer; C. O. Rock; Y. M. Zhang, *Mol. Microbiol.* **2006**, *60*, 260.
34. P. Le Couteur; J. Burreson, *Napoleon's Buttons. 17 Molecules That Changed History*; Tarcher/Penguin: New York, 2003.
35. C. Bowsher; M. Steer; A. Tobin, *Plant Biochemistry*; Garland Science: New York, 2008; p 326.
36. E. Heinz, Biosynthesis of Polyunsaturated Fatty Acids. In *Lipid Metabolism in Plants*; T. S. Moore, Jr., Ed.; CRC Press: Boca Raton, 1993; p 34.
37. K. M. Schmid; J. B. Ohlrogge, Lipid Metabolism in Plants. In *Biochemistry of Lipids, Lipoproteins and Membranes,* 4th ed.; D. E. Vance, J. E. Vance, Eds.; Elsevier: Amsterdam, 2002; pp 107–108.

38. J. Shanklin; C. Somerville, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2510.
39. G. A. Thompson; D. E. Scherer; S. Foxall-Van Aken; J. W. Kenny; H. L. Young; D. K. Shintani; J. C. Kridl; V. C. Knauf, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2578.
40. B. G. Fox; J. Shanklin; C. Somerville; E. Munck, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2486.
41. B. G. Fox; J. Shanklin; J. Y. Ay; T. M. Loehr; J. Sanders-Loehr, *J. Biochemistry* **1994**, *33*, 12776.
42. Y. Lindqvist; W. Huang; G. Schneider; J. Shanklin, *EMBO J.* **1996**, *15*, 4081.
43. Y. S. Yang; J. Broadwater; S. C. Pulver; B. G. Fox; E. I. Solomon, *J. Am. Chem. Soc.* **1999**, *121*, 2770.
44. M. H. Baik; M. Newcomb; R. A. Friesner; S. J. Lippard, *Chem. Rev.* **2003**, *103*, 2385.
45. B. Behrouzian; C. K. Savile; B. Dawson; P. H. Buist; J. Shanklin, *J. Am. Chem. Soc.* **2002**, *124*, 3277.
46. M. Moche; J. Shanklin; A. Ghoshal; Y. Lindqvist, *J. Biol. Chem.* **2003**, *278*, 25072.
47. J. E. Guy; I. A. Abreu; M. Moche; Y. Lindqvist; E. Whittle; J. Shanklin, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17220.
48. D. H. Dyer; K. S. Lyle; I. Rayment; B. G. Fox, *Protein Sci.* **2005**, *14*, 1508.
49. J. A. Haas; B. G. Fox, *Biochemistry* **2002**, *41*, 14472.
50. V. Reipa; J. Shanklin; V. Vilker, *Chem. Commun. (Camb)* **2004**, *21*, 2406.
51. P. Sobrado; K. S. Lyle; S. P. Kaul; M. M. Turco; I. Arabshahi; A. Marwah; B. G. Fox, *Biochemistry* **2006**, *45*, 4848.
52. J. K. Schwartz; P. P. Wei; K. H. Mitchell; B. G. Fox; E. I. Solomon, *J. Am. Chem. Soc.* **2008**, *130*, 7098.
53. B. J. Wallar; J. D. Lipscomb, *Biochemistry* **2001**, *40*, 2220.
54. J. A. Broadwater; B. J. Laundre; B. G. Fox, *J. Inorg. Biochem.* **2000**, *78*, 7.
55. E. Rettie; A. W. Rettenmeier; W. N. Howald; T. A. Baillie, *Science* **1987**, *235*, 890.
56. P. H. Buist; D. M. Marecak, *J. Am. Chem. Soc.* **1992**, *114*, 5073.
57. M. Akhtar; J. N. Wright, *Nat. Prod. Rep.* **1991**, *8*, 527.
58. L. Shu; J. C. Nesheim; K. Kauffmann; E. Munck; J. D. Lipscomb; L. Que, Jr., *Science* **1997**, *275*, 515.
59. D. Rinaldo; D. M. Philipp; S. J. Lippard; R. A. Friesner, *J. Am. Chem. Soc.* **2007**, *129*, 3135.
60. J. T. Groves; G. A. McClusky; R. E. White; M. J. Coon, *Biochem. Biophys. Res. Commun.* **1978**, *81*, 154.
61. V. W. Bowry; K. U. Ingold, *J. Am. Chem. Soc.* **1991**, *113*, 5699.
62. J. A. Broadwater; E. Whittle; J. Shanklin, *J. Biol. Chem.* **2002**, *277*, 15613.
63. R. J. Light; W. J. Lennarz; K. Bloch, *J. Biol. Chem.* **1962**, *237*, 1793.
64. P. H. Buist; B. Behrouzian, *J. Am. Chem. Soc.* **1996**, *118*, 6295.
65. B. Behrouzian; P. H. Buist; J. Shanklin, *J. Chem. Soc. Chem. Commun.* **2001**, 411.
66. K. S. Lyle; J. A. Haas; B. G. Fox, *Biochemistry* **2003**, *42*, 5857.
67. B. Behrouzian; D. Hodgson; C. K. Savile; B. Dawson; P. H. Buist; J. Shanklin, *Mag. Res. Chem.* **2002**, *40*, 524.
68. A. E. Tremblay; P. H. Buist; D. Hodgson; B. Dawson; E. Whittle; J. Shanklin, *Magn. Reson. Chem.* **2006**, *44*, 629.
69. R. D. White; B. G. Fox, *Biochemistry* **2003**, *42*, 7828.
70. C. E. Rogge; B. G. Fox, *Biochemistry* **2002**, *41*, 10141.
71. B. Behrouzian; P. H. Buist; J. Shanklin, *J. Chem. Soc. Chem. Commun.* **2001**, 765.
72. E. Whittle; A. E. Tremblay; P. H. Buist; J. Shanklin, *Proc. Natl. Acad. Sci.* **2008**, *105*, 14738.
73. E. B. Cahoon; A. J. Kinney, *J. Biol. Chem.* **2004**, *279*, 12495.
74. J. Zhou; W. L. Kelly; B. O. Bachmann; M. Gunsior; C. A. Townsend; E. I. Solomon, *J. Am. Chem. Soc.* **2001**, *123*, 7388.
75. P. H. Buist; B. Behrouzian, *J. Am. Chem. Soc.* **1998**, *120*, 871.
76. P. R. Ortiz de Montellano, *Trends Pharm. Sci.* **1989**, *10*, 354.
77. F. J. Van de Loo; P. Broun; S. Turner; C. Summerville, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6743.
78. D. Kumar; S. P. De Visser; S. Shaik, *J. Am. Chem. Soc.* **2004**, *126*, 5072.
79. K. Bloch, Oxygenases in Lipid Biosynthesis. In *Oxygenases and Oxygen Metabolism*; M. Nozaki, S. Yamamoto, Y. Ishimura, M. J. Coon, L. Ernster, R. W. Eastabrook, Eds.; Academic Press: New York, 1982; pp 645–653.
80. D. K. Bloomfield; K. Bloch, *Biochem. Biophys. Acta* **1958**, *30*, 220.
81. D. K. Bloomfield; K. Bloch, *J. Biol. Chem.* **1960**, *235*, 337.
82. P. Strittmatter; L. Spatz; D. Corcoran; M. J. Rogers; B. Setlow; R. Redline, *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 4565.
83. H. G. Enoch; A. Catalá; P. Strittmatter, *J. Biol. Chem.* **1976**, *251*, 5095.
84. M. Thiede; J. Ozols; P. Stittmatter, *J. Biol. Chem.* **1986**, *261*, 13230.
85. J. E. Stukey; V. M. McDonough; C. E. Martin, *J. Biol. Chem.* **1990**, *265*, 20144.
86. J. Shanklin; E. Whittle; B. G. Fox, *Biochemistry* **1994**, *33*, 12787.
87. J. Shanklin; C. Achim; H. Schmidt; B. G. Fox; E. Munck, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2981.
88. P. Sperling; E. Heinz; *Eur. J. Lipid Sci. Technol.* **2001**, *103*, 158.
89. G. J. Schroepfer; K. Bloch, *J. Biol. Chem.* **1965**, *240*, 54.
90. L. J. Morris; R. V. Harris; W. Kelly; A. T. James, *Biochem. J.* **1968**, *109*, 673.
91. A. R. Johnson; M. I. Gurr, *Lipids* **1971**, *6*, 78.
92. D. Meesapyodsuk; D. W. Reed; S. Cheevadhanarak; P. Deshnium; P. S. Covello, *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **2001**, *129*, 831.
93. B. Behrouzian; L. Fauconnot; F. Daligault; C. Nugier-Chauvin; H. Patin; P. H. Buist, *Eur. J. Biochem.* **2001**, *268*, 3545.
94. J. P. Jones; W. F. Trager, *J. Am. Chem. Soc.* **1987**, *109*, 2171.
95. D. J. Hodgson; K. Y. Y. Lao; B. Dawson; P. H. Buist, *Helv. Chim. Acta* **2003**, *86*, 3688.
96. F. Carvalho; L. T. Gauthier; D. J. Hodgson; B. Dawson; P. H. Buist, *Org. Biomol. Chem.* **2005**, *3*, 3979.
97. D. J. Hodgson; P. H. Buist, *Tetrahedron Asymmetry* **2003**, *14*, 641.
98. G. Villorbina; L. Roura; F. Camps; J. Joglar; G. Fabrias, *J. Org. Chem.* **2003**, *68*, 2820.
99. J. Shanklin; E. Whittle, *FEBS Lett.* **2003**, *545*, 188.
100. E. A. Rozhkova-Novosad; J. C. Chae; G. J. Zylstra; E. M. Bertrand; M. Alexander Ozinskas; D. Deng; L. A. Moe; J. B. van Beilen; M. Danahy; J. T. Groves; R. N. Austin, *Chem. Biol.* **2007**, *14*, 165.
101. E. Bertrand; R. Sakai; E. Rozhkova-Novosad; L. Moe; B. G. Fox; J. T. Groves; R. N. Austin, *J. Inorg. Biochem.* **2005**, *99*, 1998.

102. Z. Xin; H. Zhao; M. D. Serby; B. Liu; M. Liu; B. G. Szczepankiewicz; L. T. Nelson; H. T. Smith; T. S. Suhar; R. S. Janis; N. Cao; H. S. Camp; C. A. Collins; H. L. Sham; T. K. Surowy; G. Liu, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4298.
103. G. Liu; J. K. Lynch; J. Freeman; B. Liu; Z. Xin; H. Zhao; M. D. Serby; P. R. Kym; T. S. Suhar; H. T. Smith; N. Cao; R. Yang; R. S. Janis; J. A. Krauser; S. P. Cepa; D. W. Beno; H. L. Sham; C. A. Collins; T. K. Surowy TK; H. S. Camp, *J. Med. Chem.* **2007**, *50*, 3086.
104. M. Hoffmann; E. Hornung; S. Busch; N. Kassner; P. Ternes; G. H. Braus; I. Feussner, *J. Biol. Chem.* **2007**, *282*, 26666.
105. D. Meesapyodsuk; D. W. Reed; P. S. Covello; X. Qiu, *J. Biol. Chem.* **2007**, *282*, 20191.
106. E. B. Cahoon; J. B. Ohlrogge, *Plant Physiol.* **1994**, *104*, 827.
107. J. E. Guy; E. Whittle; D. Kumaran; Y. Lindqvist; J. Shanklin, *J. Biol. Chem.* **2007**, *282*, 19863.
108. A. E. Tremblay; E. Whittle; P. H. Buist; J. Shanklin, *Org. Biomol. Chem.* **2007**, *5*, 1270.
109. D. Lee; S. J. Lippard, *Inorg. Chem.* **2002**, *41*, 2704.
110. A. J. Fulco, *Biochim. Biophys. Acta* **1967**, *144*, 701.
111. M. C. Mansilla; P. S. Aguilar; D. Albenesi; L. E. Cybulski; S. Altabe; D. de Mendoza, *Prostag. Leukotr. Ess.* **2003**, *68*, 187.
112. P. S. Aguilar; J. E. Cronan; D. de Mendoza, *J. Bacteriol.* **1998**, *180*, 2194.
113. R. Diaz; M. C. Mansilla; A. J. Vila; D. de Mendoza, *J. Biol. Chem.* **2002**, *277*, 48099.
114. L. Fauconnot; P. H. Buist, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2879.
115. E. B. Cahoon; A. M. Cranmer; J. Shanklin, *J. Biol. Chem.* **1994**, *269*, 27519.
116. E. B. Cahoon; Y. Lindqvist; G. Schneider; J. Shanklin, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 4872.
117. L. B. Bjostad; W. L. Roelofs, *J. Biol. Chem.* **1981**, *256*, 7936.
118. W. Boland; C. Frössl; M. Schottler; M. Toth, *J. Chem. Soc. Chem. Comm.* **1993**, 1155.
119. A. Svatos; B. Kalinova; W. Boland, *Insect Biochem. Mol. Biol.* **1999**, *29*, 225.
120. W. Liu; H. Jiao; M. O'Connor; W. L. Roelofs, *Insect Biochem. Mol. Biol.* **2002**, *32*, 1489.
121. A. Pinilla; F. Camps; G. Fabrias, *Biochemistry* **1999**, *38*, 15272.
122. I. Navarro; I. Font; G. Fabrias; F. Camps, *J. Am. Chem. Soc.* **1997**, *119*, 11335.
123. S. Rodríguez; G. Hao; W. Liu; B. Piña; A. P. Rooney; F. Camps; W. L. Roelofs; G. Fabriàs, *Insect Biochem. Mol. Biol.* **2004**, *34*, 1315.
124. M. Serra; L. T. Gauthier; G. Fabrias; P. H. Buist, *Insect Biochem. Mol. Biol.* **2006**, *36*, 822.
125. A. Hoffmann-Röder; N. Krause, *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 1196.
126. S. Pyne; N. Pyne, *Biochem. J.* **2000**, *67*, 27.
127. K. Endo; T. Akiyama; S. Kobayashi; M. Okada, *Mol. Gen. Genet.* **1996**, *253*, 157.
128. C. K.-Y. Ng; K. Carr; M. R. McAinsh; B. Powell; A. M. Hetherington, *Nature* **2001**, *410*, 596.
129. S. Kim; H. Fyrst; Julie Saba, *Genetics* **2000**, *156*, 1519.
130. P. Ternes; S. Franke; U. Zahringer; P. Sperling; E. Heinz, *J. Biol. Chem.* **2002**, *277*, 25512.
131. W. Stoffel; G. Assmann; K. Bister, *Hoppe-Seyler's Z. Physiol. Chem.* **1971**, *352*, 1531.
132. C. Beckmann; J. Rattke; P. Sperling; E. Heinz; W. Boland, *Org. Biomol. Chem.* **2003**, *1*, 2448.
133. C. K. Savile; G. Fabriàs; P. H. Buist, *J. Am. Chem. Soc.* **2001**, *123*, 4382–4385.
134. G. Triola; G. Fabrias; A. Llebaria, *Angew. Chem. Int. Ed.* **2001**, *40*, 1960.
135. J. M. Kraveka; L. Li; Z. M. Szulc; J. Bielawski; B. Ogretmen; Y. A. Hannun; L. M. Obeid; A. Bielawska, *J. Biol. Chem.* **2007**, *282*, 16718.
136. G. Triola; G. Fabrias; J. Casas; A. Llebaria, *J. Org. Chem.* **2003**, *68*, 9924.
137. J. M. Munoz-Olaya; X. Matabosch; C. Bedia; M. Egido-Gabás; J. Casas; A. Llebaria; A. Delgado; G. Fabriàs, *ChemMedChem.* **2008**, *3*, 946.
138. P. Sperling; U. Zahringer; E. Heinz, *J. Biol. Chem.* **1998**, *273*, 28590.
139. F. García-Maroto; J. A. Garrido-Cárdenas; L. V. Michaelson; J. A. Napier; D. L. Alonso, *Plant Mol. Biol.* **2007**, *64*, 241.
140. C. Beckmann; J. Rattke; N. J. Oldham; P. Sperling; E. Heinz; W. Boland, *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 2298.
141. A. R. Cossins; P. A. Murray; A. Y. Gracey; J. Logue; S. Polley; M. Caddick; S. Brooks; T. Postle; N. Maclean, *Biochem. Soc. Trans.* **2002**, *30*, 1082.
142. E. Heinz, Biosynthesis of Polyunsaturated Fatty Acids. In *Lipid Metabolism in Plants*; T. S. Morre, Jr., Ed.; CRC Press: Boca Raton, 1993; pp 33–89.
143. J. L. Schwartzbeck; S. Jung; A. G. Abbott; E. Mosley; S. Lewis; G. Pries; G. L. Powell, *Phytochemistry* **2001**, *57*, 643.
144. P. S. Covello; D. W. Reed, *Plant Physiol.* **1996**, *111*, 223.
145. H. Wada; M. H. Avelange-Macherel; N. Murata, *J. Bacteriol.* **1993**, *175*, 6056.
146. S. Panpoon; D. Los; N. Murata, *Biochim. Biophys. Acts.* **1998**, *323*, 1390.
147. M. I. Gurr; J. L. Harwood; K. N. Frayn, *Lipid Biochemistry: An Introduction*, 5th ed.; Blackwell Science: Oxford, 2002; p 146.
148. O. Sayanova; R. Haslam; I. Guschina; D. Lloyd; W. W. Christie; J. L. Harwood; J. A. Napier, *J. Biol. Chem.* **2006**, *281*, 36533.
149. Z. Pan; A. M. Rimando; S. R. Baerson; M. Fishbein; S. O. Duke, *J. Biol. Chem.* **2007**, *282*, 4326.
150. D. Meesapyodsuk; D. W. Reed; C. K. Savile; P. H. Buist; P. S. Covello, *Biochemistry.* **2000**, *39*, 11948.
151. C. K. Savile; D. W. Reed; D. Meesapyodsuk; P. S. Covello; P. H. Buist, *J. Chem. Soc. Perkins Trans. 1.* **2001**, 1116.
152. J. P. Spychalla; A. J. Kinney; J. Browse, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 1142.
153. D. W. Reed; U. A. Schäfer; P. S. Covello, *Plant Physiol.* **2000**, *122*, 715.
154. F. Daligault; D. W. Reed; C. K. Savile; C. Nugier-Chauvin; H. Patin; P. S. Covello; P. H. Buist, *Phytochemistry* **2003**, *63*, 739.
155. S. L. Pereira; A. E. Leonard; P. Mukerji, *Prostag. Leukotr. Ess.,* **2003**, *68*, 97.
156. W. Stoffel; B. Holz; B. Jenke; E. Binczek; R. H. Günter; C. Kiss; I. Karakesisoglou; M. Thevis; A. A. Weber; S. Arnhold; K. Addicks; E. M. de Gyves; C. A. Sparks; O. Sayanova; P. Lazzeri; J. A. Napier; H. D. Jones, *Plant Biotechnol. J.* **2004**, *2*, 351.
157. Y.-S. Huang; V. A. Ziboh, Ed., *γ-Linolenic Acid, Recent Advances in Biotechnology and Clinical Applications*; AOCS Press: Champaign, 2000.
158. E. M. de Gyves; C. A. Sparks; O. Sayanova; P. Lazzeri; J. A. Napier; H. D. Jones, *Plant Biotechnol. J.* **2004**, *2*, 351.
159. O. Sayanova; M. A. Smith; P. Lapinskas; A. K. Stobart; G. Dobson; W. W. Christie; P. R. Shewry; J. A. Napier, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 4211.

160. B. Libisch; L. V. Michaelson; M. J. Lewis; P. R. Shewry; J. A. Napier, *Biochem. Biophys. Res. Commun.* **2000**, *279*, 779.
161. P. Sperling; M. Lee; G. Thomas; U. Zahringer; S. Stymne; E. Heinz, *Eur. J. Biochem.* **2000**, *267*, 3801.
162. N. Hastings; M. Agaba; D. R. Tocher; M. J. Leaver; J. R. Dick; J. R. Sargent; A. J. Teale, *Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14304.
163. J. E. Baenziger; I. C. P. Smith; R. J. Hill, *Chem. Phys. Lipids.* **1990**, *54*, 17.
164. L. Fauconnot; P. H. Buist, *J. Org. Chem.* **2001**, *66*, 1210.
165. K. Aitzetmuller; B. Matthaus; H. Friedrich, *Eur. J. Lipid Sci. Technol.* **2003**, *105*, 92.
166. C. Somerville; J. Browse, *Science* **1991**, *252*, 80.
167. E. B. Cahoon; T. J. Carlson; K. G. Ripp; B. J. Schweiger; G. A. Cook; S. E. Hall; A. J. Kinney, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 12935.
168. K. Fritsche; E. Homung; N. Peitzch; A. Renz; I. Feussner, *FEBS Lett.* **1999**, *462*, 249.
169. X. Qiu; D. W. Reed; H. Hong; S. L. Mackenzie; P. S. Covello, *Plant Physiol.* **2001**, *125*, 587.
170. E. B. Cahoon; K. G. Ripp; S. E. Hall; A. J. Kinney, *J. Biol. Chem.* **2001**, *276*, 2637.
171. E. Homung; C. Pernstich; I. Feussner, *Eur. J. Biochem.* **2002**, *269*, 4852.
172. J. M. Dyer; D. C. Chapital; J. C. W. Kuan; R. T. Mullen; C. Turner; T. A. McKeon; A. B. Pepperman, *Plant Physiol.* **2002**, *130*, 2027.
173. M. Iwabuchi; J. Kohno-Murase; J. Imamura, *J. Biol. Chem.* **2003**, *278*, 4603.
174. M. Lee; M. Lenman; A. Banas; M. Bafor; S. Singh; M. Schweizerm; R. Nilsson; C. Liljenberg; A. Dahlqvist; P.-O. Gummeson; S. Sjödahl; A. Green; S. Stymne, *Science* **1998**, *280*, 915.
175. C. Lu; J. G. Wallis; J. Browse, *BMC Plant Biol.* **2007**, *7*, 42.
176. P. Broun; J. Shanklin; E. Whittle; C. Somerville, *Science* **1998**, *282*, 1315.
177. I. Billault; J. R. Duan; S. Guiet; R. J. Robins, *J. Biol. Chem.* **2005**, *280*, 17645.
178. E. B. Cahoon; K. G. Ripp; S. E. Hall; B. McGonigle, *Plant Physiol.* **2002**, *128*, 615.
179. R. E. Minto; B. J. Blacklock, *Prog. Lipid. Res.* **2008**, *47*, 233.
180. E. B. Cahoon; J. A. Schnurr; E. A. Huffman; R. E. Minto, *Plant J.* **2003**, *34*, 671.
181. D. W. Reed; D. R. Polichuk; P. H. Buist; S. J. Ambrose; R. J. Sasata; C. K. Savile; A. R. S. Ross; P. S. Covello, *J. Am. Chem. Soc.* **2003**, *125*, 10635.
182. G. Villorbina; S. Rodriguez; F. Camps; G. Fabrias, *Insect Biochem. Mol. Biol.* **2003**, *33*, 155.
183. J. L. Abad; S. Rodríguez; F. Camps; G. Fabriàs, *J. Org. Chem.* **2006**, *71*, 7558.
184. C. Ip; S. Banni; E. Angioni; G. Carta; J. McGinley; H. J. Thompson; D. Barbano; D. Bauman, *J. Nutr.* **1999**, *129*, 2135.
185. L. Crombie; S. J. Holloway, *J. Chem. Soc. Perkin Trans. 1* **1985**, 2425.
186. D. W. Reed; C. K. Savile; X. Qui; P. H. Buist; P. S. Covello, *Eur. J. Biochem.* **2002**, *269*, 5024.
187. J. L. Abad; F. Camps; G. Fabrias, *Angew. Chem. Int. Ed.* **2000**, *39*, 3279.
188. S. Rodriguez; F. Camps; G. Fabrias, *J. Org. Chem.* **2001**, *66*, 8052.
189. J. L. Abad; F. Camps; G. Fabrias, *Insect Biochem. Mol. Biol.* **2001**, *31*, 799.
190. S. Rodriguez; P. Clapes; F. Camps; G. Fabrias, *J. Org. Chem.* **2002**, *67*, 2228.
191. J. L. Abad; F. Camps; G. Fabriàs, *J. Am. Chem. Soc.* **2007**, *129*, 15007.
192. M. O. Bagby; C. R. Smith, Jr.; I. A. Wolff, *J. Org. Chem.* **1965**, *30*, 4227.
193. J. S. Cowie; P. D. Landor; N. Punja; *J. Chem. Soc. Perkin Trans. I* **1972**, 2197.
194. J. G. Ondeyka; D. L. Zink; R. Young; R. Painter; S. Kodali; A. Galgoci; J. Collado; J. R. Tormo; A. Basilio; F. Vicente; J. Wang; S. B. Singh, *J. Nat. Prod.* **2006**, *69*, 377.
195. H. W. Sprecher; R. Maier; M. Barber; R. T. Holman, *Biochemistry* **1965**, *4*, 1863.
196. J. R. Nunn, *J. Chem. Soc.* **1952**, 313.
197. I. Yano; L. J. Morris; B. W. Nichols; A. T. James, *Lipids* **1972**, *7*, 35.
198. F. E. Gomez; D. E. Bauman; J. M. Ntambi; B. G. Fox, *Biochem. Biophys. Res. Commun.* **2003**, *300*, 316.
199. S. Hartmann; D. E. Minniken; H. J. Romming; M. S. Baird; C. Ratleadge; P. R. Wheeler, *Chem. Phys. Lipids* **1994**, *71*, 99.
200. L. J. Morris; S. W. Hall, *Chem. Ind. (London)* **1967**, 32.
201. X. M. Bao; S. Katz; M. Pollard; J. Ohlrogge, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7172.
202. X. M. Bao; J. J. Thelen; G. Bonaventure; J. B. Ohlrogge, *J. Biol. Chem.* **2003**, *278*, 12846.
203. L. J. Stuart; J. P. Buck; A. E. Tremblay; P. H. Buist, *Org. Lett.* **2006**, *8*, 79.

Biographical Sketch

Peter H. Buist is a Professor of Chemistry at Carleton University, Ottawa, Canada. He obtained his undergraduate and graduate degrees from the Department of Chemistry at McMaster University, Hamilton, Canada, where he studied mechanism of P450-catalyzed hydroxylation and cyclopropane fatty acid biosynthesis (1975–80). His mentors at McMaster were Bert Holland, Ian Spenser, and David Maclean. He completed a NATO-sponsored postdoctoral fellowship at the ETH-Zurich under the supervision of Duilio Arigoni (1980–82) and was involved in the application of chiral methyl group methodology to the study of sulfur insertion reactions in lipoic acid and biotin biosynthesis. He joined Carleton University as an Assistant Professor in 1982. His progress through the ranks of academia was guided by the sound advice of Karl Diedrich, Gerry Buchanan, Don Wiles, and Don Wigfield. His research program has focused on the mechanistic study of unusual enzymatic transformations that lack precedent in organic synthesis with an emphasis on fatty acid desaturation and cognate processes. He has enjoyed many fruitful collaborations with the research groups led by Pat Covello (PBI Saskatoon), Gemma Fabrias (IIQAB-CSIC Barcelona), Henri Patin (Ecole Nationale Supérieure de Chimie de Rennes), and John Shanklin (Brookhaven National Labs, NY).

1.03 Prostaglandin Endoperoxide Synthases: Structure, Function, and Synthesis of Novel Lipid Signaling Molecules

Melissa V. Turman and Lawrence J. Marnett, Vanderbilt University School of Medicine, Nashville, TN, USA

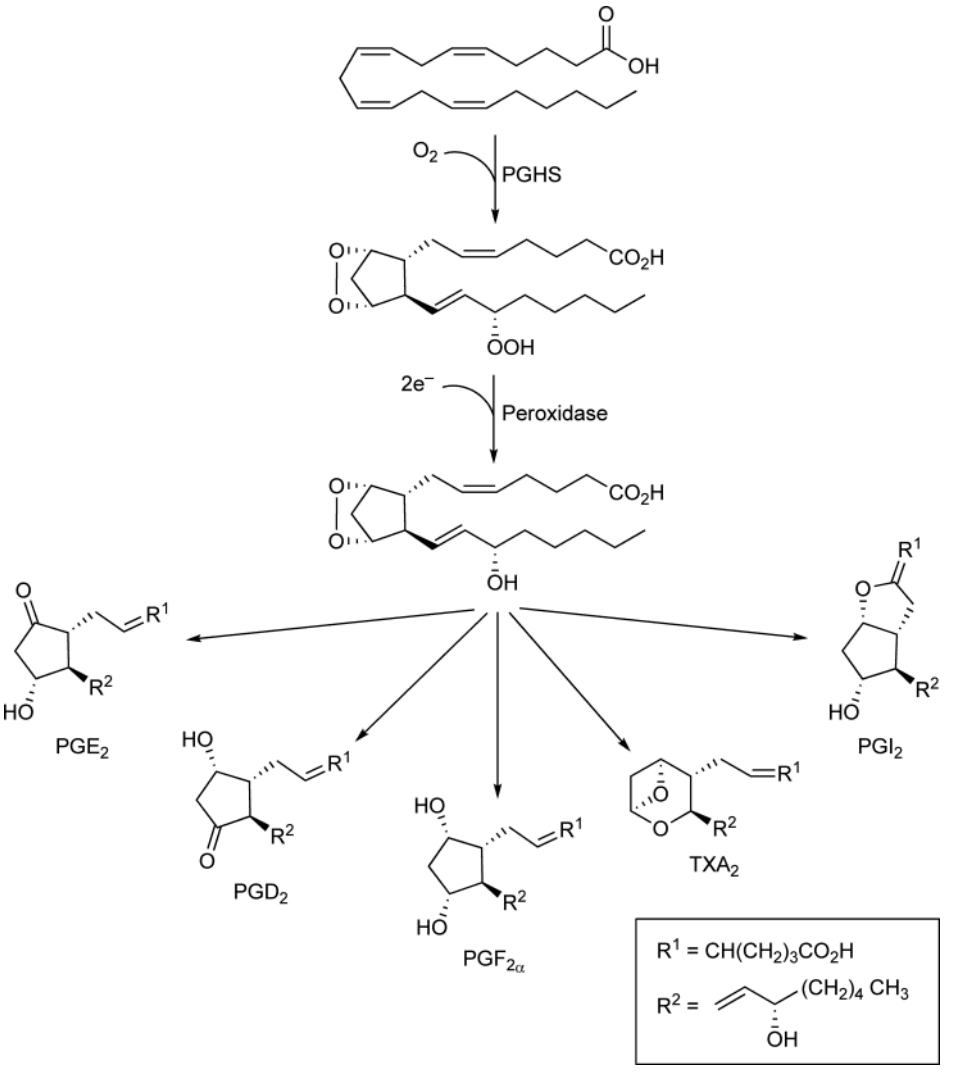
© 2010 Elsevier Ltd. All rights reserved.

1.03.1	**Introduction**	36
1.03.2	**Physiological Functions and Regulation of PGHS Isoforms**	37
1.03.2.1	Tissue Expression and Physiological Roles of PGHS Isoforms	37
1.03.2.2	Regulation of PGHS by Reactive Nitrogen Species	37
1.03.2.2.1	Activation by peroxynitrite	38
1.03.2.2.2	Tyrosine nitration	38
1.03.2.2.3	S-Nitrosation	38
1.03.2.3	Regulation of PGHS Protein Degradation	40
1.03.2.4	Lessons from Targeted Deletion and Exchange of PGHS-1 and PGHS-2	40
1.03.3	**Structural Determinants for Arachidonic Acid Metabolism by PGHS**	41
1.03.3.1	Structure of PGHS	41
1.03.3.2	Partnering between Monomers in PGHS Homodimers	41
1.03.3.3	Molecular and Kinetic Mechanisms of Prostaglandin Biosynthesis	43
1.03.3.4	Molecular Determinants of Substrate Binding and Prostaglandin Biosynthesis	45
1.03.3.4.1	High-affinity substrate binding	46
1.03.3.4.2	Positioning substrate for hydrogen abstraction	46
1.03.3.4.3	Control of stereochemistry in prostaglandin biosynthesis	46
1.03.3.4.4	Control of carbon ring cyclization in prostaglandin biosynthesis	46
1.03.4	**Role of PGHS in the Biosynthesis of Novel Acidic Eicosanoids**	47
1.03.4.1	Oxygenation of Arachidonic Acid by Acetylated PGHS-2	47
1.03.4.2	Interactions of PGHS with ω-3 Fatty Acids	49
1.03.4.2.1	α-Linolenic acid	49
1.03.4.2.2	Bis-dioxygenation of eicosapentaenoic acid	49
1.03.4.2.3	Monodioxygenation of ω-3 fatty acids by PGHS-2 and acetylated PGHS-2	50
1.03.4.3	Oxygenation of Hydroxyeicosatetraenoic Acids	50
1.03.4.3.1	3-Hydroxyeicosatetraenoic acid	50
1.03.4.3.2	5-Hydroxyeicosatetraenoic acid	52
1.03.4.3.3	20-Hydroxyeicosatetraenoic acid	52
1.03.5	**Biosynthesis, Degradation, and Pharmacology of Prostaglandin-Glycerylesters**	53
1.03.5.1	Structural Requirements for Oxygenation of 2-Arachidonoylglycerol by PGHS	53
1.03.5.2	Metabolism of PGH_2-Glycerylester and Its Products	53
1.03.5.3	Synthesis of Prostaglandin-Glycerylesters in Cells and *In Vivo*	54
1.03.5.4	Pharmacologic Actions of Prostaglandin-Glycerylesters	54
1.03.6	**Biosynthesis, Degradation, and Pharmacology of Prostaglandin-Ethanolamides**	55
1.03.6.1	Structural Determinants for Oxygenation of Arachidonoylethanolamide by PGHS	55
1.03.6.2	Metabolism of PGH_2-Ethanolamide and Its Products	55
1.03.6.3	Synthesis of Prostaglandin-Ethanolamides in Cells and *In Vivo*	55
1.03.6.4	Pharmacologic Actions of Prostaglandin-Ethanolamides	56
1.03.7	**Metabolism of Other Arachidonoyl Amides by PGHS**	56
1.03.8	**Conclusions**	56
References		57

1.03.1 Introduction

Arachidonic acid metabolism provides a pathway for the generation of diverse, fast-acting, short-lived signaling molecules. Cytosolic phospholipase A_2 releases arachidonic acid from the phospholipid pool in cellular membranes.[1] Once liberated, multiple oxygenases can act on arachidonate to introduce a single atom of oxygen or one or two molecules of oxygen. Cytochromes P-450 catalyze three monooxygenase reactions with arachidonate: allylic oxidations forming hydroxyeicosatetraenoates (HETEs); ω/ω-1 hydroxylations of the aliphatic chain, also forming HETEs; and olefin epoxidations, yielding epoxyeicosatrienoic acids (EETs).[2] Lipoxygenases (LOXs) remove a bis-allylic hydrogen from arachidonate and control the stereo- and regiochemistry of addition of molecular oxygen, yielding hydroperoxyeicosatetraenoates (HpETEs) that can be reduced to HETEs by peroxidases.[3] Prostaglandin endoperoxide synthase (PGHS, also referred to as cyclooxygenase (COX) or PGG/H synthase) catalyzes the bis-dioxygenation of arachidonate, generating PGH_2, the central intermediate for prostanoids.[4]

By virtue of its key role in prostanoid biosynthesis, PGHS is involved in many physiological and pathophysiological roles. The enzyme incorporates two equivalents of molecular oxygen into arachidonic acid to form the hydroperoxy-endoperoxide prostaglandin G_2 (PGG_2) (**Scheme 1**).[5–7] The hydroperoxide is reduced



Scheme 1 Biosynthesis of prostanoids from arachidonic acid.

by a peroxidase to the corresponding alcohol, PGH_2. *In vivo*, PGH_2 is converted to PGE_2, PGD_2, $PGF_{2\alpha}$, prostacyclin I_2 (PGI_2), and thromboxane A_2 (TXA_2) by a variety of synthases (**Scheme 1**).[8] In the absence of prostanoid synthases, PGH_2 degrades nonenzymatically to form PGE_2 and PGD_2.[5] These downstream signaling molecules bind to G-protein coupled receptors (GPCRs) to mediate their effects, which include maintenance of vascular tone, platelet aggregation, and gastric cytoprotection.[9] PGs are also key mediators of inflammatory responses and hyperalgesia. Nonsteroidal anti-inflammatory drugs (NSAIDs), including acetylsalicylic acid (ASA), indomethacin, and celecoxib, work by blocking fatty acid oxygenation by PGHS.[10–12] Although NSAIDs are very effective anti-inflammatory and analgesic agents, their utility is somewhat limited due to potential gastrointestinal and cardiovascular toxicities.[13,14] This chapter will focus on the structure and function of PGHS isoforms and the synthesis of novel eicosanoids derived from alternative substrates.

1.03.2 Physiological Functions and Regulation of PGHS Isoforms

PGHS is a membrane-bound homodimer[15–17] and a bifunctional protein, possessing fatty acid oxygenase and peroxidase activities.[16,18,19] There are two functional isoforms of PGHS in mammals, PGHS-1 and PGHS-2.[20–22] Alternative splice variants of PGHS genes have been identified, but to date, translation of these variants to catalytically active proteins has not been established in humans.[23] PGHS-1 and PGHS-2 catalyze the oxygenation of arachidonate with similar catalytic efficiencies,[24] and both are activated by peroxides, although much lower concentrations are required to activate PGHS-2 than PGHS-1.[25] The two isoforms share considerable protein sequence and structure similarity[20–22,26] and exhibit similar patterns of subcellular localization to the endoplasmic reticulum and nuclear envelope.[27] However, there are evident differences in the expression, regulation, and functions of PGHS-1 and PGHS-2.

1.03.2.1 Tissue Expression and Physiological Roles of PGHS Isoforms

PGHS-1 is primarily the constitutive isoform of the enzyme. In contrast, PGHS-2 is induced by numerous physiological and pathophysiological stimuli[20–22] and is overexpressed at an early stage in multiple cancers.[28–30] Thus, the paradigm has been that PGHS-1 contributes largely to homeostatic functions, whereas PGHS-2 is a major mediator of pathophysiological functions.[31] It is increasingly evident that this is an oversimplified view. Induction of PGHS is not always deleterious. For example, PGHS-1 is induced during differentiation of some cell and tissue types,[32,33] and inducible expression of PGHS-2 is physiologically important in wound healing,[34] female reproduction,[35] and vascular biology.[36] Furthermore, constitutive expression of PGHS-2 has been observed in the kidney[37] and some epithelial cell types.[38,39] In the central nervous system, PGHS-2 is constitutively expressed at high levels in the cortex, hippocampus, hypothalamus, and spinal cord,[40,41] and PGHS-1 is induced under certain pathophysiological conditions.[42–44] Using PGHS-1 and PGHS-2 knockout mice, recent studies of bacterial lipopolysaccharide (LPS)-induced inflammation in the brain reveal a proinflammatory role for PGHS-1 and an anti-inflammatory role for PGHS-2.[45,46] Although the underlying mechanism for these effects is not well understood, it illustrates the changing landscape of our understanding of the role of PGHS isoforms in health and disease.

1.03.2.2 Regulation of PGHS by Reactive Nitrogen Species

Nitric oxide ($^{\bullet}NO$) mediates a wide range of physiological processes including vascular homeostasis[47] and immune responses,[48,49] primarily via activation of soluble guanylate cyclase.[50] However, under many pathophysiological conditions, cytotoxic responses can also be effected by nitric oxide.[51] Nitric oxide is a diffusible, membrane-permeable, reactive gas generated *in vivo* by nitric oxide synthases (NOSs).[52] Diffusion-limited reaction of superoxide with nitric oxide gives rise to the inorganic peroxide peroxynitrite ($ONOO^-$) (**Scheme 2**), which is protonated to peroxynitrous acid (ONOOH) or couples with carbon dioxide at physiological concentrations to yield nitrosoperoxycarbonate ($ONOOCO_2^-$).[53–56] Either peroxynitrous acid or nitrosoperoxycarbonate can undergo homolysis, yielding hydroxyl radical and nitrogen dioxide ($^{\bullet}NO_2$) or

$$^{\bullet}NO + O_2^{\bullet -} \rightleftharpoons ONOO^-$$
$$ONOO^- + CO_2 \rightleftharpoons ONOOCO_2^-$$
$$ONOOCO_2^- \rightleftharpoons {}^{\bullet}NO_2 + CO_3^{\bullet -}$$
$$ONOOH \rightleftharpoons {}^{\bullet}NO_2 + {}^{\bullet}OH$$

Scheme 2 Reactions of selected reactive nitrogen species.

carbonate radical, respectively.[57–59] Nitric oxide and its products are referred to as reactive nitrogen species; they can covalently modify proteins, lipids, and DNA.[60–64]

Transcriptional regulation of NOS isoforms and PGHS isoforms is quite similar, and the proteins are often coexpressed in cells and *in vivo*. Endothelial NOS (eNOS) and neuronal NOS (nNOS) are constitutive isoforms and are associated with basal expression of PGHS. PGHS-1 and eNOS are both expressed in endothelial cells,[65–67] whereas nNOS is primarily expressed in the brain and spinal cord,[68] where constitutive expression of PGHS-2 is observed.[40,41] The promoter regions of genes encoding PGHS-2 and inducible NOS (iNOS) have many of the same response elements,[69,70] and the proteins are induced concomitantly in response to several inflammatory stimuli.[71,72] There is strong evidence for cross talk between nitric oxide and prostaglandin biosynthetic pathways.[73] However, the effects of nitric oxide on prostaglandin biosynthesis are seemingly paradoxical, stimulating prostaglandin synthesis in some cases and inhibiting it in others. Given the reactivity of nitric oxide, other reactive nitrogen species may mediate the observed effects. Nitric oxide either augments or inhibits PGHS protein expression, but the mechanism of transcriptional and posttranscriptional regulation have not been characterized in detail.[73–75] Reactive nitrogen species can also directly modulate PGHS activity by multiple mechanisms, which, in part, are dependent upon the identity and concentration of reactive nitrogen species.

1.03.2.2.1 Activation by peroxynitrite

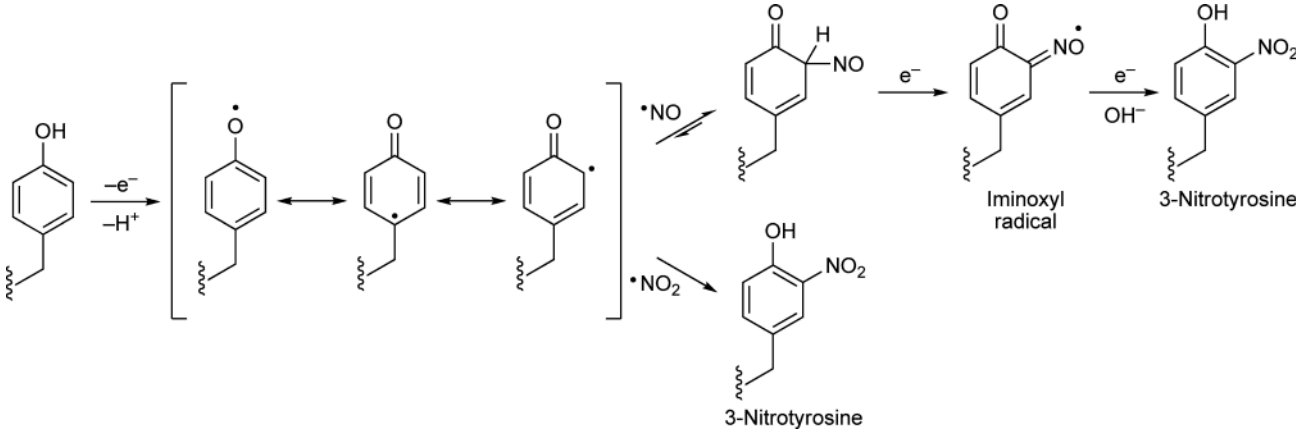
Activation of the peroxidase of PGHS is required for initiation of fatty acid oxygenation, and this can be achieved by numerous hydrophobic organic peroxides.[76,77] *In vitro* peroxynitrous acid serves as an excellent peroxidase substrate for both PGHS-1 and PGHS-2 and stimulates prostaglandin formation by PGHS.[78] Peroxynitrous acid can provide the peroxide tone required to activate PGHS-2 in LPS-treated RAW264.7 immortalized macrophages[78] and vascular smooth muscle cells.[79] Direct activation of PGHS by peroxynitrous acid provides one mechanism by which reactive nitrogen species may stimulate prostaglandin synthesis.

1.03.2.2.2 Tyrosine nitration

Under certain conditions, nitric oxide and peroxynitrous acid have been reported to inactivate PGHS, with inactivation correlating to tyrosine nitration.[80–82] *In vitro*, PGHS is selectively nitrated at Tyr385.[83,84] During the course of catalysis by PGHS, a radical is generated at Tyr385, a step that is absolutely required for prostaglandin synthesis. The tyrosyl radical can be trapped by nitric oxide to generate a tyrosine iminoxyl radical (**Scheme 3**).[85] This short-lived intermediate is further oxidized to yield 3-nitrotyrosine. Alternatively, PGHS may be nitrated in the presence of peroxynitrous acid.[83,86,87] Nitrogen dioxide released from the homolysis of peroxynitrous acid or nitrosoperoxycarbonate (**Scheme 2**) can react with the tyrosyl radical, directly forming 3-nitrotyrosine (**Scheme 3**). Because of the catalytic requirement for the residue, nitration of Tyr385 prevents metabolism of fatty acid substrates by PGHS.

1.03.2.2.3 S-Nitrosation

Protein S-nitrosation is now recognized as an important posttranslational modification of many proteins.[61] S-Nitrosation of PGHS-2 has been observed in cells and in animals, often under conditions where iNOS or nNOS are concomitantly expressed.[88–90] iNOS and nNOS can bind directly to the C-terminus of PGHS-2 but not PGHS-1, and nitric oxide generated by NOS can directly modify PGHS-2.[88,90] It seems that multiple cysteines within PGHS-2 may be S-nitrosated, but modification of Cys540 leads to an increase in PG formation. Both *in vivo* and in cells, inhibition of the associated NOS prevents activation of PGHS-2, even



Scheme 3 Formation of 3-nitrotyrosine from nitric oxide and nitrogen dioxide.

though levels of PGHS-2 protein are unaffected.[88–90] S-Nitrosation of PGHS-2 appears to be an important regulatory modification *in vivo*. S-Nitrosation and activation of PGHS-2 have been implicated in mediating glutamate/*N*-methyl-D-aspartate receptor-dependent neurotoxicity.[88] In a model of myocardial ischemia–reperfusion, the protective effects of atorvastatin involve induction of iNOS and PGHS-2, and activation of PGHS-2 by S-nitrosation.[89] However, the mechanism by which S-nitrosation activates PGHS-2 has not been elucidated.

1.03.2.3 Regulation of PGHS Protein Degradation

Expression of PGHS is tightly controlled by transcriptional regulators and factors that alter mRNA stability.[91,92] PGHS-2 is further regulated at the protein level by proteolysis. PGHS-2 is a substrate for the ubiquitin–proteasome system and exhibits a significantly shorter half-life than PGHS-1 in many cell types.[93,94] PGHS-2 contains a 27-amino acid instability motif at its C-terminus, consisting of residues 586 through 612.[95] The instability motif includes a 19-amino acid insert (residues 594 through 612) containing a conserved asparagine (Asn594) that is found in all PGHS-2 protein sequences but not in PGHS-1. Removal of the instability motif extends the half-life of PGHS-2; conversely, insertion of the motif into PGHS-1 speeds its degradation.[95,96]

Glycosylation of PGHS-2 at Asn594 is necessary but not sufficient for proteasome-dependent degradation.[96] This glycosylation event tags the protein for entry into an endoplasmic reticulum-associated degradation (ERAD) pathway.[95,96] The protein is then ubiquitinated and shuttled to the 26S proteasome. Degradation may be regulated in part by the COP9 signalosome, a multiprotein complex that determines the stability of many regulatory proteins.[97] Timing of PGHS-2 degradation seems to be controlled by a helix–loop–helix configuration in the instability motif, which is proposed to prevent constitutive glycosylation of Asn594 by steric hindrance.[95]

A second pathway of PGHS-2 degradation is dependent upon substrate turnover.[95] Degradation by this mechanism can be inhibited by NSAIDs and by a mutation in PGHS-2 that reduces its catalytic activity. As with the proteasome-dependent pathway, PGHS-1 is resistant to substrate-dependent degradation. The mechanism for substrate-induced degradation is unknown but is not proteasomal or lysosomal.

1.03.2.4 Lessons from Targeted Deletion and Exchange of PGHS-1 and PGHS-2

Targeted deletion of genes is useful for uncovering the physiological functions of proteins of interest and, in the case of PGHS, has yielded unexpected results, unveiling previously unrecognized roles for isoforms. Deletion of PGHS-1 in C57/Bl6 mice results in nearly complete abrogation of prostaglandin synthesis in most tissues.[98] Yet the animals display no apparent phenotype aside from impaired platelet aggregation, which occurs because PGHS-1 is the primary contributor to TXA_2 synthesis. In contrast, deletion of PGHS-2 is detrimental to the organism.[98–100] Many mice die within days of birth, mainly due to patent ductus arteriosus, which is failed closure of the shunt that connects the aorta and pulmonary artery and bypasses the lungs during gestation.[101] Surviving mice exhibit chronic renal failure, myocardial fibrosis, susceptibility to peritonitis, and ulcer formation.[99,100] Female mice are largely infertile. The inflammatory response to LPS is ablated, but responses to many other inflammatory stimuli that induce PGHS-2 are unaffected by deletion of this isoform. In fact, inflammation persists longer in the PGHS-2 knockout than wild-type animals.[99]

To better understand the physiological role of PGHS-2 *in vivo*, additional genetic models have been developed. In one model, PGHS-2 was mutated to encode a Y385F mutation, which imparts wild-type peroxidase activity but eliminates fatty acid oxygenase activity, in effect mimicking chronic PGHS-2-selective inhibition.[102] This mutation fully compensates for the patent ductus arteriosus phenotype. It is proposed that this compensation results from formation of heterodimers between inactive Y385F PGHS-2 and active, wild-type PGHS-1. PGHS-1 monomer would generate prostaglandins, whereas Y385F PGHS-2 would control the spatiotemporal regulation of prostaglandin production. Compared to wild-type mice, PGI_2 and PGE_2 levels are reduced in Y385F PGHS-2 mice. Prostaglandin production in macrophages from Y385F PGHS-2 mice is increased in response to LPS, but not to the extent of wild-type animals. These mice exhibit increased adult mortality, impaired postnatal renal development, and impaired female reproductive capacity associated with PGHS-2 deletion.

In another mouse model (PGHS-1>PGHS-2), PGHS-2 was deleted and the PGHS-1 cDNA inserted into the vacant PGHS-2 locus, placing it under the control of the PGHS-2 promoter.[103] Thus, PGHS-1 displays its normal expression profile but is also constitutively expressed and induced as PGHS-2 would be. Normal closure of the ductus arteriosus occurs in PGHS-1>PGHS-2 animals. Female reproduction is mostly but not completely compensated in this model. PGE_2 and TXA_2 levels match those observed in wild-type animals, but PGI_2 levels are only partially compensated, suggesting that prostacyclin synthase may preferentially couple to PGHS-2 over PGHS-1. Renal pathologies associated with PGHS-2 deletion are observed in old but not young PGHS-1>PGHS-2 mice, and chronic peritonitis and ulceration persist. LPS-stimulated prostaglandin formation by PGHS-1>PGHS-2 macrophages mimics wild-type macrophages at high but not low ($<1\ \mu mol\ l^{-1}$) arachidonate concentrations, possibly owing to differences in peroxide activation requirements between PGHS-1 and PGHS-2. The importance of posttranslational modifications, protein degradation pathways, and metabolism of alternative substrates in mediating PGHS-2 knockout and PGHS-1>PGHS-2 phenotypes remains unclear.

1.03.3 Structural Determinants for Arachidonic Acid Metabolism by PGHS

1.03.3.1 Structure of PGHS

Although PGHS-1 and PGHS-2 drive distinct physiological and pathophysiological processes, the overall architecture and catalytic mechanisms of the two isoforms are very similar.[15,26] Each monomer contains three distinct structural motifs: the epidermal growth factor (EGF)-like domain, the membrane-binding domain, and the globular catalytic domain[15] (**Figure 1**). The EGF-like domain (residues 34 through 72 in ovine PGHS-1) consists of a pair of two-stranded β-sheets with three intradomain disulfide bonds. The membrane-binding domain of PGHS (residues 73 through 116) provides a rather unique mode for anchoring a protein to the membrane. Most membrane-bound proteins possess one or more hydrophobic helices that traverse the membrane. In contrast, PGHS possesses four amphipathic helices (A through D), three of which (helices B through D) lie approximately orthogonal to each other. Owing to the size, orientation, and amphipathic nature of these helices, the membrane-binding domain is only able to associate with one leaflet of the phospholipid bilayer. The remainder of the protein constitutes the α-helical globular catalytic domain, which is structurally homologous to myeloperoxidase. This domain contains both the fatty acid oxygenase and peroxidase active sites. The peroxidase site consists of a solvent-exposed cleft at the top of the catalytic domain where heme binds. The oxygenase active site is buried in the globular catalytic domain. To enter the oxygenase site, ligands must travel through the open core of the membrane-binding domain, an area referred to as the lobby (**Figure 2**). At the top of the lobby, Arg120, Tyr355, and Glu524 form a constriction site. A hydrophobic channel that extends up from the constriction site forms the oxygenase site.

Within the primary shell of the fatty acid oxygenase site, sequence identity is greater than 85% between PGHS-1 and PGHS-2.[26,107] However, the volume of the PGHS-2 oxygenase site is approximately 20% larger than that of PGHS-1 (**Figure 2**). Leu352, Ser353, Tyr355, Phe518, and Val523 form a pocket adjacent to the constriction site of PGHS-2, a pocket that is less accessible in PGHS-1. Residues 523 and 434, which are valine in PGHS-2 and isoleucine in PGHS-1, govern accessibility of this pocket. The greater bulk of isoleucine at position 523 directly hinders access to this pocket. Residue 434 packs against a gatekeeper residue, Phe518. The additional bulk of isoleucine at residue 434 alters the position of Phe518, preventing favorable binding to PGHS-1 of some molecules, including the diarylheterocycle class of PGHS-2-selective inhibitors. Furthermore, opening of the side pocket in PGHS-2 allows access to an additional hydrophilic interaction at Arg513, histidine in PGHS-1. These structural differences do not have a significant impact on binding of arachidonic acid or kinetics of its oxygenation.[24,108]

1.03.3.2 Partnering between Monomers in PGHS Homodimers

The precise role of the dimer structure of PGHS remained elusive for decades, but its importance was alluded to by key findings. Dissociation of PGHS homodimers, even into monomers still possessing secondary structure, results in loss of catalytic activity that cannot be regained under renaturing conditions.[109]

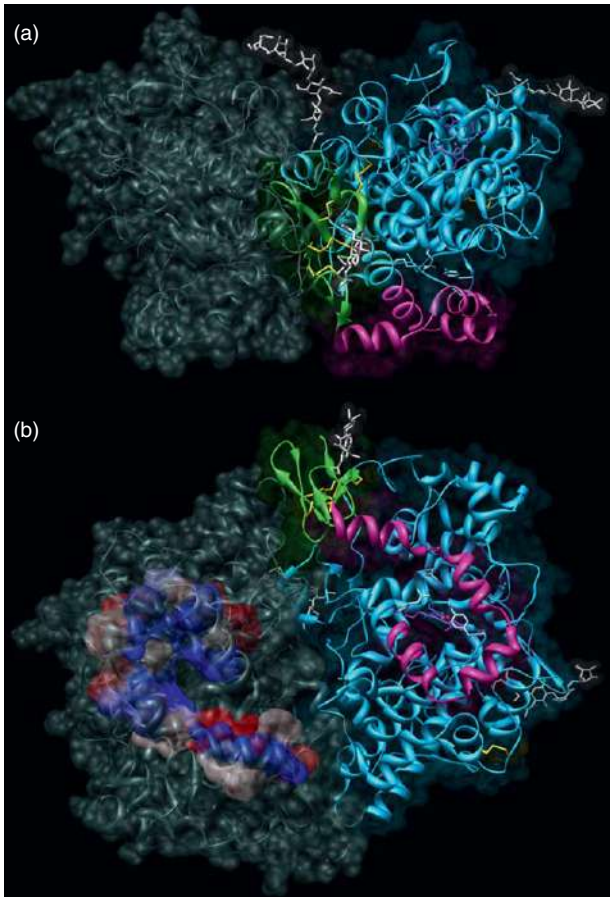


Figure 1 Domain architecture of the PGHS homodimer. (a) Two monomers of PGHS (monomer A colored gray) are related by a C2 axis of symmetry. The three domains are highlighted in monomer B: EGF-like domain (green), membrane-binding domain (magenta), and globular catalytic domain (blue). Each monomer contains four conserved disulfide bonds (yellow) and three conserved glycosylation sites (white). (b) The membrane-binding domain consists of four amphipathic helices that interact with a single leaflet of the phospholipid bilayer. The residues of the membrane-binding domain of monomer A are colored by hydrophobicity (polar residues red, hydrophobic residues blue). This figure was generated from PDB entry 1PRH using UCSF Chimera.[104]

Complete inhibition of PGHS by slow, tight-binding inhibitors, such as flurbiprofen and indomethacin, can be achieved with only one inhibitor molecule bound to each dimer (half-of-sites reactivity), suggesting cross talk between subunits within a dimer.[110,111]

Cooperativity between monomers has recently been investigated using heterodimers composed of a native PGHS-2 subunit and a mutant PGHS-2 subunit.[110,112] G533A is unable to bind arachidonic acid in a catalytically competent conformation, and G533A homodimers exhibit only 5% oxygnase activity in comparison to wild type.[113] However, native–G533A heterodimers are as active as native–native homodimers, suggesting that only one monomer functions at any given time.[110] Furthermore, native–G533A and native–native dimers undergo the same number of turnovers prior to self-inactivation, indicating that inactivation of one monomer prevents catalysis at the second monomer. Partnering between subunits appears to occur between oxygenase sites only. Similar studies have been conducted using G533A and Q203R, a mutation that eliminates peroxidase activity but retains fatty acid oxygenase activity. Cross talk between the peroxidase site of one monomer and the oxygenase site of its partner subunit is not observed.[112] Formation of heterodimers of PGHS-1 and PGHS-2 has been demonstrated *in vitro* and in cells.[102] However, the molecular and kinetic properties of PGHS-1/PGHS-2 heterodimers have not been defined.

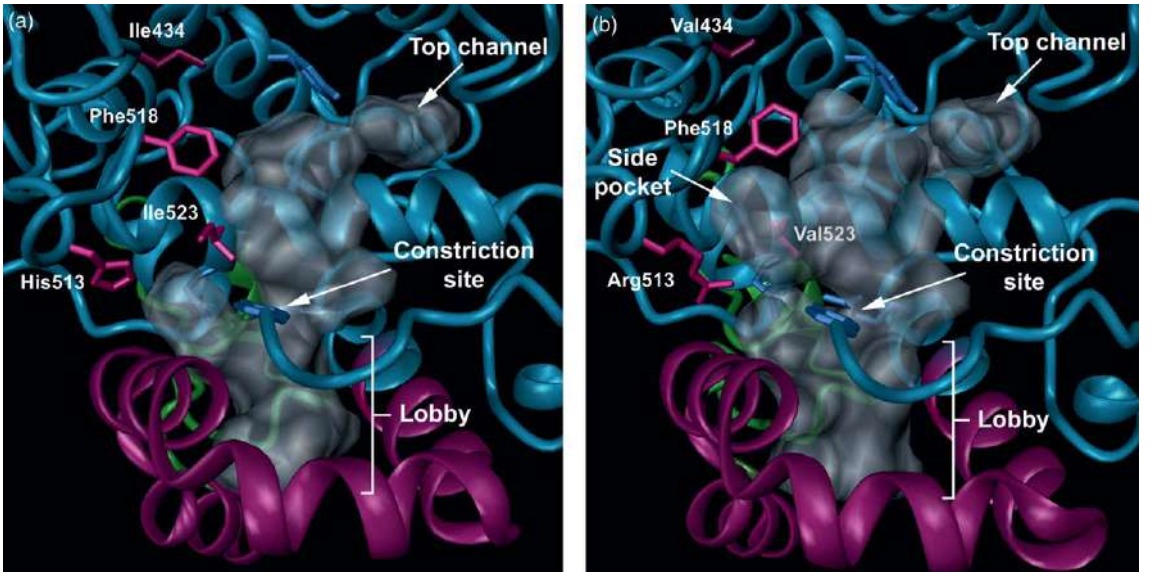


Figure 2 Fatty acid oxygenase sites and lobbies of (a) PGHS-1 and (b) PGHS-2. The solvent-accessible surface area enclosed by the membrane-binding domain, also referred to as the lobby, narrows at the constriction site, consisting of Arg120, Tyr355, and Glu524, before continuing to the fatty acid oxygenase active site. Although the active sites of PGHS isoforms are homologous, two critical differences at residue 434 and 523 open up the side pocket in PGHS-2, allowing access to residue 513. Solvent-accessible surfaces were calculated using MSMS.[105] Molecular graphics images were produced using DINO.[106]

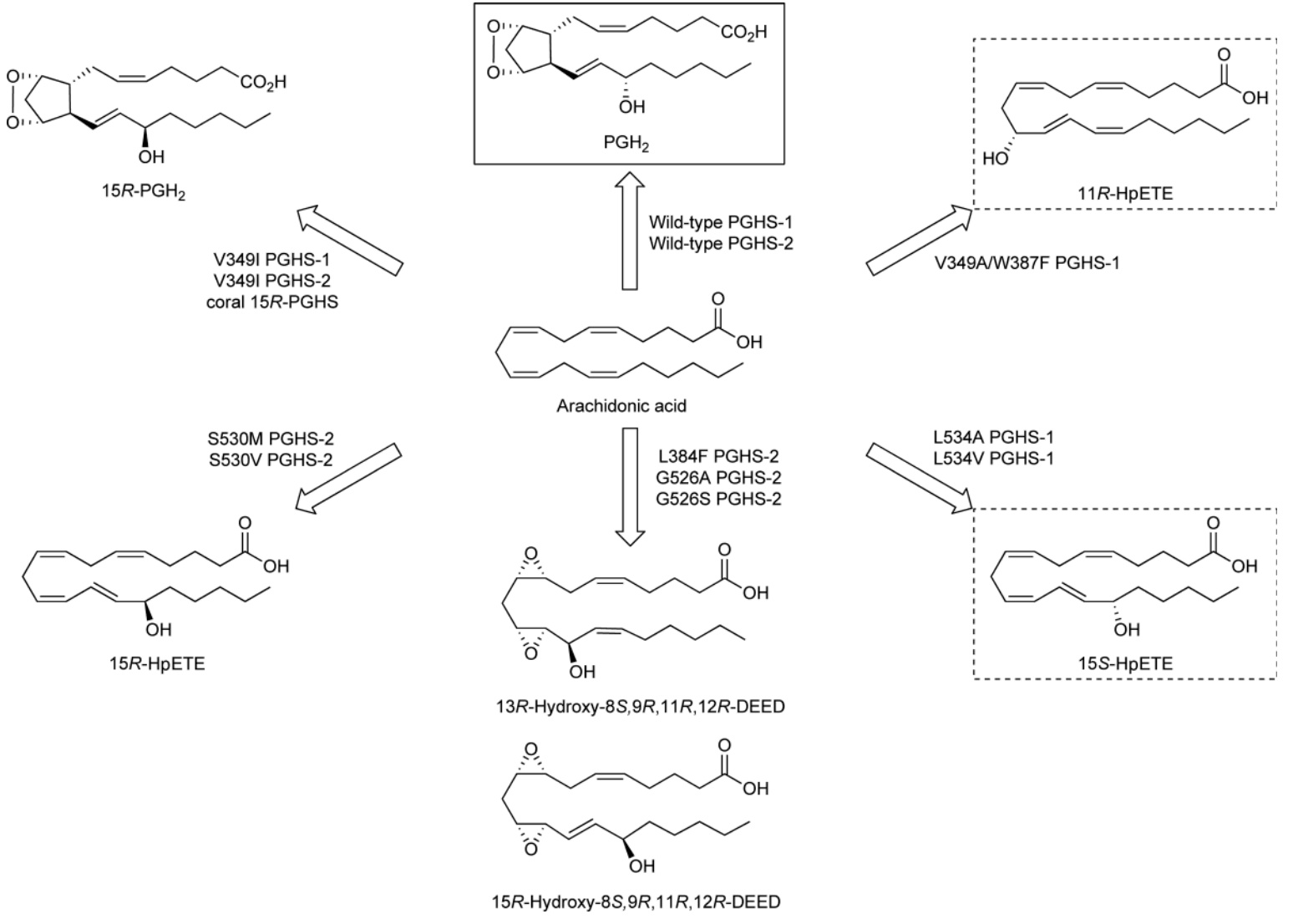
1.03.3.3 Molecular and Kinetic Mechanisms of Prostaglandin Biosynthesis

Even though PGHS transforms an achiral substrate to a product with five chiral centers, the catalytic role of the enzyme in this process is surprisingly simple and closely related to mechanisms of autooxidation of polyunsaturated fatty acids. PGHS serves as a radical initiator for oxygenation of substrates.[114–117] Oxidation of the heme moiety by peroxide gives rise to an oxo–ferryl complex (Fe^{4+}=O) with a protoporphyrin cation radical. Although other tyrosine residues reside within electron transfer distance of the heme, the iron–oxo complex selectively oxidizes Tyr385 to a radical. Selectivity may arise from stabilization of Tyr385$^{\bullet}$ by hydrogen bonding to a neighboring residue, Tyr348.[118]

Tyr385$^{\bullet}$ abstracts the bis-allylic, pro-*S* hydrogen from C-13 of arachidonic acid to generate a resonance-stabilized, carbon-centered radical.[116,117] Trapping of the radical at C-11 by molecular oxygen yields a peroxyl radical that cyclizes to C-9 to yield the 9,11-endoperoxide and a carbon-centered radical at C-8, which adds to C-12 to generate the cyclopentane ring and an allylic carbon radical. This intermediate is susceptible to addition of molecular oxygen at either C-13 or C-15, although the major product arises from coupling at C-15.[119] The peroxyl radical formed at this stage may then abstract a hydrogen atom from Tyr385, giving rise to PGG_2 and propagating the tyrosyl radical, which can initiate the cycle again.

Although PGG_2 is the major product of arachidonic acid oxidation, products arising from a single dioxygenation reaction are also observed. Once the first equivalent of molecular oxygen is trapped, the 11-peroxyl radical may abstract hydrogen from the protein before cyclization can occur, generating 11(*R*)-hydroperoxyeicosatetraenoic acid (11*R*-HpETE)[119] (**Scheme 4**). It is also possible for oxygen to insert at C-15 instead of at C-11, creating a peroxyl radical intermediate that is unable to cyclize, resulting in the formation of 15*S*-HpETE (**Scheme 4**). Typically, HpETEs are relatively minor products of the oxygenase reaction.

Detailed steady-state kinetics of PGHS-1 studied as a function of substrate concentrations, pH, and solvent viscosity, as well as by oxygen kinetic isotope effects are consistent with a sequential kinetic mechanism of catalysis.[120] This kinetic mechanism requires formation of a ternary complex of enzyme, fatty acid, and molecular oxygen as an intermediate prior to the first kinetically irreversible step. For bis-dioxygenation of arachidonate, the first irreversible step has been proposed to be cyclization of the five-membered endoperoxide



Scheme 4 Products of arachidonate oxygenation by wild-type and mutant PGHS. The main product of oxygenation by wild-type PGHS is PGH_2 (solid box), although minor products are 11*R*-HpETE and 15*S*-HpETE (dashed boxes). Mutations in the oxygenase site can alter product profiles as shown.

ring or of the bicylic endoperoxide. Abstraction of the bis-allylic hydrogen appears to be reversible, and at least part of the specificity of the oxygenase reaction arises from competition between trapping of molecular oxygen and reversible hydrogen transfer. If hydrogen abstraction were irreversible, then competing pathways would favor formation of HpETEs.

1.03.3.4 Molecular Determinants of Substrate Binding and Prostaglandin Biosynthesis

The regio- and stereochemistry of the oxygenase reaction is strictly controlled by the conformation of arachidonic acid in the active site.[121,122] The X-ray structure of arachidonic acid bound to ovine PGHS-1 reveals that the substrate binds in an inverted L-shaped orientation[122] (**Figure 3**). The carboxylic acid of arachidonate appears to form an ionic bond with Arg120 and a hydrogen bond with Tyr355. The aliphatic chain extends up through the hydrophobic channel and kinks at the 9,10-bond, placing the pro-*S* hydrogen atom of C-13 by the catalytic residue, Tyr385, and projecting C-12 through C-20 into the top channel. The conformation of the aliphatic chain is strictly controlled by an estimated 47 hydrophobic interactions (**Figure 3**). The nature and importance of protein–ligand interactions in the formation of PGG_2 have been explored with numerous site-directed mutants of PGHS.

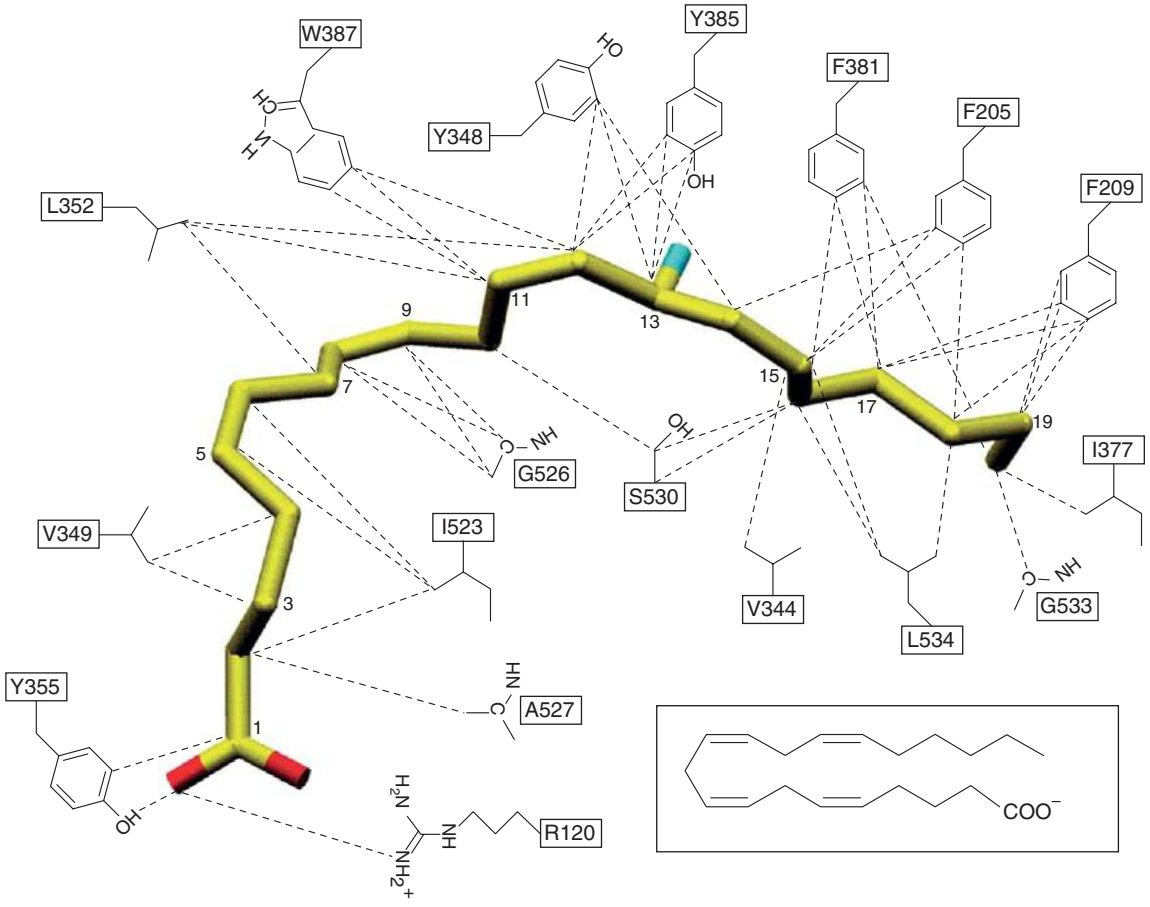


Figure 3 A schematic of interactions between arachidonate and PGHS channel residues from the cocrystal structure of arachidonate bound to PGHS-1. Carbon atoms of arachidonate are yellow, oxygen atoms red, and the 13-pro-*S* hydrogen blue. All dashed lines represent interactions within 4.0 Å between the specific side chain atom of the protein and arachidonate. (Inset) A schematic of the chemical structure of arachidonate. Reproduced from M. G. Malkowski; S. L. Ginell; W. L. Smith; R. M. Garavito, *Science* **2000**, *289*, 1933–1937. Copyright 2000. Reprinted with permission from AAAS.

1.03.3.4.1 High-affinity substrate binding

Interactions of PGHS with the carboxylic acid of arachidonate are key determinants mediating substrate affinity. Tyr355 is positioned to hydrogen bond with the carboxylic acid of arachidonate and to form a hydrophobic contact with C-1 of arachidonate.[122] Tyr355 helps optimize substrate affinity and conformation; loss of this interaction results in modest increases in K_m.[123] Arg120 appears to form an ionic bond with the substrate's carboxylic acid.[122] Mutation of Arg120 to Gln in PGHS-1 almost completely abolishes the oxygenase activity.[123] In fact, the loss of the ionic interaction reduces the affinity of substrate binding more than three orders of magnitude for PGHS-1. However, a very different effect is observed with PGHS-2.[124] R120Q PGHS-2 exhibits wild-type activity, with no significant differences in K_M or v_{max}, suggesting that a hydrogen bond may substitute for an ionic bond in PGHS-2.

1.03.3.4.2 Positioning substrate for hydrogen abstraction

Positioning of C-13 of arachidonic acid with respect to Tyr385 is absolutely critical to formation of PGG_2. The tyrosyl radical formed by oxidation of Tyr385 serves as the free radical initiator of the oxygenase reaction, and mutation of this residue to phenyalanine results in complete loss of oxygenase activity.[125] Mutations that alter the position of the substrate in the top channel or that sterically hinder access of the substrate to the top channel prevent alignment of the bis-allylic hydrogen with Tyr385, and catalysis cannot occur. In its productive conformation, the ω-carbon of arachidonate is positioned above Gly533.[122] Mutation of Gly533 to valine shortens the top channel and blocks turnover of arachidonic acid.[110] However, G533V PGHS-2 is able to oxygenate the shorter C_{18} substrate linolenic acid.

1.03.3.4.3 Control of stereochemistry in prostaglandin biosynthesis

Stereochemistry of substrate oxygenation is strictly controlled by PGHS. Study of coral enzymes provided insight into the key determinants of this specificity. Mammals synthesize exclusively 15*S*-prostaglandins. In contrast, the coral *Plexaura homomalla* may generate either 15*S*- or 15*R*-PGs (**Scheme 4**), depending on the location from which the coral is isolated.[126,127] Genes encoding PGHS from *P. homomalla* harvested from different regions reveal that the proteins are 97% identical, and within the oxygenase site, only residue 349 differs.[128] This residue is isoleucine in *P. homomalla* that produces 15*R*-prostaglandins but is valine in *P. homomalla* that produces 15*S*-prostaglandins; valine is present in mammalian PGHS isozymes. Mutation of residue 349 in *P. homomalla* PGHS isozymes alters the stereochemical specificity by approximately 30%. Similarly, mutation of Val349 to isoleucine in human PGHS-1, human PGHS-2, and murine PGHS-2 results in production of 40–65% 15*R*-prostaglandins.[129]

Although not naturally occurring, an even more dramatic effect on stereochemical specificity is observed upon site-directed mutagenesis of Ser530, which resides across from Val349 in the active site of PGHS.[129] S530T PGHS-2 generates 15*R*-prostaglandins almost exclusively, and S530M and S530V PGHS-2 produce more than 80% 15*R*-prostaglandins. Because both larger polar and nonpolar residues cause this shift in product profile, it is most likely that Ser530 controls stereochemistry by steric effects.

1.03.3.4.4 Control of carbon ring cyclization in prostaglandin biosynthesis

Many of the hydrophobic interactions between PGHS and arachidonate are critical for determining the specificity of the oxygenase reaction. The first step of the oxygenase mechanism involves formation of a resonance-stabilized, carbon-centered radical. However, disruption of van der Waals interactions often produces changes in the ratio of oxygenation products.

Mutation of Ser530 can affect the cyclization of the peroxyl radical. Substitution of alanine for Ser530 has little effect on stereochemistry or cyclization in either PGHS-1 or PGHS-2.[129–131] Mutation of Ser530 to threonine, valine, methionine, leucine, or isoleucine in PGHS-1 significantly attenuates oxygenase activity, suggesting that the additional bulk blocks the active site and prevents access of the substrate to the catalytic tyrosyl radical.[129,131,132] The addition of bulk to position 530 in PGHS-2 has varying effects. Mutation to asparagine has no effect on metabolism of arachidonic acid or the ratio of products.[133] Mutation to glutamine, leucine, or isoleucine abrogates oxygenase activity, as observed with PGHS-1.[129,133] However, mutation to valine or methionine alters the product profile for PGHS-2, such that few prostaglandins are made and the major product is 15*R*-HETE (**Scheme 4**).[129,133,134]

In addition to its role in controlling stereochemistry, Val349 is also a critical determinant of cyclization of the 11-peroxyl radical to PGG_2.[132] Mutation of Val349 to alanine does not affect the stereochemical specificity of the reaction but shifts the reaction toward 11*R*-HETE (**Scheme 4**). A similar effect is observed with W387F. The double mutant V349A/W387F PGHS-1 generates over 80% 11*R*-HETE.[135] The crystal structure of arachidonic acid bound to the cobalt (III)-protoporphyrin-IX derivative of V349A/W387F ovine PGHS-1 provides a structural basis for the altered product profile, demonstrating the importance of hydrophobic contacts with the substrate. Although electron density was not observed for C-13 through C-20, modeling studies indicate that the pro-*S* hydrogen of C-13 would be positioned within 3 Å of Tyr385, a distance allowing for hydrogen abstraction. However, positioning of the first 12 carbons of arachidonate reveals a striking difference in the conformation of arachidonic acid bound to wild-type PGHS-1 and the double mutant. From the crystal structure of arachidonic acid bound to wild-type ovine PGHS-1, Val349 interacts with C-2 and C-3 of the substrate.[122] When alanine is substituted at this position, C-3 through C-6 of the aliphatic chain of arachidonic acid shift and rotate into the opening that is created.[135] This rotation induces large shifts in more rigid sections of arachidonic acid. Steric interactions between Trp387 and C-11 and C-12 of arachidonic acid likely direct the formation of the endoperoxide bridge. Mutation of Trp387 to phenylalanine alters the steric interactions and changes the rotational freedom of neighboring bonds, generating a conformation for the 11-peroxyl radical that makes attack of C-9 unfavorable.

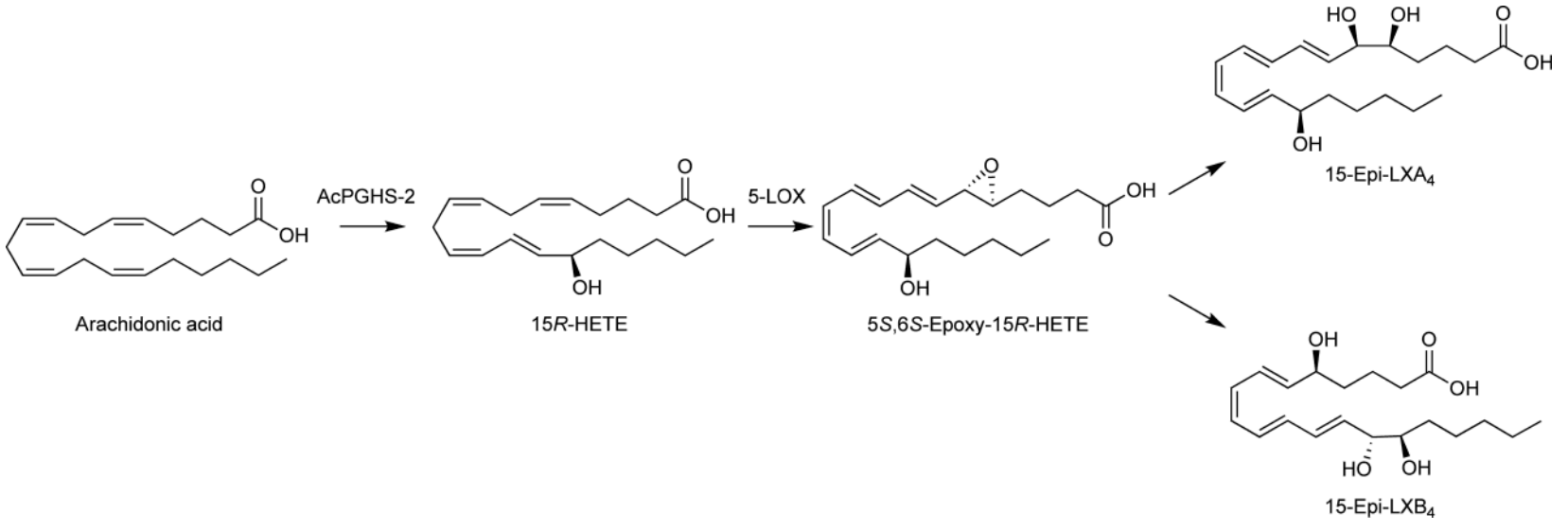
Although there is no crystal structure available with arachidonic acid bound to PGHS-2 in a productive conformation, a structure of PGH_2 bound to apo-PGHS-2 provides additional insight into requirements for PGG_2 formation.[136] The active site closely resembles that of PGHS-1 with arachidonic acid bound, but two side-chain rotamers are observed for Leu384 and Tyr385. In this structure, Tyr385 flips up into the vacant heme-binding site. Modeling indicates that shifting the carboxylate of PGH_2 to form an ionic bond with Arg120 could restore the conformation of Tyr385, and this translation can readily be accommodated by the active site without additional movement of protein side chains. To accommodate the endoperoxide moiety, the side chain of Leu384 rotates approximately 60°. Overlaying this structure with the rotamer observed in the structure of PGHS-1 with arachidonic acid bound, it is evident that, without rotation of Leu384, there would be a steric clash with the endoperoxide bridge. Although Leu384 has no direct contact with arachidonic acid, mutation of this residue attenuates PGG_2 formation.[132,137] L384W and L384F PGHS-2 generate a higher proportion of 15- and 11-HETE than wild-type PGHS-2.[137] Furthermore, these mutants produce diepoxyalcohols that arise from alternate reactions of the C-8 radical on the 9,11-endoperoxide (**Scheme 4**). Mutation of a neighboring residue, Gly526, to alanine, serine, threonine, or valine has a similar effect, producing higher levels of the diepoxyalcohols.

1.03.4 Role of PGHS in the Biosynthesis of Novel Acidic Eicosanoids

1.03.4.1 Oxygenation of Arachidonic Acid by Acetylated PGHS-2

ASA irreversibly inhibits prostaglandin biosynthesis by acetylation of PGHS.[10,138] The acetyl group of ASA is transferred to the active site residue Ser530.[139–142] Although Ser530 is not required for catalysis, covalent modification of this residue completely blocks arachidonic acid metabolism by PGHS-1[143] by hindering substrate access to Tyr385.[144] Unlike PGHS-1, ASA-acetylated PGHS-2 is still able to metabolize arachidonic acid, but it produces 15-HETE instead of prostaglandins.[143] Unmodified PGHS-2 generates small amounts of 15*S*-HpETE in the course of catalysis, but acetylated PGHS-2 produces 15*R*-HpETE with no detectable *S*-isomer, as is observed with S530M PGHS-2.[133,134,145,146] The extra space of the side pocket allows PGHS-2 to bind arachidonic acid in a productive conformation for subsequent conversion to 15*R*-HpETE.[147] V434I/R513H/V523I PGHS-2 is a triple mutant that replaces the conserved substitutions in this pocket with the corresponding PGHS-1 consensus residues. This triple mutant exhibits wild-type activity in the absence of inhibitor, but following acetylation by ASA, very little oxidation of arachidonic acid is observed.

The conversion of arachidonic acid to 15*R*-HpETE by acetylated PGHS-2 may have important physiological impact. Using 15-HpETE or 15-HETE as a substrate, 5-lipoxygenase can produce 5*S*,6*S*-epoxy-15-HETE. This intermediate is metabolized by epoxide hydrolases to yield 5*S*,6*R*,15-trihydroxyeicosatetraenoic acid (lipoxin A_4, LXA_4) or 5*S*,14*R*,15-trihydroxyeicosatetraenoic acid (lipoxin B_4, LXB_4) (**Scheme 5**).



Scheme 5 Biosynthesis of aspirin-triggered lipoxins. AcPGHS-2, ASA-acetylated PGHS-2.

Stereochemistry at the 15-position is simply determined by the configuration of 15-H(p)ETE. 15*S*-H(p)ETE is the major product of 15-lipoxygenase, whereas 15*R*-H(p)ETE is generated by either ASA-acetylated PGHS-2 or cytochromes P-450. If the configuration is 15*R*, the lipoxins are referred to as 15-epi-LXA_4 and 15-epi-LXB_4. ASA triggers biosynthesis of four 15*R*-epimers of lipoxins in a mixed culture of human endothelial cells and polymorphonuclear leukocytes.[148] In a randomized human trial of healthy volunteers, plasma levels of 15-epi-LXA_4 are increased somewhat following daily dosing with low-dose ASA (81 mg), revealing that 15-epi-LXA_4 is produced at clinically relevant doses.[149]

15-Epi-LXA_4 and stable analogs of this product have potent anti-inflammatory properties. Lipoxins halt neutrophil recruitment, adhesion, and infiltration;[150–152] reduce expression of genes upregulated by inflammatory responses;[153–156] stimulate prostacyclin[157] and nitric oxide generation;[158] and reduce angiogenesis.[159–161] Many of the anti-inflammatory effects of LXA_4 and 15-epi-LXA_4 are mediated by expression of suppressor of cytokine signaling-2 (SOCS-2),[162] which is activated by binding of these ligands to ALX, a GPCR,[155] and the aryl hydrocarbon receptor, a nuclear receptor.[163] In addition to these direct actions, there is evidence for cross talk with the leukotriene B_4 receptor[164] and growth factor receptors.[159,165] 15-Epi-LXA_4 may account for some of the anti-inflammatory effects and beneficial cardiovascular outcomes associated with low-dose ASA treatments.

1.03.4.2 Interactions of PGHS with ω-3 Fatty Acids

The high ratio of ω-6:ω-3 fatty acid in Western diets, now approaching 20:1, has been correlated to increased incidence of cardiovascular and inflammatory diseases.[166] ω-6 and ω-3 fatty acids differ in the position of the double bonds with respect to the terminal methyl group. Counting from the methyl end of the carbon chain, the first double bond of ω-6 fatty acids, such as arachidonic acid, occurs between carbons 6 and 7; the first double bond of ω-3 fatty acids, such as eicosapentaenoic acid (EPA), occurs between carbons 3 and 4. ω-3 Fatty acids exhibit potent immunomodulatory effects.[167,168] In humans, ω-3 supplementation suppresses expression of cytokines and clotting factors and alters the profile of eicosanoids produced. Specifically, PGE_2, TXA_2, and LTB_4 levels are reduced while levels of PGI_2 and the EPA metabolites PGI_3 and LTB_5 are increased. The molecular mechanisms for the biological effects of ω-3 fatty acids continue to be investigated. Some effects may be mediated through the direct interaction of ω-3 fatty acids with PGHS and by products of their oxidation by PGHS.

1.03.4.2.1 α-Linolenic acid

Although PGHS exhibits a marked preference for 20-carbon substrates, it is able to metabolize 18-carbon substrates including γ-linolenic acid (18:3, ω-6) and linoleic acid (18:2), albeit less efficiently than arachidonic acid. α-Linolenic acid (ALA) is an 18-carbon ω-3 fatty acid that is a very poor substrate for PGHS-1, displaying a k_{cat}/K_m that is 500-fold lower than that for arachidonic acid.[169] PGHS-2 utilizes this substrate more efficiently with only a 10-fold reduction in k_{cat}/K_m compared to arachidonic acid. PGHS-2 acts as an LOX with ALA, incorporating a single equivalent of molecular oxygen to generate 12-hydroxy-9,13,15-octadecatrienoic acid. ALA can inhibit arachidonic acid oxidation by microsomal PGHS-1 and purified PGHS-2[170] and weakly inhibits PGE_2 synthesis in cells.[171]

1.03.4.2.2 Bis-dioxygenation of eicosapentaenoic acid

EPA is structurally similar to arachidonic acid, containing an additional double bond between carbons 17 and 18. Despite the structural similarity, EPA is not oxygenated by PGHS-1 unless the protein is activated by lipid hydroperoxides.[172] Even in the presence of 15-HpETE, oxidation of EPA is only about 10% of that observed with arachidonate. However, peroxide-dependent metabolism of EPA by PGHS-1 does occur in cells.[173,174] EPA inhibits arachidonate oxygenation by PGHS-1 *in vitro* with half-maximal inhibition achieved at equimolar EPA/AA concentrations.[172,174] The cocrystal structure of EPA complexed with Co(III)-PPIX-reconstituted PGHS-1 reveals a molecular basis for reduced efficiency of EPA oxidation.[175] The fifth double bond of EPA reduces the flexibility of the carbon chain, as compared to arachidonic acid. This results in misalignment of C-13 with respect to Tyr385, and modeling studies place the pro-*S* hydrogen 4.1 Å from the phenolic oxygen, a

distance calculated to be 2.7 Å for arachidonate. Increasing the volume of the top channel by substituting leucine for Phe205 affords a moderate increase in EPA oxidation.

PGHS-2 is much better than PGHS-1 at metabolizing EPA.[169,172] *In vitro*, the k_{cat}/K_M of EPA is only twofold lower than that of arachidonate, but PGHS-2 exhibits a marked preference for arachidonate over EPA when the substrates are added concomitantly.[172] Even at a fivefold molar excess, EPA only modestly inhibits arachidonate oxygenation, whereas at equimolar concentrations, arachidonate significantly attenuates EPA metabolism. PGH_3 (**Scheme 6**) is the major product of EPA oxidation by PGHS and can be converted by prostaglandin synthases to 3-series prostaglandins. Both hematopoietic- and lipocalin-prostaglandin D synthases (H-PGDS and L-PGDS, respectively) and microsomal prostaglandin E synthase-1 (mPGES-1) metabolize PGH_3 but not as readily as PGH_2. Data are not available for other PGE synthases or the numerous prostaglandin F synthases. Prostacyclin synthase and thromboxane synthase exhibit similar activity for PGH_3 and PGH_2.

In rodents and humans, ingestion of ω-3 fatty acids reduces production of PGE_2, TXA_2, and LTB_4.[176] Concomitantly TXA_3 and LTB_5 increase;[176] these EPA-derived products are weaker agonists of their respective receptors than their arachidonate-derived counterparts.[177–179] PGI_3 levels also increase but without a concomitant decrease in PGI_2.[180,181] Both PGI_3 and PGI_2 are vasodilators and inhibitors of platelet aggregation.[174] The alteration of prostanoid profile may in part explain the cardiovascular and anti-inflammatory benefits of a diet high in ω-3 fatty acids.

In cell culture and in animal models, many cancer lines exhibit reduced proliferation following treatment with ω-3 fatty acids,[182] and in some cases, increased production of PGE_3 has been noted.[183–185] In many cases, the antiproliferative effect is abrogated by treatment with PGHS inhibitors. Treatment of A549, a lung cancer cell line, and some pancreatic cancer cell lines with PGE_3 achieved similar effects as treatment with EPA.[183,184] Although the mechanisms by which ω-3 fatty acids act on cancer cell lines are likely varied, this provides strong evidence for the physiological importance of EPA-derived prostaglandins.

1.03.4.2.3 Monodioxygenation of ω-3 fatty acids by PGHS-2 and acetylated PGHS-2

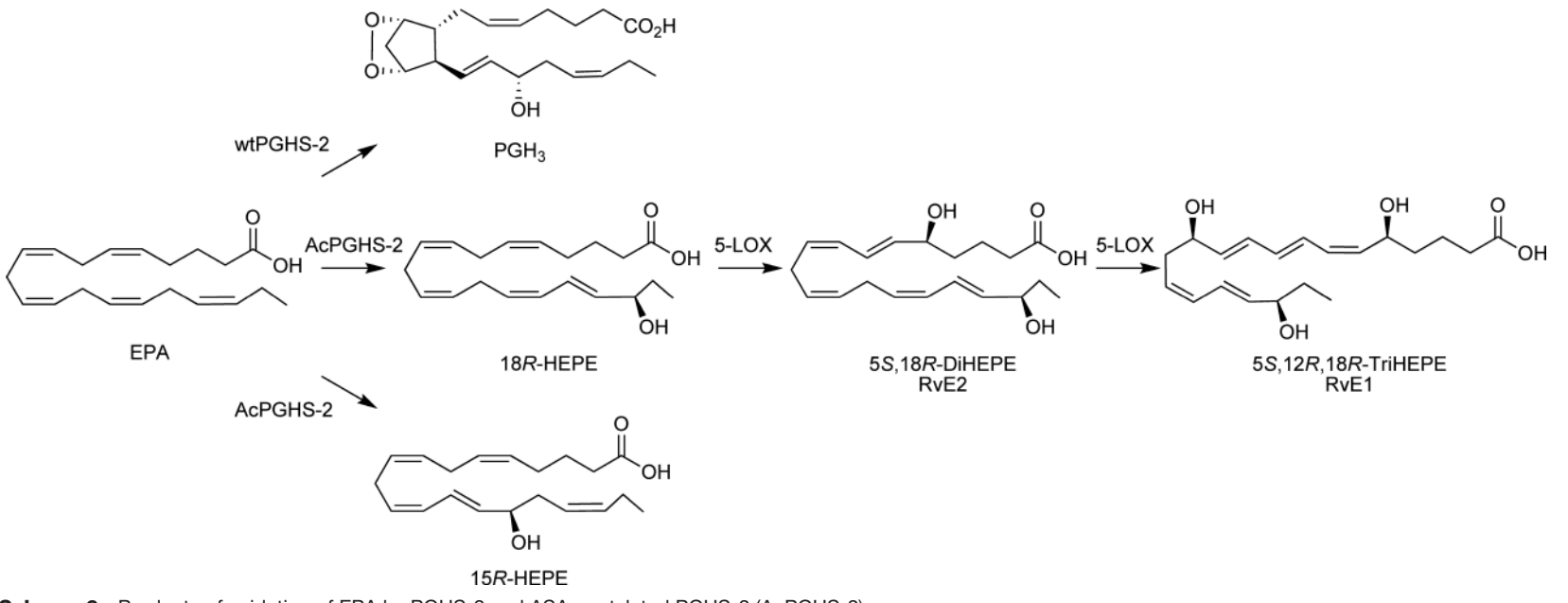
PGHS-2 also catalyzes the monodioxygenation of EPA and docosahexaenoic acid (DHA). Although PGH_3 is the major product of EPA oxidation by PGHS-2, the enzyme also produces small quantities of hydroxylated EPA, predominantly 11*R*-hydroxyeicosapentaenoic acid (11*R*-HEPE). Acetylation of PGHS-2 inhibits PGH_3 synthesis and shifts the monooxygenated products to 18*R*-HEPE and 15*R*-HEPE.[186] Unlike EPA, DHA is not converted to prostaglandin-like products by PGHS-2.[187] Rather the major product of oxidation is 13-hydroxydocosahexaenoic acid (13-HDHA), which is shifted to 17*R*-HDHA following treatment of PGHS-2 with ASA.[188] The action of LOXs on 15*R*-HEPE, 18*R*-HEPE, and 17*R*-HDHA can produce di- and tri-hydroxy derivatives, referred to as resolvins, the E-series from EPA and the D-series from DHA.[186,188–191] As with lipoxins, resolvins promote the resolution phase of inflammation, reducing neutrophil infiltration, and regulating chemokine and cytokine production.[192] Resolvin E1 binds to the leukotriene B_4 receptor, attenuating LTB_4 proinflammatory signaling, and to an orphan receptor ChemR23, reducing the nuclear factor-κB (NF-κB) response.[193] To date, receptors for the D-series resolvins have not been identified.

1.03.4.3 Oxygenation of Hydroxyeicosatetraenoic Acids

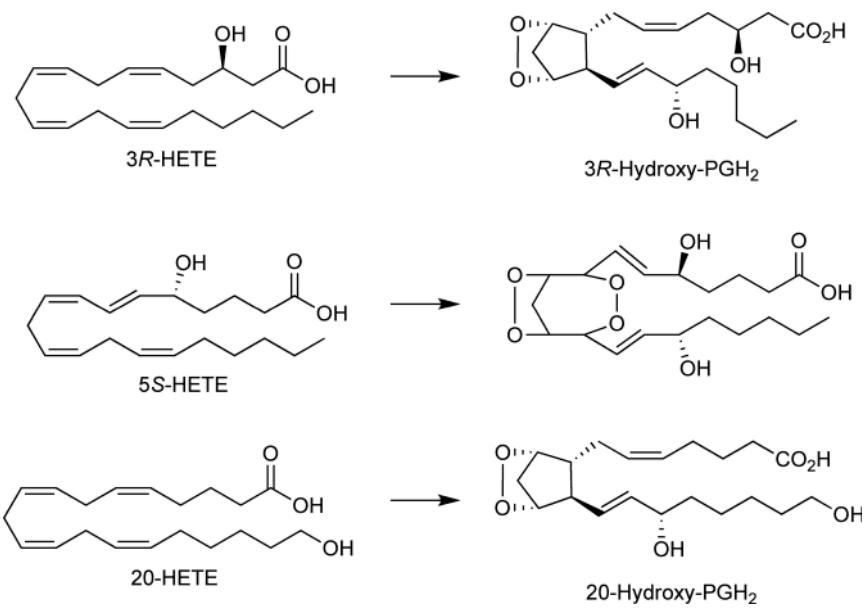
HETEs are monooxygenated products of arachidonic acid generated by LOXs and cytochromes P-450, and in small part, PGHS.[2] Owing to rearrangement of the double bonds, some major HETEs, namely 11-, 12-, and 15-HETE, do not possess a bis-allylic hydrogen at C-13 and likely do not serve as substrates for PGHS. Other HETE isomers retain the double bond configuration neighboring C-13, which allows them to be oxygenated by PGHS, giving rise to novel classes of eicosanoids.

1.03.4.3.1 3-Hydroxyeicosatetraenoic acid

Candida albicans and related fungi generate 3*R*-HETE from arachidonic acid via a β-oxidation-related pathway that has not been fully elucidated.[194–196] As with other HETE isomers, 3*R*-HETE activates TRPV1 receptors.[197] PGHS-1 and PGHS-2 oxygenate 3*R*-HETE with comparable efficiencies, albeit 10-fold less efficiently than arachidonic acid.[198] The major products of PGHS-2 oxidation of 3*R*-HETE are derived from nonenzymatic degradation of 3-hydroxy-PGH_2, although 3,11- and 3,15-diHETE are also observed (**Scheme 7**).



Scheme 6 Products of oxidation of EPA by PGHS-2 and ASA-acetylated PGHS-2 (AcPGHS-2).



Scheme 7 Oxygenation of HETEs by PGHS.

Acetylation of PGHS-2 by ASA favors conversion of 3*R*-HETE to 3,15-diHETE, a product that is a potent activator of maxi-K ion channels in human trabecular meshwork cells. HeLa cells infected with *C. albicans* produce 3-hydroxy-PGE_2 following treatment with arachidonic acid. 3-Hydroxy-PGE_2 induces interleukin (IL)-6 production in A549 cells via the EP_3 receptor and stimulates synthesis of cAMP in Jurkat cells through action at the EP_4 receptor.

1.03.4.3.2 5-Hydroxyeicosatetraenoic acid

5*S*-HETE is generated by the monodioxygenation of arachidonic acid by 5-lipoxygenase. PGHS-1 is unable to further oxygenate 5*S*- or 5*R*-HETE.[199] In contrast, 5*S*-HETE, but not 5*R*-HETE is efficiently metabolized by PGHS-2. The major product of 5*S*-HETE oxygenation by purified PGHS-2 contains five additional oxygen atoms. This requires incorporation of three equivalents of molecular oxygen followed by removal of one atom of oxygen, presumably by the peroxidase activity of PGHS. The product was identified as a dihydroxy bicyclic diendoperoxide. The absolute configuration is predicted to exist as 5*S*,15*S*-dihydroxy-9*S*,11*R*-8*S*,12*S*-diendoperoxy-6*E*,13*E*-eicosadienoic acid (**Scheme 7**). PGHS-2 and 5-lipoxygenase are coexpressed and activated under similar conditions, most notably in activated macrophages where both enzymes localize to the nuclear envelope.[27,200,201] Although formation of this unusual bicyclic diendoperoxide has not been confirmed in cells, convergence of 5-lipoxygenase and PGHS-2 pathways may give rise to previously unidentified eicosanoids.

1.03.4.3.3 20-Hydroxyeicosatetraenoic acid

20-HETE, generated by the ω-hydroxylation of arachidonic acid by cytochromes P-450 CYP4F2 and CYP4A1,[202,203] is a critical modulator of renal function and blood pressure.[204] Depending upon the tissue examined, 20-HETE can act as a potent vasoconstrictor, vasodilator, or bronchodilator, but the activity can be blocked by PGHS inhibitors.[205–208] PGHS converts 20-HETE to 20-hydroxy-PGG_2 and 20-hydroxy-PGH_2 (**Scheme 7**),[206] and the downstream 20-hydroxy-prostanoids seem to mediate the biological activity of 20-HETE, at least in part.[206–208] The relative efficiencies of 20-HETE metabolism by PGHS isoforms have not been established. Neither the conversion of 20-hydroxy-PGH_2 by downstream synthases nor the receptors mediating the biological properties of 20-hydroxy-prostanoids have been examined.

1.03.5 Biosynthesis, Degradation, and Pharmacology of Prostaglandin-Glycerylesters

2-Arachidonoylglycerol (2-AG) is an endogenous ligand of the cannabinoid receptors (CB_1 and CB_2), members of a family of GPCRs best known for mediating effects of the active components of marijuana.[209,210] The brain and peripheral tissues contain abundant levels of 2-AG.[211] *In vivo*, 2-AG is inactivated by serine hydrolases, predominantly monacylglycerol lipase (MGL)[212] and, to a lesser extent, two currently uncharacterized enzymes, ABHD6 and ABHD12.[213] The glycerol moiety is also susceptible to nonenzymatic, base-catalyzed acyl migration, converting 2-AG to the 1(3)-isomer.[214] The fatty acyl chain of 2-AG is identical to that of arachidonate, suggesting it might undergo some of the same reactions as the fatty acid; indeed oxidation of 2-AG by LOXs and PGHS has been established.[215–217]

1.03.5.1 Structural Requirements for Oxygenation of 2-Arachidonoylglycerol by PGHS

PGHS-2, but not PGHS-1, converts 2-AG to PGH_2-glycerylester (PGH_2-G) with efficiency similar to that of arachidonate.[217] The stereo- and regiochemistry of the reaction is the same as observed with arachidonate.[108] Among arachidonoyl esters, PGHS-2 displays a marked preference for 2-AG. Metabolism of 1-AG or the unsubstituted ethanolester of arachidonate occurs at half the rate of 2-AG oxygenation. Interactions between 2-AG and PGHS-2 have been probed using site-directed mutagenesis.[108,217] Tyr385 is required for oxygenation of 2-AG, and truncation of the top channel of the active site (G533V) prevents metabolism as well.[217] The constriction site provides important stabilizing interactions for 2-AG binding, as it does for arachidonate, although the contributions of individual residues are distinct. As described, R120Q has little effect on binding of arachidonic acid to PGHS-2, whereas E524L and Y355F reduce binding affinity.[124] With R120Q and E524L mutants, 2-AG metabolism is almost completely abrogated,[108,217] whereas mutation of Tyr355 has little effect on oxygenation of 2-AG.[108] As with arachidonic acid, acetylation of PGHS-2 or mutation of Ser530 to methionine favors monooxygenation of 2-AG, with a marked preference for oxygen insertion at C-15 of the acyl chain.

Isoform selectivity of 2-AG metabolism is largely determined by interactions in the side pocket of PGHS (**Figure 2**).[108] A particularly critical element is the identity of residue 513, which is arginine in PGHS-2. Mutation of this single residue to histidine, the substitution found in ovine PGHS-1, reduces 2-AG oxygenation by more than 50%. An additional reduction in 2-AG oxidation is observed in a triple mutant, V434I/R513H/V523I (VRV), which replaces all three residues contributing to isoform differences in the side pocket with the corresponding ovine PGHS-1 residues. However, even the VRV mutant does not completely reduce 2-AG oxygenation to the level of PGHS-1, suggesting that additional determinants of substrate selectivity exist.

1.03.5.2 Metabolism of PGH_2-Glycerylester and Its Products

PGH_2-G can be converted to nearly the full range of prostanoid-glyceryl ester species that are observed for PGH_2.[217,218] In a coupled assay with PGHS-2 and human thromboxane synthase, the conversion of 2-AG to TXA_2-G is less than 5% as efficient as the conversion of arachidonate to TXA_2;[218] it is proposed that the thromboxane pathway is not physiologically relevant for 2-AG. In the systems studied, efficacy of conversion of PGH_2-G to other prostanoids is within twofold of PGH_2. Because many studies have been conducted in whole cells or using microsomal preps, the efficiencies and specificities of prostaglandin synthase isoforms have not been determined. PGE_2-G is a substrate for 15-hydroxy prostaglandin dehydrogenase (15-HPGDH), although it is metabolized less efficiently than PGE_2; oxidation of $PGF_{2\alpha}$-G by 15-HPGDH is not observed.[219]

The stability of prostaglandin-glycerylesters (PG-Gs) to hydrolysis is a question of import when searching for products *in vivo*.[219] PGE_2-G is rapidly hydrolyzed in rat plasma with a half-life of approximately 15 s. In human plasma, though, minimal hydrolysis is detected after 10 min, and PGE_2-G has a half-life of 7 min in human whole blood. No hydrolysis of PGE_2-G is observed in cerebrospinal fluid. PG-Gs also seem to be stable

to hydrolysis in cell culture systems. PG-Gs are rapidly hydrolyzed in dog brain homogenates, but they are not good substrates for the endocannabinoid hydrolyzing enzymes, monoacylglycerol lipase (MGL) and fatty acid amide hydrolase (FAAH).[220]

1.03.5.3 Synthesis of Prostaglandin-Glycerylesters in Cells and *In Vivo*

PG-Gs are produced in PGHS-expressing cells. Following addition of exogenous 2-AG, RAW264.7 cells generate PGD_2-G,[217,218] whereas HCA-7 cells produce PGE_2-G and $PGF_{2\alpha}$-G.[218] In response to zymosan, a stimulus that induces PGHS-2, murine resident peritoneal macrophages generate PGE_2-G and PGI_2-G from endogenous 2-AG stores.[221] Although arachidonic acid and 2-AG are released at a ratio of 10:1, prostaglandin levels are 1000-fold higher than those of PG-Gs.

PG-G synthesis does not seem to be fully dependent upon PGHS-2 in mouse resident peritoneal macrophages. Treatment of these macrophages with a PGHS-2-selective inhibitor reduces PG-G levels but not completely.[221] Macrophages harvested from PGHS-2-knockout animals generate approximately the same amount of PG-Gs as wild type, but synthesis of PG-Gs is almost completely ablated in macrophages from PGHS-1 knockout mice.[222] This finding was quite unexpected, given the marked preference of PGHS-2 for 2-AG *in vitro*. However, it may reflect the fact that resident murine peritoneal macrophages express very high levels of PGHS-1 in the absence of activation by LPS or zymosan. The ability of murine PGHS-1 to catalyze 2-AG oxidation has not been examined *in vitro*. However, residue 513 is glutamine in murine PGHS-1 instead of histidine, found in human and ovine PGHS-1, [130] so it is conceivable that mouse PGHS-1 is more efficient at oxidizing 2-AG than ovine PGHS-1.

Recently, endogenous formation of PGE_2-G in the rat hindpaw was described.[223] As in studies with murine resident peritoneal macrophages, the levels of PGE_2-G are 1000-fold lower than PGE_2. Notably, PGE_2-G levels are not increased by inflammatory stimuli, although PGE_2 levels are elevated. Injection of PGE_2-G into the hindpaw reveals that PGE_2-G is rapidly metabolized, with less than 10% remaining immediately following injection. Some hydrolysis of PGE_2-G to PGE_2 is observed, but quantitative studies to determine the metabolic fate of PGE_2-G have not been conducted.

1.03.5.4 Pharmacologic Actions of Prostaglandin-Glycerylesters

The effects of PGE_2-G in a macrophage-like cell line (RAW264.7) have been examined. Treatment of RAW264.7 cells with PGE_2-G results in Ca^{2+} mobilization that is dose dependent and occurs in the picomolar range.[224] The response is dependent upon the generation of IP_3 from hydrolysis of PIP_2. The concomitant hydrolysis product, diacylglycerol, activates protein kinase C (PKC)-dependent phosphorylation of extracellular signal-regulated kinase (ERK). PGE_2-G activates NF-κB at low concentrations but suppresses activation at higher doses.[223] PGE_2-G displays only low-affinity binding to canonical prostanoid receptors, and it is unlikely that signaling is mediated through them.[224]

Effects of PGE_2-G on the nervous system have also been investigated. In cultured neurons, synaptic transmission induced by neurotransmitters can be modulated by PGE_2-G. PGE_2-G dose dependently enhances γ-aminobutyric acid (GABA)-receptor mediated inhibitory synaptic transmission and abolishes depolarization-induced suppression of inhibition.[225] These actions are the opposite of those of 2-AG and are largely independent of prostanoid and endocannabinoid receptors. Mitogen-activated protein kinase (MAPK) and IP_3 pathways mediate the action of PGE_2-G, but the effects are not dependent upon PKC or protein kinase A. Similarly, excitatory glutamatergic synaptic signaling in cultured neurons is enhanced by PGE_2-G, which is dependent upon MAPK and IP_3.[226] Phosphorylation of ERK, p38 MAPK, and NF-κB is stimulated by PGE_2-G. Potentiation of excitatory glutamatergic synaptic transmission by PGE_2-G can induce neuronal injury and death.[223] In another study, *in vivo* administration of PGE_2-G augments pain responses in the peripheral nervous system. The responses can be partially but not fully blocked by EP receptor antagonists. These studies provide evidence for the existence of unique receptors for PG-Gs. Further work is required to define the receptors mediating effects of PG-Gs and to determine what agents stimulate PG-G synthesis *in vivo*.

1.03.6 Biosynthesis, Degradation, and Pharmacology of Prostaglandin-Ethanolamides

Arachidonoylethanolamide (AEA, also referred to as anandamide) was the first molecule identified as an endogenous ligand of the cannabinoid receptors.[227] AEA also binds to numerous voltage- and ligand-gated ion channels.[228] *In vivo*, levels of AEA are tightly regulated by FAAH, which catalyzes the hydrolysis of AEA to arachidonic acid and ethanolamine.[229,230] As with 2-AG, the acyl chain of AEA is susceptible to oxidative metabolism by LOXs and PGHS-2.[12,231–233] AEA can also be oxidized by cytochromes P-450.[234–237] The biosynthesis, degradation, and pharmacology of prostaglandin-ethanolamides (PG-EAs) have been studied in great detail.

1.03.6.1 Structural Determinants for Oxygenation of Arachidonoylethanolamide by PGHS

In vitro and in intact cells, PGHS-2 oxygenates AEA to yield PGH_2-EA.[233,238] The efficiency of AEA metabolism by PGHS-2 is approximately fivefold lower than for arachidonic acid, owing largely to an increase in the K_M for the substrate. AEA is not metabolized by human PGHS-1[233] but is a substrate, albeit a poor one, for ovine PGHS-1.[239] An extensive structure–activity study of neutral arachidonoyl amides was conducted, comparing the ability of PGHS-1 and PGHS-2 to catalyze their oxygenation.[239] In the series, substitution of the hydroxyl of AEA with either methoxy or methyl reduced metabolism by either isoform. Addition of one or two methyl groups at the 1- or 2-position of the ethanolamide had no effect, or in some cases, improved oxygenation of the substrate by PGHS-1. Substitution of ethanol at the 1- or 2-position also improved PGHS-2-dependent metabolism.

Site-directed mutagenesis of PGHS-2 has illuminated important interactions for the binding and metabolism of AEA. The effects of mutations are quite similar to those observed with 2-AG. As with 2-AG, Arg120 provides an important stabilizing interaction, and mutation to glutamine reduces AEA metabolism threefold.[239] E524L affects binding of AEA and arachidonate to a similar extent,[239] and hydrogen bonding to Tyr355 is not required at all for AEA oxygenation.[239,240] Unlike 2-AG, isoform selectivity of AEA oxygenation is solely dependent upon the identity of residue 513.[239] Mutation of Arg513 to histidine in PGHS-2 reduces the rate of AEA oxidation to that of ovine PGHS-1. Mutagenesis studies combined with the structure–activity series indicate that the acyl chain of AEA binds in a similar orientation as arachidonic acid, with the ethanolamide binding in the side pocket of PGHS-2 and forming a hydrogen bond with Arg513.

1.03.6.2 Metabolism of PGH_2-Ethanolamide and Its Products

PGH_2-EA is susceptible to similar transformations as PGH_2 *in vitro*.[218] As with PGH_2-G, PGH_2-EA is a poor substrate for thromboxane synthase, and biosynthesis of TXA_2-EA likely does not occur under physiological conditions. With the exception of a newly characterized PGFS, the specificity and efficiency of PGH_2-EA conversion by prostanoid synthases have not been rigorously defined. PGE_2-EA is oxidized to 15-keto-PGE_2-EA by 15-HPGDH. PG-EAs are much more stable to metabolic inactivation than PG-Gs. Unlike AEA, PG-EAs are very poor substrates for FAAH.[220] In human and rat plasma, PGE_2-EA exhibits a half-life greater than 5 h.[219] The major reaction in plasma is dehydration on the cyclopentane ring to PGB_2-EA. *In vivo*, PGE_2-EA is still detectable in rat plasma up to 2 h after administration, with a half-life of approximately 6 min.

1.03.6.3 Synthesis of Prostaglandin-Ethanolamides in Cells and *In Vivo*

Similar to 2-AG, cultured cells expressing PGHS-2 can convert exogenous AEA to PG-EAs. PGE_2-EA is produced by human foreskin fibroblast cells,[233] PGE_2-EA and PGD_2-EA by RAW264.7 cells,[218,241] and PGE_2-EA and $PGF_{2\alpha}$-EA in HCA-7 cells.[218] Following administration of AEA to mice, PGE_2/D_2-EA are found at low levels in the kidney and lung.[242] In FAAH knockout mice, $PGF_{2\alpha}$-EA, PGE_2-EA, and PGD_2-EA are detected in the kidney and lung, and $PGF_{2\alpha}$-EA is also found in the liver.

1.03.6.4 Pharmacologic Actions of Prostaglandin-Ethanolamides

The biological effects of $PGF_{2\alpha}$-EA and bimatoprost, a structural analog, have been extensively characterized. These compounds reduce intraocular pressure *in vivo*,[243] and bimatoprost is currently marketed to control intraocular pressure in glaucoma and ocular hypertension.[244] $PGF_{2\alpha}$-EA and bimatoprost do not exhibit high-affinity binding to canonical prostanoid receptors.[243,245] Although their effects require a pathway distinct to that of $PGF_{2\alpha}$, deletion of the gene encoding the FP receptor ablates the action of $PGF_{2\alpha}$-EA and bimatoprost.[246,247] These amides may mediate their effects via binding to an FP heterodimer consisting of a molecule of canonical receptor and a molecule of alternatively spliced receptor.[248]

PGE_2-EA may have a role in inflammatory signaling.[249] AEA and PGE_2-EA suppress expression of the proinflammatory cytokine IL-12 by activated microglial cells and RAW264.7 cells. This effect is mediated in part by activation of a repressor site, GA-12, within the promoter region of the p40 subunit of IL-12. Effects seem to be mediated in part by EP_2 receptor,[249] even though PGE_2-EA exhibits low affinity for the EP receptors.[250] Additional study of the pathways mediating PG-EA actions is warranted.

1.03.7 Metabolism of Other Arachidonoyl Amides by PGHS

Over the past decade, additional fatty acyl amides have been identified as potential signaling molecules, and the ability of PGHS to metabolize many of the arachidonoyl derivatives has been established. PGHS-2 is able to metabolize a range of arachidonoyl amino acids.[251,252] Of these amino acid conjugates, it most efficiently metabolizes *N*-arachidonoyl glycine (NAGly), and the structural determinants for oxygenation of NAGly by PGHS-2 closely mirror those for 2-AG.[252] Although not the most efficient substrate for PGHS-2, NAGly is the most isoform-selective substrate discovered to date. *N*-Arachidonoyl alanine and *N*-arachidonoyl-γ-aminobuytric acid can also serve as substrates for PGHS-2 *in vitro*.[251] A taurine conjugate of arachidonic acid, bearing a sulfonic acid, is not metabolized by either PGHS isoform.[253] Bulky, neutral substitutions such as observed with *N*-arachidonoyl dopamine serve as poor substrates for PGHS as well.[251] Although oxidation of these arachidonoyl amides has been established *in vitro*, it remains to be seen whether this occurs in cells or *in vivo* under physiologically relevant conditions.

1.03.8 Conclusions

PGHS is clearly a central player in many physiological and pathophysiological processes. Yet, genetic alterations of PGHS in animal models have revealed unexpected isoform-specific functions. Modification of PGHS by reactive nitrogen species and specific control of protein degradation may have important contributions to the spatiotemporal regulation of PGHS activity. Although the conversion of arachidonic acid to prostanoids mediates many of the biological effects of PGHS, the enzymes' ability to generate novel classes of eicosanoids from other substrates may provide additional pathways to drive distinct physiological processes. Many of these transformations are efficiently catalyzed by PGHS-2 but not PGHS-1. Findings from various cellular and *in vivo* studies allude to the existence of PGHS-2-specific signaling pathways. Such pathways may contribute to profound phenotypes observed in PGHS-2-deficient animals and could explain why PGHS-1 cannot fully complement a PGHS-2-knockout even when under the control of the PGHS-2 promoter. Characterization of stimuli that induce synthesis of novel eicosanoids and the pathways regulated by them is a challenge that lies ahead. However, as the roles of enzyme isoforms continue to be defined, the possible impact of alternative substrates and their metabolism by PGHS must be considered.

Abbreviations

15-HPGDH	15-hydroxy prostaglandin dehydrogenase
ALA	α-linolenic acid
ASA	acetylsalicylic acid; aspirin

DHA	docosahexaenoic acid
EET	epoxyeicosatrienoic acid
eNOS	endothelial NOS
EPA	eicosapentaenoic acid
ERK	extracellular signal-regulated kinase
FAAH	fatty acid amide hydrolase
GPCR	G-protein coupled receptor
HDHA	hydroxydocosahexaenoic acid
HEPE	hydroxyeicosapentaenoic acid
HETE	hydroxyeicosatetraenoic acid
HpETE	hydroperoxyeicosatetraenoic acid
IL	interleukin
iNOS	inducible NOS
LOX	lipoxygenase
LPS	bacterial lipopolysaccharide
MAPK	mitogen-activated protein kinase
MGL	monoacylglycerol lipase
NF-κB	nuclear factor-κB
nNOS	neuronal/brain NOS
NSAID	nonsteroidal anti-inflammatory drug
PG	prostaglandin
PG-EA	prostaglandin-ethanolamide
PG-G	prostaglandin-glycerylester
PGHS	prostaglandin endoperoxide synthase
PGI_2	prostacyclin I_2
PKC	protein kinase C
TXA_2	thromboxane A_2

References

1. J. D. Clark; L.-L. Lin; R. W. Kriz; C. S. Ramesha; L. A. Sultzman; A. Y. Lin; N. Milona; J. L. Knopf, *Cell* **1991**, *65*, 1043–1051.
2. J. H. Capdevila; J. R. Falck, *Prostaglandins Other Lipid Mediat.* **2002**, *68–69*, 325–344.
3. S. Yamamoto, *Free Radic. Biol. Med.* **1991**, *10*, 149–159.
4. L. J. Marnett, *Curr. Opin. Chem. Biol.* **2000**, *4*, 545–552.
5. D. H. Nugteren; E. Hazelhof, *Biochim. Biophys. Acta* **1973**, *326*, 448–461.
6. M. Hamberg; B. Samuelsson, *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 899–903.
7. M. Hamberg; J. Svensson; T. Wakabayashi; B. Samuelsson, *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 345–349.
8. R. J. A. Helliwell; L. F. Adams; M. D. Mitchell, *Prostaglandins Leukot. Essent. Fatty Acids* **2004**, *70*, 101–113.
9. C. L. Bos; D. J. Richel; T. Ritsema; M. P. Peppelenbosch; H. H. Versteeg, *Int. J. Biochem. Cell Biol.* **2004**, *36*, 1187–1205.
10. J. R. Vane, *Nat. New. Biol.* **1971**, *231*, 232–235.
11. W. L. Smith; W. E. Lands, *J. Biol. Chem.* **1971**, *246*, 6700–6702.
12. T. D. Penning; J. J. Talley; S. R. Bertenshaw; J. S. Carter; P. W. Collins; S. Docter; M. J. Graneto; L. F. Lee; J. W. Malecha; J. M. Miyashiro; R. S. Rogers; D. J. Rogier; S. S. Yu; G. D. Anderson; E. G. Burton; J. N. Cogburn; S. A. Gregory; C. M. Koboldt; W. E. Perkins; K. Seibert; A. W. Veenhuizen; Y. Y. Zhang; P. C. Isakson, *J. Med. Chem.* **1997**, *40*, 1347–1365.
13. A. Tanaka; S. Hase; T. Miyazawa; R. Ohno; K. Takeuchi, *J. Pharmacol. Exp. Ther.* **2002**, *303*, 1248–1254.
14. P. M. Kearney; C. Baigent; J. Godwin; H. Halls; J. R. Emberson; C. Patrono, *BMJ* **2006**, *332*, 1302–1308.
15. D. Picot; P. J. Loll; R. M. Garavito, *Nature* **1994**, *367*, 243–249.
16. F. J. Van der Ouderaa; M. Buytenhek; D. H. Nugteren; D. A. Van Dorp, *Biochim. Biophys. Acta* **1977**, *487*, 315–331.
17. F. J. van der Ouderaa; M. Buytenhek; F. J. Slikkerveer; D. A. van Dorp, *Biochim. Biophys. Acta* **1979**, *572*, 29–42.
18. T. Miyamoto; N. Ogino; S. Yamamoto; O. Hayaishi, *J. Biol. Chem.* **1976**, *251*, 2629–2636.
19. M. Hemler; W. E. Lands, *J. Biol. Chem.* **1976**, *251*, 5575–5579.
20. M. K. O'Banion; V. D. Winn; D. A. Young, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4888–4892.
21. W. L. Xie; J. G. Chipman; D. L. Robertson; R. L. Erikson; D. L. Simmons, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2692–2696.
22. T. Hla; K. Neilson, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7384–7388.
23. K. L. T. Roos; D. L. Simmons, *Biochem. Biophys. Res. Commun.* **2005**, *338*, 62–69.

24. J. K. Gierse; S. D. Hauser; D. P. Creely; C. Koboldt; S. H. Rangwala; P. C. Isakson; K. Seibert, *Biochem. J.* **1995**, *305*, 479–484.
25. R. J. Kulmacz; L. H. Wang, *J. Biol. Chem.* **1995**, *270*, 24019–24023.
26. R. G. Kurumbail; A. M. Stevens; J. K. Gierse; J. J. McDonald; R. A. Stegeman; J. Y. Pak; D. Gildehaus; J. M. Miyashiro; T. D. Penning; K. Seibert; P. C. Isakson; W. C. Stallings, *Nature* **1996**, *384*, 644–648.
27. A. G. Spencer; J. W. Woods; T. Arakawa; I. I. Singer; W. L. Smith, *J. Biol. Chem.* **1998**, *273*, 9886–9893.
28. H. Sano; Y. Kawahito; R. L. Wilder; A. Hashiramoto; S. Mukai; K. Asai; S. Kimura; H. Kato; M. Kondo; T. Hla, *Cancer Res.* **1995**, *55*, 3785–3789.
29. K. T. Wilson; S. Fu; K. S. Ramanujam; S. J. Meltzer, *Cancer Res.* **1998**, *58*, 2929–2934.
30. E. de Moraes; N. A. Dar; C. V. de Moura Gall; P. Hainaut, *Int. J. Cancer* **2007**, *121*, 929–937.
31. K. Seibert; Y. Zhang; K. Leahy; S. Hauser; J. Masferrer; W. Perkins; L. Lee; P. Isakson, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12013–12017.
32. T. S. Brannon; A. J. North; L. B. Wells; P. W. Shaul, *J. Clin. Invest.* **1994**, *93*, 2230–2235.
33. C. J. Smith; J. D. Morrow; L. J. Roberts; L. J. Marnett, *Biochem. Biophys. Res. Commun.* **1993**, *192*, 787–793.
34. M. K. Jones; H. Wang; B. M. Peskar; E. Levin; R. M. Itani; I. J. Sarfeh; A. S. Tarnawski, *Nat. Med.* **1999**, *5*, 1418–1423.
35. H. Lim; B. C. Paria; S. K. Das; J. E. Dinchuk; R. Langenbach; J. M. Trzaskos; S. K. Dey, *Cell* **1997**, *91*, 197–208.
36. J. N. Topper; J. Cai; D. Falb; M. A. Gimbrone, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10417–10422.
37. R. C. Harris; Y. A. McKanna; H. R. Jacobson; R. N. DuBois; M. D. Breyer, *J. Clin. Invest.* **1994**, *94*, 2504–2510.
38. T. J. Smith; T. A. Jennings; D. Sciaky; H. J. Cao, *J. Biol. Chem.* **1999**, *274*, 15622–15632.
39. R. W. Walenga; M. Kester; E. Coroneos; S. Butcher; R. Dwivedi; C. Statt, *Prostaglandins* **1996**, *52*, 341–359.
40. C. D. Breder; C. B. Saper, *Brain Res.* **1996**, *713*, 64–69.
41. C. D. Breder; D. Dewitt; R. P. Kraig, *J. Comp. Neurol.* **1995**, *355*, 296–315.
42. J. M. Schwab; R. Beschorner; R. Meyermann; F. Gozalan; H. J. Schluesener, *J. Neurosurg.* **2002**, *96*, 892–899.
43. O. Pepicelli; E. Fedele; M. Berardi; M. Raiteri; G. Levi; A. Greco; M. A. Ajmone-Cat; L. Minghetti, *J. Neurochem.* **2005**, *93*, 1561–1567.
44. E. Candelario-Jalil; A. Gonzalez-Falcon; M. Garcia-Cabrera; D. Alvarez; S. Al-Dalain; G. Martinez; O. S. Leon; J. E. Springer, *J. Neurochem.* **2003**, *86*, 545–555.
45. S. Aid; R. Langenbach; F. Bosetti, *J. Neuroinflammation* **2008**, *5*, 17.
46. S. H. Choi; R. Langenbach; F. Bosetti, *FASEB J.* **2008**, *22*, 1491–1501.
47. S. Moncada; E. A. Higgs, *Br. J. Pharmacol.* **2006**, *147*, S193–S201.
48. C. Bogdan, *Nat. Immunol.* **2001**, *2*, 907–916.
49. P. Tripathi; P. Tripathi; L. Kashyap; V. Singh, *FEMS Immunol. Med. Microbiol.* **2007**, *51*, 443–452.
50. T. C. Bellamy; J. Garthwaite, *Mol. Cell Biochem.* **2002**, *230*, 165–176.
51. P. Pacher; J. S. Beckman; L. Liaudet, *Physiol. Rev.* **2007**, *87*, 315–424.
52. W. K. Alderton; C. E. Cooper; R. G. Knowles, *Biochem. J.* **2001**, *357*, 593–615.
53. N. V. Blough; O. C. Zafiriou, *Inorg. Chem.* **1985**, *24*, 3502–3504.
54. J. S. Beckman; T. W. Beckman; J. Chen; P. A. Marshall; B. A. Freeman, *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 1620–1624.
55. R. Kissner; T. Nauser; P. Bugnon; P. G. Lye; W. H. Koppenol, *Chem. Res. Toxicol.* **1997**, *10*, 1285–1292.
56. S. V. Lymar; J. K. Hurst, *J. Am. Chem. Soc.* **1995**, *117*, 8867–8868.
57. M. G. Bonini; R. Radi; G. Ferrer-Sueta; A. M. D. C. Ferreira; O. Augusto, *J. Biol. Chem.* **1999**, *274*, 10802–10806.
58. G. Merenyi; J. Lind, *Chem. Res. Toxicol.* **1998**, *11*, 243–246.
59. O. Augusto; R. M. Gatti; R. Radi, *Arch. Biochem. Biophys.* **1994**, *310*, 118–125.
60. R. Radi, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 4003–4008.
61. D. T. Hess; A. Matsumoto; S. O. Kim; H. E. Marshall; J. S. Stamler, *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 150–166.
62. A. Trostchansky; H. Rubbo, *Free Radic. Biol. Med.* **2008**, *44*, 1887–1896.
63. S. Burney; J. L. Caulfield; J. C. Niles; J. S. Wishnok; S. R. Tannenbaum, *Mutat. Res.* **1999**, *424*, 37–49.
64. J. C. Niles; J. S. Wishnok; S. R. Tannenbaum, *Nitric Oxide* **2006**, *14*, 109–121.
65. S. Lamas; P. A. Marsden; G. K. Li; P. Tempst; T. Michel, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 6348–6352.
66. S. P. Janssens; A. Shimouchi; T. Quertermous; D. B. Bloch; K. D. Bloch, *J. Biol. Chem.* **1992**, *267*, 14519–14522.
67. J. A. Mitchell; P. Akarasereenont; C. Thiemermann; R. J. Flower; J. R. Vane, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11693–11697.
68. D. S. Bredt; P. M. Hwang; S. H. Snyder, *Nature* **1990**, *347*, 768–770.
69. H. Kleinert; A. Pautz; K. Linker; P. M. Schwarz, *Eur. J. Pharmacol.* **2004**, *500*, 255–266.
70. K.-S. Chun; Y.-J. Surh, *Biochem. Pharmacol.* **2004**, *68*, 1089–1100.
71. J. R. Vane; J. A. Mitchell; I. Appleton; A. Tomlinson; D. Bishop-Bailey; J. Croxtall; D. A. Willoughby, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 2046–2050.
72. J. A. Corbett; G. Kwon; J. Turk; M. L. McDaniel, *Biochemistry* **1993**, *32*, 13767–13770.
73. V. Mollace; C. Muscoli; E. Masini; S. Cuzzocrea; D. Salvemini, *Pharmacol. Rev.* **2005**, *57*, 217–252.
74. T. Tetsuka; D. Daphna-Iken; B. W. Miller; Z. Guan; L. D. Baier; A. R. Morrison, *J. Clin. Invest.* **1996**, *97*, 2051–2056.
75. A. Habib; C. Bernard; M. Lebret; C. Creminon; B. Esposito; A. Tedgui; J. Maclouf, *J. Immunol.* **1997**, *158*, 3845–3851.
76. H. W. Cook; W. E. Lands, *Biochem. Biophys. Res. Commun.* **1975**, *65*, 464–471.
77. M. E. Hemler; G. Graff; W. E. Lands, *Biochem. Biophys. Res. Commun.* **1978**, *85*, 1325–1331.
78. L. M. Landino; B. C. Crews; M. D. Timmons; J. D. Morrow; L. J. Marnett, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 15069–15074.
79. S. Schildknecht; M. Bachschmid; V. Ullrich, *FASEB J.* **2005**, *19*, 1169–1171.
80. R. S. Deeb; H. Shen; C. Gamss; T. Gavrilova; B. D. Summers; R. Kraemer; G. Hao; S. S. Gross; M. Laine; N. Maeda; D. P. Hajjar; R. K. Upmacis, *Am. J. Pathol.* **2006**, *168*, 349–362.
81. S. Schildknecht; K. Heinz; A. Daiber; J. Hamacher; C. Kavakli; V. Ullrich; M. Bachschmid, *Biochem. Biophys. Res. Commun.* **2006**, *340*, 318–325.
82. C. Boulos; H. Jiang; M. Balazy, *J. Pharmacol. Exp. Ther.* **2000**, *293*, 222–229.

83. R. S. Deeb; G. Hao; S. S. Gross; M. Laine; J. H. Qiu; B. Resnick; E. J. Barbar; D. P. Hajjar; R. K. Upmacis, *J. Lipid Res.* **2006**, *47*, 898–911.
84. D. C. Goodwin; M. R. Gunther; L. C. Hsi; B. C. Crews; T. E. Eling; R. P. Mason; L. J. Marnett, *J. Biol. Chem.* **1998**, *273*, 8903–8909.
85. M. R. Gunther; L. C. Hsi; J. F. Curtis; J. K. Gierse; L. J. Marnett; T. E. Eling; R. P. Mason, *J. Biol. Chem.* **1997**, *272*, 17086–17090.
86. A. Trostchansky; V. B. O'Donnell; D. C. Goodwin; L. M. Landino; L. J. Marnett; R. Radi; H. Rubbo, *Free Radic. Biol. Med.* **2007**, *42*, 1029–1038.
87. R. S. Deeb; M. J. Resnick; D. Mittar; T. McCaffrey; D. P. Hajjar; R. K. Upmacis, *J. Lipid Res.* **2002**, *43*, 1718–1726.
88. J. Tian; S. F. Kim; L. Hester; S. H. Snyder, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 10537–10540.
89. S. Atar; Y. Ye; Y. Lin; S. Y. Freeberg; S. P. Nishi; S. Rosanio; M. H. Huang; B. F. Uretsky; J. R. Perez-Polo; Y. Birnbaum, *Am. J. Physiol. Heart Circ. Physiol.* **2006**, *290*, H1960–H1968.
90. S. F. Kim; D. A. Huri; S. H. Snyder, *Science* **2005**, *310*, 1966–1970.
91. T. Tanabe; N. Tohnai, *Prostaglandins Other Lipid Mediat.* **2002**, *68–69*, 95–114.
92. Y.-J. Kang; U. R. Mbonye; C. J. DeLong; M. Wada; W. L. Smith, *Prog. Lipid Res.* **2007**, *46*, 108–125.
93. P. Rockwell; H. Yuan; R. Magnusson; M. E. Figueiredo-Pereira, *Arch. Biochem. Biophys.* **2000**, *374*, 325–333.
94. J. Shao; H. Sheng; H. Inoue; J. D. Morrow; R. N. DuBois, *J. Biol. Chem.* **2000**, *275*, 33951–33956.
95. U. R. Mbonye; C. Yuan; C. E. Harris; R. S. Sidhu; I. Song; T. Arakawa; W. L. Smith, *J. Biol. Chem.* **2008**, *283*, 8611–8623.
96. U. R. Mbonye; M. Wada; C. J. Rieke; H. Y. Tang; D. L. Dewitt; W. L. Smith, *J. Biol. Chem.* **2006**, *281*, 35770–35778.
97. H. Neuss; X. Huang; B. K. Hetfeld; R. Deva; P. Henklein; S. Nigam; J. W. Mall; W. Schwenk; W. Dubiel, *J. Mol. Med.* **2007**, *85*, 961–970.
98. R. Langenbach; S. G. Morham; H. F. Tiano; C. D. Loftin; B. I. Ghanayem; P. C. Chulada; J. F. Mahler; C. A. Lee; E. H. Goulding; K. D. Kluckman; H. S. Kim; O. Smithies, *Cell* **1995**, *83*, 483–492.
99. J. E. Dinchuk; B. D. Car; R. J. Focht; J. J. Johnston; B. D. Jaffee; M. B. Covington; N. R. Contel; V. M. Eng; R. J. Collins; P. M. Czerniak; S. A. Gorry; J. M. Trzaskos, *Nature* **1995**, *378*, 406–409.
100. S. G. Morham; R. Langenbach; C. D. Loftin; H. F. Tiano; N. Vouloumanos; J. C. Jennette; J. F. Mahler; K. D. Kluckman; A. Ledford; C. A. Lee; O. Smithies, *Cell* **1995**, *83*, 473–482.
101. C. D. Loftin; D. B. Trivedi; H. F. Tiano; J. A. Clark; C. A. Lee; J. A. Epstein; S. G. Morham; M. D. Breyer; M. Nguyen; B. M. Hawkins; J. L. Goulet; O. Smithies; B. H. Koller; R. Langenbach, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 1059–1064.
102. Y. Yu; J. Fan; X. S. Chen; D. Wang; A. J. Klein-Szanto; R. L. Campbell; G. A. FitzGerald; C. D. Funk, *Nat. Med.* **2006**, *12*, 699–704.
103. Y. Yu; J. Fan; Y. Hui; C. A. Rouzer; L. J. Marnett; A. J. Klein-Szanto; G. A. FitzGerald; C. D. Funk, *J. Biol. Chem.* **2007**, *282*, 1498–1506.
104. E. F. Pettersen; T. D. Goddard; C. C. Huang; G. S. Couch; D. M. Greenblatt; E. C. Meng; T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.
105. M. F. Sanner; A. J. Olson; J. C. Spehner, *Biopolymers* **1996**, *38*, 305–320.
106. A. Philippsen, *DINO: Visualizing Structural Biology* **2002**.
107. C. Luong; A. Miller; J. Barnett; J. Chow; C. Ramesha; M. F. Browner, *Nat. Struct. Biol.* **1996**, *3*, 927–933.
108. K. R. Kozak; J. J. Prusakiewicz; S. W. Rowlinson; C. Schneider; L. J. Marnett, *J. Biol. Chem.* **2001**, *276*, 30072–30077.
109. G. Xiao; W. Chen; R. J. Kulmacz, *J. Biol. Chem.* **1998**, *273*, 6801–6811.
110. C. Yuan; C. J. Rieke; G. Rimon; B. A. Wingerd; W. L. Smith, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 6142–6147.
111. R. J. Kulmacz; W. E. Lands, *J. Biol. Chem.* **1985**, *260*, 12572–12578.
112. J. Liu; S. A. Seibold; C. J. Rieke; I. Song; R. I. Cukier; W. L. Smith, *J. Biol. Chem.* **2007**, *282*, 18233–18244.
113. S. W. Rowlinson; B. C. Crews; C. A. Lanzo; L. J. Marnett, *J. Biol. Chem.* **1999**, *274*, 23305–23310.
114. R. Dietz; W. Nastainczyk; H. H. Ruf, *Eur. J. Biochem.* **1988**, *171*, 321–328.
115. R. Karthein; R. Dietz; W. Nastainczyk; H. H. Ruf, *Eur. J. Biochem.* **1988**, *171*, 313–320.
116. J. Schreiber; T. E. Eling; R. P. Mason, *Arch. Biochem. Biophys.* **1986**, *249*, 126–136.
117. R. P. Mason; B. Kalyanaraman; B. E. Tainer; T. E. Eling, *J. Biol. Chem.* **1980**, *255*, 5019–5022.
118. J. C. Wilson; G. Wu; A. L. Tsai; G. J. Gerfen, *J. Am. Chem. Soc.* **2005**, *127*, 1618–1619.
119. M. Hecker; V. Ullrich; C. Fischer; C. O. Meese, *Eur. J. Biochem.* **1987**, *169*, 113–123.
120. A. Mukherjee; D. W. Brinkley; K. M. Chang; J. P. Roth, *Biochemistry* **2007**, *46*, 3975–3989.
121. L. J. Marnett; K. R. Maddipati, Prostaglandin H Synthase. In *Peroxidases: Chemistry and Biology*; J. Everse, K. Everse, M. Grisham, Eds.; CRC Press: Boca Raton, FL, 1991; pp 293–334.
122. M. G. Malkowski; S. L. Ginell; W. L. Smith; R. M. Garavito, *Science* **2000**, *289*, 1933–1937.
123. D. K. Bhattacharyya; M. Lecomte; C. J. Rieke; M. Garavito; W. L. Smith, *J. Biol. Chem.* **1996**, *271*, 2179–2184.
124. C. J. Rieke; A. M. Mulichak; R. M. Garavito; W. L. Smith, *J. Biol. Chem.* **1999**, *274*, 17109–17114.
125. A. Tsai; L. C. Hsi; R. J. Kulmacz; G. Palmer; W. L. Smith, *J. Biol. Chem.* **1994**, *269*, 5085–5091.
126. R. J. Light; B. Samuelsson, *Eur. J. Biochem.* **1972**, *28*, 232–240.
127. W. P. Schneider; R. D. Hamilton; L. E. Rhuland, *J. Am. Chem. Soc.* **1972**, *94*, 2122–2123.
128. K. Valmsen; W. E. Boeglin; I. Jarving; C. Schneider; K. Varvas; A. R. Brash; N. Samel, *Eur. J. Biochem.* **2004**, *271*, 3533–3538.
129. C. Schneider; W. E. Boeglin; J. J. Prusakiewicz; S. W. Rowlinson; L. J. Marnett; N. Samel; A. R. Brash, *J. Biol. Chem.* **2002**, *277*, 478–485.
130. D. L. DeWitt; E. A. el-Harith; S. A. Kraemer; M. J. Andrews; E. F. Yao; R. L. Armstrong; W. L. Smith, *J. Biol. Chem.* **1990**, *265*, 5192–5198.
131. T. Shimokawa; W. L. Smith, *J. Biol. Chem.* **1992**, *267*, 12387–12392.
132. E. D. Thuresson; K. M. Lakkides; C. J. Rieke; Y. Sun; B. A. Wingerd; R. Micielli; A. M. Mulichak; M. G. Malkowski; R. M. Garavito; W. L. Smith, *J. Biol. Chem.* **2001**, *276*, 10347–10357.

133. M. Lecomte; O. Laneuville; C. Ji; D. L. DeWitt; W. L. Smith, *J. Biol. Chem.* **1994**, *269*, 13207–13215.
134. J. A. Mancini; G. P. O'Neill; C. Bayly; P. J. Vickers, *FEBS Lett.* **1994**, *342*, 33–37.
135. C. A. Harman; C. J. Rieke; R. M. Garavito; W. L. Smith, *J. Biol. Chem.* **2004**, *279*, 42929–42935.
136. J. R. Kiefer; J. L. Pawlitz; K. T. Moreland; R. A. Stegeman; W. F. Hood; J. K. Gierse; A. M. Stevens; D. C. Goodwin; S. W. Rowlinson; L. J. Marnett; W. C. Stallings; R. G. Kurumbail, *Nature* **2000**, *405*, 97–101.
137. C. Schneider; W. E. Boeglin; A. R. Brash, *J. Biol. Chem.* **2004**, *279*, 4404–4414.
138. G. J. Roth; N. Stanford; P. W. Majerus, *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 3073–3076.
139. G. J. Roth; E. T. Machuga; J. Ozols, *Biochemistry* **1983**, *22*, 4672–4675.
140. J. P. Merlie; D. Fagan; J. Mudd; P. Needleman, *J. Biol. Chem.* **1988**, *263*, 3550–3553.
141. D. L. DeWitt; W. L. Smith, *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1412–1416.
142. C. Yokoyama; T. Takai; T. Tanabe, *FEBS Lett.* **1988**, *231*, 347–351.
143. E. A. Meade; W. L. Smith; D. L. DeWitt, *J. Biol. Chem.* **1993**, *268*, 6610–6614.
144. P. J. Loll; D. Picot; R. M. Garavito, *Nat. Struct. Biol.* **1995**, *2*, 637–643.
145. M. J. Holtzman; J. Turk; L. P. Shornick, *J. Biol. Chem.* **1992**, *267*, 21438–21445.
146. G. P. O'Neill; J. A. Mancini; S. Kargman; J. Yergey; M. Y. Kwan; J. P. Falgueyret; M. Abramovitz; B. P. Kennedy; M. Ouellet; W. Cromlish; S. Culp; J. F. Evans; A. W. Ford-Hutchinson; P. J. Vickers, *Mol. Pharmacol.* **1994**, *45*, 245–254.
147. S. W. Rowlinson; B. C. Crews; D. C. Goodwin; C. Schneider; J. K. Gierse; L. J. Marnett, *J. Biol. Chem.* **2000**, *275*, 6586–6591.
148. J. Claria; C. N. Serhan, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9475–9479.
149. N. Chiang; E. A. Bermudez; P. M. Ridker; S. Hurwitz; C. N. Serhan, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15178–15183.
150. T. Takano; C. B. Clish; K. Gronert; N. Petasis; C. N. Serhan, *J. Clin. Invest.* **1998**, *101*, 819–826.
151. C. B. Clish; J. A. O'Brien; K. Gronert; G. L. Stahl; N. A. Petasis; C. N. Serhan, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 8247–8252.
152. T. Takano; S. Fiore; J. F. Maddox; H. R. Brady; N. A. Petasis; C. N. Serhan, *J. Exp. Med.* **1997**, *185*, 1693–1704.
153. L. Jozsef; C. Zouki; N. A. Petasis; C. N. Serhan; J. G. Filep, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13266–13271.
154. A. T. Gewirtz; L. S. Collier-Hyams; A. N. Young; T. Kucharzik; W. J. Guilford; J. F. Parkinson; I. R. Williams; A. S. Neish; J. L. Madara, *J. Immunol.* **2002**, *168*, 5260–5267.
155. K. Gronert; T. Martinsson-Niskanen; S. Ravasi; N. Chiang; C. N. Serhan, *Am. J. Pathol.* **2001**, *158*, 3–9.
156. M. Pouliot; C. N. Serhan, *J. Periodontal Res.* **1999**, *34*, 370–373.
157. M. E. Brezinski; M. A. Gimbrone; K. C. Nicolaou; C. N. Serhan, *FEBS Lett.* **1989**, *245*, 167–172.
158. M. J. Paul-Clark; T. van Cao; N. Moradi-Bidhendi; D. Cooper; D. W. Gilroy, *J. Exp. Med.* **2004**, *200*, 69–78.
159. P. F. Cezar-de-Mello; A. M. Vieira; V. Nascimento-Silva; C. G. Villela; C. Barja-Fidalgo; I. M. Fierro, *Br. J. Pharmacol.* **2008**, *153*, 956–965.
160. P. F. Cezar-de-Mello; V. Nascimento-Silva; C. G. Villela; I. M. Fierro, *Oncogene* **2006**, *25*, 122–129.
161. I. M. Fierro; J. L. Kutok; C. N. Serhan, *J. Pharmacol. Exp. Ther.* **2002**, *300*, 385–392.
162. F. S. Machado; J. E. Johndrow; L. Esper; A. Dias; A. Bafica; C. N. Serhan; J. Aliberti, *Nat. Med.* **2006**, *12*, 330–334.
163. C. M. Schaldach; J. Riby; L. F. Bjeldanes, *Biochemistry* **1999**, *38*, 7594–7600.
164. N. Chiang; K. Gronert; C. B. Clish; J. A. O'Brien; M. W. Freeman; C. N. Serhan, *J. Clin. Invest.* **1999**, *104*, 309–316.
165. D. Mitchell; S. J. O'Meara; A. Gaffney; J. K. G. Crean; B. T. Kinsella; C. Godson, *J. Biol. Chem.* **2007**, *282*, 15606–15618.
166. A. P. Simopoulos, *Exp. Biol. Med.* **2008**, *233*, 674–688.
167. A. P. Simopoulos, *J. Am. Coll. Nutr.* **2002**, *21*, 495–505.
168. P. C. Calder, *Am. J. Clin. Nutr.* **2006**, *83*, S1505–S1519.
169. O. Laneuville; D. K. Breuer; N. Xu; Z. H. Huang; D. A. Gage; J. T. Watson; M. Lagarde; D. L. DeWitt; W. L. Smith, *J. Biol. Chem.* **1995**, *270*, 19330–19336.
170. T. Ringbom; U. Huss; A. Stenholm; S. Flock; L. Skattebol; P. Perera; L. Bohlin, *J. Nat. Prod.* **2001**, *64*, 745–749.
171. E. Horia; B. A. Watkins, *J. Nutr. Biochem.* **2005**, *16*, 184–192.
172. M. Wada; C. J. DeLong; Y. H. Hong; C. J. Rieke; I. Song; R. S. Sidhu; C. Yuan; M. Warnock; A. H. Schmaier; C. Yokoyama; E. M. Smyth; S. J. Wilson; G. A. FitzGerald; R. M. Garavito; D. X. Sui; J. W. Regan; W. L. Smith, *J. Biol. Chem.* **2007**, *282*, 22254–22266.
173. I. Morita; R. Takahashi; Y. Saito; S. Murota, *J. Biol. Chem.* **1983**, *258*, 10197–10199.
174. P. Needleman; A. Raz; M. S. Minkes; J. A. Ferrendelli; H. Sprecher, *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 944–948.
175. M. G. Malkowski; E. D. Thuresson; K. M. Lakkides; C. J. Rieke; R. Micielli; W. L. Smith; R. M. Garavito, *J. Biol. Chem.* **2001**, *276*, 37547–37555.
176. S. Fischer; P. C. Weber, *Biochem. Biophys. Res. Commun.* **1983**, *116*, 1091–1099.
177. P. Needleman; M. Minkes; A. Raz, *Science* **1976**, *193*, 163–165.
178. D. W. Goldman; W. C. Pickett; E. J. Goetzl, *Biochem. Biophys. Res. Commun.* **1983**, *117*, 282–288.
179. T. H. Lee; J. M. Menica-Huerta; C. Shih; E. J. Corey; R. A. Lewis; K. F. Austen, *J. Biol. Chem.* **1984**, *259*, 2383–2389.
180. H. R. Knapp; N. Salem, Jr., *Prostaglandins* **1989**, *38*, 509–521.
181. S. Fischer; P. C. Weber, *Nature* **1984**, *307*, 165–168.
182. S. C. Larsson; M. Kumlin; M. Ingelman-Sundberg; A. Wolk, *Am. J. Clin. Nutr.* **2004**, *79*, 935–945.
183. P. Yang; D. Chan; E. Felix; C. Cartwright; D. G. Menter; T. Madden; R. D. Klein; S. M. Fischer; R. A. Newman, *J. Lipid Res.* **2004**, *45*, 1030–1039.
184. H. Funahashi; M. Satake; S. Hasan; H. Sawai; R. A. Newman; H. A. Reber; O. J. Hines; G. Eibl, *Pancreas* **2008**, *36*, 353–362.
185. J. Vanamala; A. Glagolenko; P. Yang; R. J. Carroll; M. E. Murphy; R. A. Newman; J. R. Ford; L. A. Braby; R. S. Chapkin; N. D. Turner; J. R. Lupton, *Carcinogenesis* **2008**, *29*, 790–796.
186. C. N. Serhan; C. B. Clish; J. Brannon; S. P. Colgan; N. Chiang; K. Gronert, *J. Exp. Med.* **2000**, *192*, 1197–1204.
187. W. Liu; D. Cao; S. F. Oh; C. N. Serhan; R. J. Kulmacz, *FASEB J.* **2006**, *20*, 1097–1108.
188. C. N. Serhan; S. Hong; K. Gronert; S. P. Colgan; P. R. Devchand; G. Mirick; R.-L. Moussignac, *J. Exp. Med.* **2002**, *196*, 1025–1037.
189. B. K. Lam; H. Aizan; Y. Sho; T. Yasushi; P. Y. K. Wong, *Biochim. Biophys. Acta* **1987**, *917*, 398–405.

190. E. Tjonahen; S. F. Oh; J. Siegelman; S. Elangovan; K. B. Percarpio; S. Hong; M. Arita; C. N. Serhan, *Chem. Biol.* **2006**, *13*, 1193–1202.
191. Y. P. Sun; S. F. Oh; J. Uddin; R. Yang; K. Gotlinger; E. Campbell; S. P. Colgan; N. A. Petasis; C. N. Serhan, *J. Biol. Chem.* **2007**, *282*, 9323–9334.
192. C. N. Serhan; N. Chiang; T. E. Van Dyke, *Nat. Rev. Immunol.* **2008**, *8*, 349–361.
193. M. Arita; T. Ohira; Y. P. Sun; S. Elangovan; N. Chiang; C. N. Serhan, *J. Immunol.* **2007**, *178*, 3912–3917.
194. R. Deva; R. Ciccoli; T. Schewe; J. L. F. Kock; S. Nigam, *Biochim. Biophys. Acta* **2000**, *1486*, 299–311.
195. P. Venter; J. L. Kock; G. S. Kumar; A. Botha; D. J. Coetzee; P. J. Botes; R. K. Bhatt; J. R. Falck; T. Schewe; S. Nigam, *Lipids* **1997**, *32*, 1277–1283.
196. S. R. Fox; M. Hamberg; J. Friend; C. Ratledge, *Lipids* **2000**, *35*, 1205–1214.
197. K. Starowicz; S. Nigam; V. Di Marzo, *Pharmacol. Ther.* **2007**, *114*, 13–33.
198. R. Ciccoli; S. Sahi; S. Singh; H. Prakash; M. P. Zafiriou; G. Ishdorj; J. L. Kock; S. Nigam, *Biochem. J.* **2005**, *390*, 737–747.
199. C. Schneider; W. E. Boeglin; H. Yin; D. F. Stec; M. Voehler, *J. Am. Chem. Soc.* **2006**, *128*, 720–721.
200. M. Romano; J. Claria, *FASEB J.* **2003**, *17*, 1986–1995.
201. M. Luo; S. M. Jones; M. Peters-Golden; T. G. Brock, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12165–12170.
202. P. K. Powell; I. Wolf; R. Jin; J. M. Lasker, *J. Pharmacol. Exp. Ther.* **1998**, *285*, 1327–1336.
203. J. M. Lasker; W. B. Chen; I. Wolf; B. P. Bloswick; P. D. Wilson; P. K. Powell, *J. Biol. Chem.* **2000**, *275*, 4118–4126.
204. J. C. McGiff; J. Quilley, *Curr. Opin. Nephrol. Hypertens.* **2001**, *10*, 231–237.
205. B. Escalante; W. C. Sessa; J. R. Falck; P. Yadagiri; M. L. Schwartzman, *J. Pharmacol. Exp. Ther.* **1989**, *248*, 229–232.
206. M. L. Schwartzman; J. R. Falck; P. Yadagiri; B. Escalante, *J. Biol. Chem.* **1989**, *264*, 11658–11662.
207. X. Fang; F. M. Faraci; T. L. Kaduce; S. Harmon; M. L. Modrick; S. Hu; S. A. Moore; J. R. Falck; N. L. Weintraub; A. A. Spector, *Am. J. Physiol. Heart Circ. Physiol.* **2006**, *291*, H2301–H2307.
208. E. R. Jacobs; R. M. Effros; J. R. Falck; K. M. Reddy; W. B. Campbell; D. Zhu, *Am. J. Physiol. Lung Cell Mol. Physiol.* **1999**, *276*, L280–L288.
209. N. Stella; P. Schweitzer; D. Piomelli, *Nature* **1997**, *388*, 773–778.
210. T. Sugiura; T. Kodaka; S. Nakane; T. Miyashita; S. Kondo; Y. Suhara; H. Takayama; K. Waku; C. Seki; N. Baba; Y. Ishima, *J. Biol. Chem.* **1999**, *274*, 2794–2801.
211. S. Kondo; H. Kondo; S. Nakane; T. Kodaka; A. Tokumura; K. Waku; T. Sugiura, *FEBS Lett.* **1998**, *429*, 152–156.
212. T. P. Dinh; D. Carpenter; F. M. Leslie; T. F. Freund; I. Katona; S. L. Sensi; S. Kathuria; D. Piomelli, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10819–10824.
213. J. L. Blankman; G. M. Simon; B. F. Cravatt, *Chem. Biol.* **2007**, *14*, 1347–1356.
214. C. A. Rouzer; K. Ghebreselasie; L. J. Marnett, *Chem. Phys. Lipids* **2002**, *119*, 69–82.
215. J. S. Moody; K. R. Kozak; C. Ji; L. J. Marnett, *Biochemistry* **2001**, *40*, 861–866.
216. K. R. Kozak; R. A. Gupta; J. S. Moody; C. Ji; W. E. Boeglin; R. N. DuBois; A. R. Brash; L. J. Marnett, *J. Biol. Chem.* **2002**, *277*, 23278–23286.
217. K. R. Kozak; S. W. Rowlinson; L. J. Marnett, *J. Biol. Chem.* **2000**, *275*, 33744–33749.
218. K. R. Kozak; B. C. Crews; J. D. Morrow; L. H. Wang; Y. H. Ma; R. Weinander; P. J. Jakobsson; L. J. Marnett, *J. Biol. Chem.* **2002**, *277*, 44877–44885.
219. K. R. Kozak; B. C. Crews; J. L. Ray; H. H. Tai; J. D. Morrow; L. J. Marnett, *J. Biol. Chem.* **2001**, *276*, 36993–36998.
220. A. Vila; A. Rosengarth; D. Piomelli; B. Cravatt; L. J. Marnett, *Biochemistry* **2007**, *46*, 9578–9585.
221. C. A. Rouzer; L. J. Marnett, *J. Biol. Chem.* **2005**, *280*, 26690–26700.
222. C. A. Rouzer; S. Tranguch; H. Wang; H. Zhang; S. K. Dey; L. J. Marnett, *Biochem. J.* **2006**, *399*, 91–99.
223. S. S. Hu; H. B. Bradshaw; J. S. Chen; B. Tan; J. M. Walker, *Br. J. Pharmacol.* **2008**, *153*, 1538–1549.
224. C. S. Nirodi; B. C. Crews; K. R. Kozak; J. D. Morrow; L. J. Marnett, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 1840–1845.
225. N. Sang; J. Zhang; C. Chen, *J. Physiol.* **2006**, *572*, 735–745.
226. N. Sang; J. Zhang; C. Chen, *J. Neurochem.* **2007**, *102*, 1966–1977.
227. W. A. Devane; L. Hanus; A. Breuer; R. G. Pertwee; L. A. Stevenson; G. Griffin; D. Gibson; A. Mandelbaum; A. Etinger; R. Mechoulam, *Science* **1992**, *258*, 1946–1949.
228. M. Oz, *Pharmacol. Ther.* **2006**, *111*, 114–144.
229. B. F. Cravatt; D. K. Giang; S. P. Mayfield; D. L. Boger; R. A. Lerner; N. B. Gilula, *Nature* **1996**, *384*, 83–87.
230. B. F. Cravatt; K. Demarest; M. P. Patricelli; M. H. Bracey; D. K. Giang; B. R. Martin; A. H. Lichtman, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 9371–9376.
231. N. Ueda; K. Yamamoto; S. Yamamoto; T. Tokunaga; E. Shirakawa; H. Shinkai; M. Ogawa; T. Sato; I. Kudo; K. Inoue; H. Takizawa; T. Nagano; M. Hirobe; N. Matsuki; H. Saito, *Biochim. Biophys. Acta* **1995**, *1254*, 127–134.
232. W. S. Edgemond; C. J. Hillard; J. R. Falck; C. S. Kearn; W. B. Campbell, *Mol. Pharmacol.* **1998**, *54*, 180–188.
233. M. Yu; D. Ives; C. S. Ramesha, *J. Biol. Chem.* **1997**, *272*, 21181–21186.
234. L. M. Bornheim; K. Y. Kim; B. Chen; M. A. Correia, *Biochem. Pharmacol.* **1995**, *50*, 677–686.
235. M. D. F. P.G. Katarina Stark, *FEBS J.* **2008**, *275*, 3706–3717.
236. N. T. Snider; A. M. Kornilov; U. M. Kent; P. F. Hollenberg, *J. Pharmacol. Exp. Ther.* **2007**, *321*, 590–597.
237. N. T. Snider; M. J. Sikora; C. Sridar; T. J. Feuerstein; J. M. Rae; P. F. Hollenberg, *J. Pharmacol. Exp. Ther.* **2008**, *327*, 538–545.
238. W. Yang; J. Ni; D. F. Woodward; D. D. S. Tang-Liu; K.-H. J. Ling, *J. Lipid Res.* **2005**, *46*, 2745–2751.
239. K. R. Kozak; J. J. Prusakiewicz; S. W. Rowlinson; D. R. Prudhomme; L. J. Marnett, *Biochemistry* **2003**, *42*, 9041–9049.
240. O.-Y. So; L. E. Scarafia; A. Y. Mak; O. H. Callan; D. C. Swinney, *J. Biol. Chem.* **1998**, *273*, 5801–5807.
241. S. H. Burstein; R. G. Rossetti; B. Yagen; R. B. Zurier, *Prostaglandins Other Lipid Mediat.* **2000**, *61*, 29–41.
242. A. Weber; J. Ni; K.-H. J. Ling; A. Acheampong; D. D. S. Tang-Liu; R. Burk; B. F. Cravatt; D. Woodward, *J. Lipid Res.* **2004**, *45*, 757–763.
243. D. F. Woodward; A. H. P. Krauss; J. Chen; R. K. Lai; C. S. Spada; R. M. Burk; S. W. Andrews; L. Shi; Y. Liang; K. M. Kedzie; R. Chen; D. W. Gil; A. Kharlamb; A. Archeampong; J. Ling; C. Madhu; J. Ni; P. Rix; J. Usansky; H. Usansky; A. Weber; D. Welty; W. Yang; D. D. S. Tang-Liu; M. E. Garst; B. Brar; L. A. Wheeler; L. J. Kaplan, *Surv. Ophthalmol.* **2001**, *45*, S337–S345.

244. D. L. Eisenberg; C. B. Toris; C. B. Camras, *Surv. Opthalmol.* **2002**, *47*, S105–S115.
245. I. Matias; J. Chen; L. De Petrocellis; T. Bisogno; A. Ligresti; F. Fezza; A. H. P. Krauss; L. Shi; C. E. Protzman; C. Li; Y. Liang; A. L. Nieves; K. M. Kedzie; R. M. Burk; V. Di Marzo; D. F. Woodward, *J. Pharmacol. Exp. Ther.* **2004**, *309*, 745–757.
246. J. G. Crowston; J. D. Lindsey; C. A. Morris; L. Wheeler; F. A. Medeiros; R. N. Weinreb, *Invest. Ophthalmol. Vis. Sci.* **2005**, *46*, 4571–4577.
247. T. Ota; M. Aihara; S. Narumiya; M. Araie, *Invest. Ophthalmol. Vis. Sci.* **2005**, *46*, 4159–4163.
248. Y. Liang; D. F. Woodward; V. M. Guzman; C. Li; D. F. Scott; J. W. Wang; L. A. Wheeler; M. E. Garst; K. Landsverk; G. Sachs; A. H. Krauss; C. Cornell; J. Martos; S. Pettit; H. Fliri, *Br. J. Pharmacol.* **2008**, *154*, 1079–1093.
249. F. Correa; F. Docagne; D. Clemente; L. Mestre; C. Becker; C. Guaza, *Biochem. J.* **2008**, *409*, 761–770.
250. R. A. Ross; S. J. Craib; L. A. Stevenson; R. G. Pertwee; A. Henderson; J. Toole; H. C. Ellington, *J. Pharmacol. Exp. Ther.* **2002**, *301*, 900–907.
251. J. J. Prusakiewicz; M. V. Turman; A. Vila; H. L. Ball; A. H. Al-Mestarihi; V. D. Marzo; L. J. Marnett, *Arch. Biochem. Biophys.* **2007**, *464*, 260–268.
252. J. J. Prusakiewicz; P. J. Kingsley; K. R. Kozak; L. J. Marnett, *Biochem. Biophys. Res. Commun.* **2002**, *296*, 612–617.
253. M. V. Turman; P. J. Kingsley; C. A. Rouzer; B. F. Cravatt; L. J. Marnett, *Biochemistry* **2008**, *47*, 3917–3925.

Biographical Sketches

Melissa V. Turman received her B.S. in biochemistry from the University of North Carolina-Greensboro in 2003. She recently received her Ph.D. in chemistry from Vanderbilt University under the direction of Lawrence J. Marnett. Her main research interests focus on the interaction of substrate analogs with prostaglandin endoperoxide synthases and the molecular determinants of inhibition of these enzymes. She is the recipient of a Ruth L. Kirschstein National Research Service Award from the National Institutes of Health/ National Institute of Drug Addiction.

Lawrence J. Marnett received his undergraduate degree from Rockhurst College in 1969 and his Ph.D. in chemistry from Duke University in 1973 under the direction of Ned Porter. He did his postdoctoral work at the Karolinska Institute with Bengt Samuelsson and at Wayne State University with Paul Schaap. He began his academic career in 1975 at Wayne State

University, where he rose to professor of chemistry. In 1989, he moved to Vanderbilt University as the Mary Geddes Stahlman Professor of Cancer Research, professor of biochemistry, chemistry, and pharmacology. Marnett's research program focuses on the role of PGHS-2 in cancer and inflammation and on the contribution of inflammation and oxidative stress to the generation of DNA damage and mutation and alterations in cell signaling. Marnett has received an Outstanding Investigator Award and a MERIT Award from the National Cancer Institute, the Harvey Branscomb Distinguished Professorship, and the Stanley Cohen Prize at Vanderbilt University. He is a Fellow of the American Association for the Advancement of Science. He is the author of over 400 research publications and 13 patents. He is the founding and current editor-in-chief of the American Chemical Society journal, *Chemical Research in Toxicology* and is the director of the Vanderbilt Institute of Chemical Biology.

1.04 Mycolic Acid/Cyclopropane Fatty Acid/Fatty Acid Biosynthesis and Health Relations

David B. Kastrinsky, Nicholas S. McBride, Keriann M. Backus, Jason J. LeBlanc, and Clifton E. Barry, III, National Institutes of Health, Bethesda, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

1.04.1	**Introduction**	66
1.04.1.1	The Structure of a Mycobacterial Cell	67
1.04.2	**Structural Variation, Biosynthesis, and Function of the Core Cell Envelope Mycolic Acids**	67
1.04.2.1	Structural Varieties of Mycolic Acids	67
1.04.2.2	Biosynthesis of the Fatty Acid Core	70
1.04.2.3	Modifications to the Meromycolate Core Structure	72
1.04.2.4	Structural Effects of Mycolic Acids on the Cell Envelope	78
1.04.2.4.1	Thiacetazone and its impact on cyclopropane synthesis	80
1.04.2.5	Immunopathogenic Effects of Alterations in Mycolic Acid Structure	81
1.04.2.5.1	Variability of mycolic acid structures in the BCG vaccine	81
1.04.3	**Trehalose-Based Glycolipids**	82
1.04.3.1	Overview	82
1.04.3.2	Structural Features of TDM	82
1.04.3.2.1	Synthesis	84
1.04.3.2.2	Biophysical studies of TDM properties	85
1.04.3.3	Biosynthesis	85
1.04.3.3.1	The role of antigen 85 in TDM biosynthesis	85
1.04.3.3.2	Crystal structure	86
1.04.3.3.3	Enzymology	87
1.04.3.3.4	Inhibitor development	88
1.04.3.4	The Immunopathology of TDM	89
1.04.3.4.1	The structure–activity relationship of granuloma formation	90
1.04.3.4.2	Complement binding	91
1.04.3.4.3	The ‘cytokine storm’ response to TDM	91
1.04.3.4.4	NO production	92
1.04.3.4.5	TDM arrests phagosome maturation	92
1.04.3.4.6	Putting it all together – A model of granuloma formation	92
1.04.3.4.7	Angiogenesis: VEGF formation and granuloma stability	92
1.04.3.4.8	Antitumor properties	92
1.04.3.4.9	Biological properties of antigen 85	92
1.04.3.4.10	Vaccine potential of TDM and antigen 85	93
1.04.3.4.11	Humoral responses to trehalose dimycolates	94
1.04.3.5	CD1 – A Mammalian Lipid Recognition System	94
1.04.3.5.1	CD1 subtypes and lipid partners	95
1.04.3.6	Di-, Tri-, and Polyacyl Trehaloses	97
1.04.3.7	Sulfolipids	97
1.04.3.7.1	Sulfolipid biosynthesis	99
1.04.4	**Phthiocerol-Based Glycolipids**	101
1.04.4.1	Polyketide Synthases in Mycobacteria	101
1.04.4.2	Phthiocerol Dimycocerosates	101
1.04.4.3	The Phenolic Glycolipids	105
1.04.4.3.1	PDIM-less mutant strains of Mtb as vaccine candidates	106
1.04.5	**Lipoarabinomannan and Phosphatidylinositides**	108

1.04.5.1	Introduction to the Polar Glycolipids	108
1.04.5.2	Biosynthesis of LAM	108
1.04.5.3	Immunopathogenesis	111
1.04.6	**Exotic Glycolipids and Glycopeptidolipids**	112
1.04.6.1	Glycopeptidolipids and Sliding Biofilms	112
1.04.6.1.1	Biofilms and sliding	114
1.04.6.1.2	Biosynthesis of GPLs	114
1.04.6.2	Mycobactins	116
1.04.6.2.1	Structure	116
1.04.6.2.2	Biosynthesis	118
1.04.6.3	Iron Acquisition in Mtb	120
1.04.6.3.1	Current models for mycobacterial iron trafficking	120
1.04.6.3.2	Mycobactins and their analogues as antimicrobials	123
1.04.6.3.3	Other therapeutic uses of mycobacterial siderophores	124
1.04.6.3.4	Immunology of mycobactin	124
1.04.6.4	Mycolactones	126
1.04.6.4.1	Structure	127
1.04.6.4.2	Function	129
1.04.6.4.3	Biosynthesis	130
1.04.7	**Conclusions and Outlook**	131
References		133

1.04.1 Introduction

Mycobacteria have played a truly unique role in the evolution of humans and in shaping our history. They have emerged from swamps, marshes, bogs, lagoons, and even our water supplies to cause multiple well-known human diseases such as tuberculosis (TB) and leprosy, alongside lesser-known diseases such as Buruli ulcer, 'swimming pool granuloma', and atypical mycobacterial infections in the immunocompromised.[1] TB is perhaps the most serious of these infectious diseases, infecting more than 7 million humans every year with fatal outcomes for nearly 3 million.[2] Nonetheless, mycobacterial diseases other than TB are caused by more than 30 different organisms within this genus, including *Mycobacterium abscessus*, organisms of the *Mycobacterium avium–intracellulare* group, *Mycobacterium ulcerans*, and *Mycobacterium leprae*. Dating back thousands of years, these pathogens have caused suffering and death for millions every year.

Mycobacteria occupy a unique place among disease-causing organisms. They are truly a 'chemists' pathogen' because their pathogenicity derives primarily from interactions of lipid-containing factors produced by the bacilli and not from interactions between proteinaceous factors produced by the bacilli and the host. Scientists and physicians who study bacterial pathogenesis have historically been molecular biologists, cellular biologists, physicians, and protein biochemists who were ill-equipped to deal with virulence factors that were primarily lipids and carbohydrates, including some of truly extraordinary complexity. Until recently, these efforts have resulted in studies of mycobacterial pathogenesis characterized by two approaches: the 'traditional' approach of genetic knockouts and phenotypic evaluation, and the chemical approach of characterization and total synthesis without an in-depth biological analysis of the resulting molecules.

In this chapter, the major areas where the field has made substantial progress in understanding the chemical biology of mycobacterial pathogenesis and virulence throughout the past decade are surveyed. In particular, what we have learned about some of the extraordinary and complex natural products produced by the tubercle bacillus, commonly referred to as *Mycobacterium tuberculosis* (*Mtb*), and the mechanisms underlying their role in disease will be focused on. Several concepts will be discussed: that lipid and glycolipid virulence factors have rich and complex interactions with the human host; that lipidic virulence factors are responsive to the same evolutionary pressures that drive protein variation and therefore show structural variability between species and strains akin to that seen with protein virulence factors; and how these topics have challenged a new

generation of scientists interested in bacterial pathogens and human health. As a result, fundamental insights into human immunology and plausible routes of intervention for developing new treatment, prevention, and diagnostic strategies for mycobacterial diseases are beginning to emerge.

1.04.1.1 The Structure of a Mycobacterial Cell

Unlike proteins, lipids and glycolipids have no intrinsic structure in an aqueous milieu. They derive their structure by virtue of interactions of the individual species with other hydrophobic molecules in the cell envelope. This introduces a considerable challenge because of the complexity of the cell wall of the bacillus with which most of the molecules are either peripherally or intimately associated. Although the chemical composition of this envelope has been known for decades and extensively reviewed,[3–5] the three-dimensional structure remains elusive because of the lack of useful tools to characterize it. Unlike other biological membrane systems, the mycobacterial cell wall does not conform to a 'fluid mosaic' model of membrane structure with two-dimensional diffusion of components across a sea of perfectly aligned amphipathic molecules. Instead the structure of a single cell is composed literally of a single polymeric molecule with three concentric layers: peptidoglycan, arabinogalactan, and mycolic acids.

Our understanding of the physical organization of this polymer has been influenced heavily by the limited biophysical techniques that can be applied to the problem. X-ray diffraction experiments illustrated that the main chains of the mycolic acids were organized parallel to each other and freeze-fracture electron microscopy studies supported visually the notion that a second hydrophobic barrier existed outside of the plasma membrane.[6] Early electron microscopic evaluation of the coarse structure in thin sections cut from plastic-embedded bacilli revealed an electron-transparent zone subtended by a darkly stained boundary.[7,8] These observations inspired a model first proposed by Minnikin of the covalently attached mycolic acids stretched out full length with ancillary lipids and surface materials interdigitated between the asymmetric meromycolate chains to fill the void left by the shorter α-branches.[3,6,9,10]

This model of the cell envelope dominated the field for more than 20 years but has been substantially challenged by two recent observations. The first is that for the major mycobacterial water-filled porin, MspA, to span the membrane and to function properly, the phospholipid bilayer must be narrower than originally proposed.[11–13] The second conflicting observation derives from electron microscopy of cryopreserved cells. This revealed that the mycobacterial membrane is no more than 8 nm in thickness (compared with 7 nm of the plasmalemma).[14,15] These recent observations provide strong experimental evidence that the cell wall is no thicker than the plasma membrane and have engendered speculation about new possibilities for the physical arrangement of the complex substructures of the envelope. A consensus view has yet to emerge concerning the precise arrangement of these complex molecules.

The hydrophobic mycolates seem chemically unattractive as the molecules are in direct contact with the hydrophilic environment. Outside of this layer, but still in intimate contact, reside the numerous amphipathic molecules that are the direct interface for the pathogenic mycobacterium in its encounters with potential victims. This matrix of materials has been called the capsule and is home to such structurally diverse families of glycolipids as trehalose mono- and dimycolate, the phthiocerol dimycocerosate waxes, sulfolipids, lipoarabinomannas, and glycopeptidolipids (GPLs) that are the subject of this review.[3,10,16] The major families of the envelope lipids starting with their structures will be systematically worked through. Then their biosynthesis and putative functions and interactions with the human host from the literature published during the past 10 years will be surveyed.

1.04.2 Structural Variation, Biosynthesis, and Function of the Core Cell Envelope Mycolic Acids

1.04.2.1 Structural Varieties of Mycolic Acids

The inner leaflet of the cell wall of *Mtb* is composed of a single molecule, a biopolymer called mycolic acid–arabinogalactan–peptidoglycan (MAGP) as shown in **Figure 1**.

In terms of weight and volume, the most significant part of this complex is composed of mycolic acids and thus it is not surprising that mycobacteria and related genera have been termed the 'mycolata'. The presence or

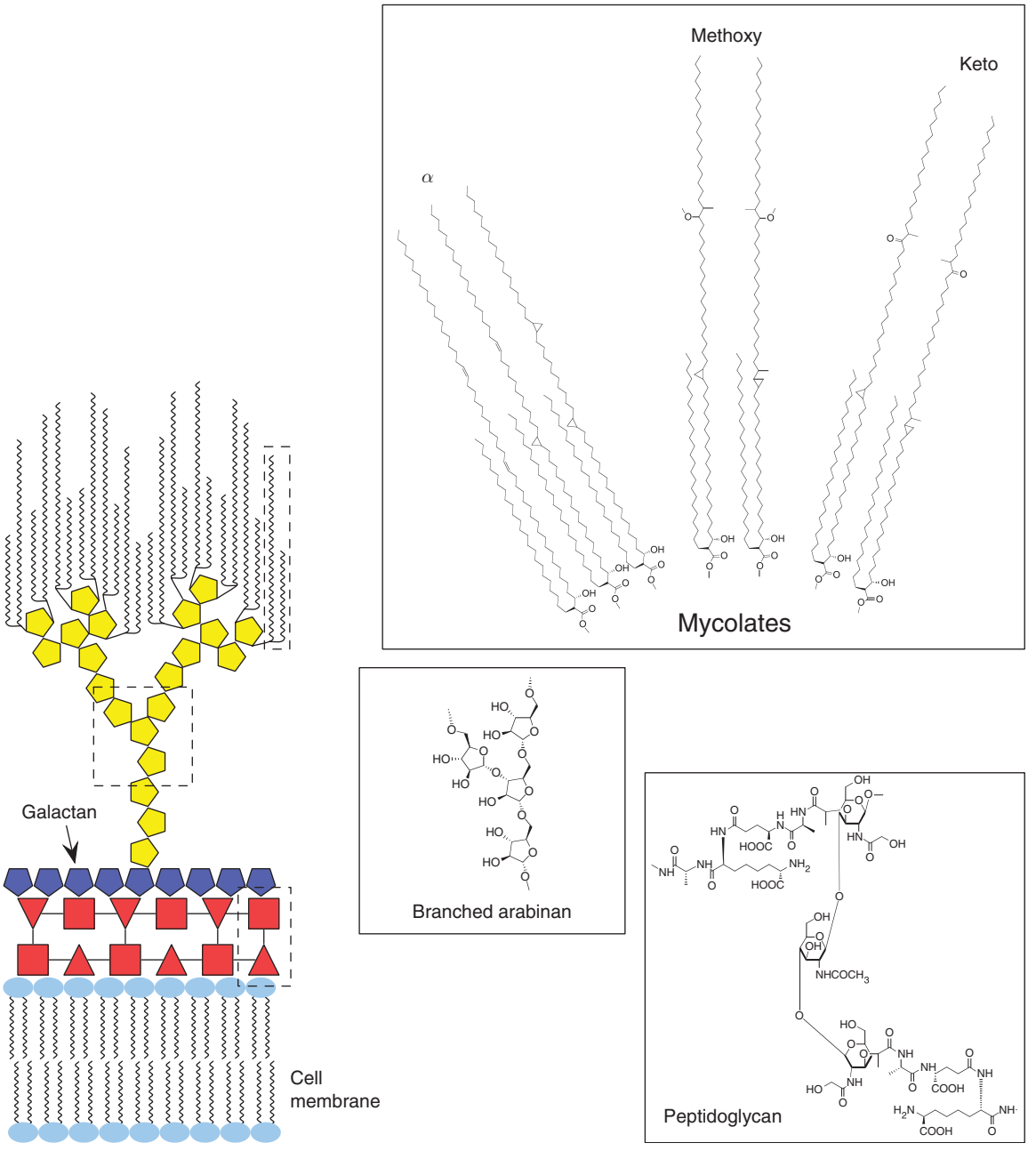


Figure 1 Inner leaflet of the *Mtb* cell wall showing the layers of mycolic acids, arabinan, galactan, and peptidoglycan outside of the plasma membrane.

absence of mycolic acids, α-alkyl-β-hydroxy acids with $C_{>30}$, is a genus-specific trait and the fine details of their structure are used in chemotaxonomic approaches to species identification. The mycolic acids are found in two very important molecules in the cell; bound to arabinogalactan in the inner leaflet and esterified as trehalose-6,6′-dimycolate (TDM) in the noncovalently bound outer leaflet. There do not appear to be any structural differences between mycolic acids present in TDM and those anchored in the cell wall. In several detailed studies of mycolic acid structure and function, the structures of mycolic acids throughout the mycolata vary based on three distinctions (see **Figure 2**):

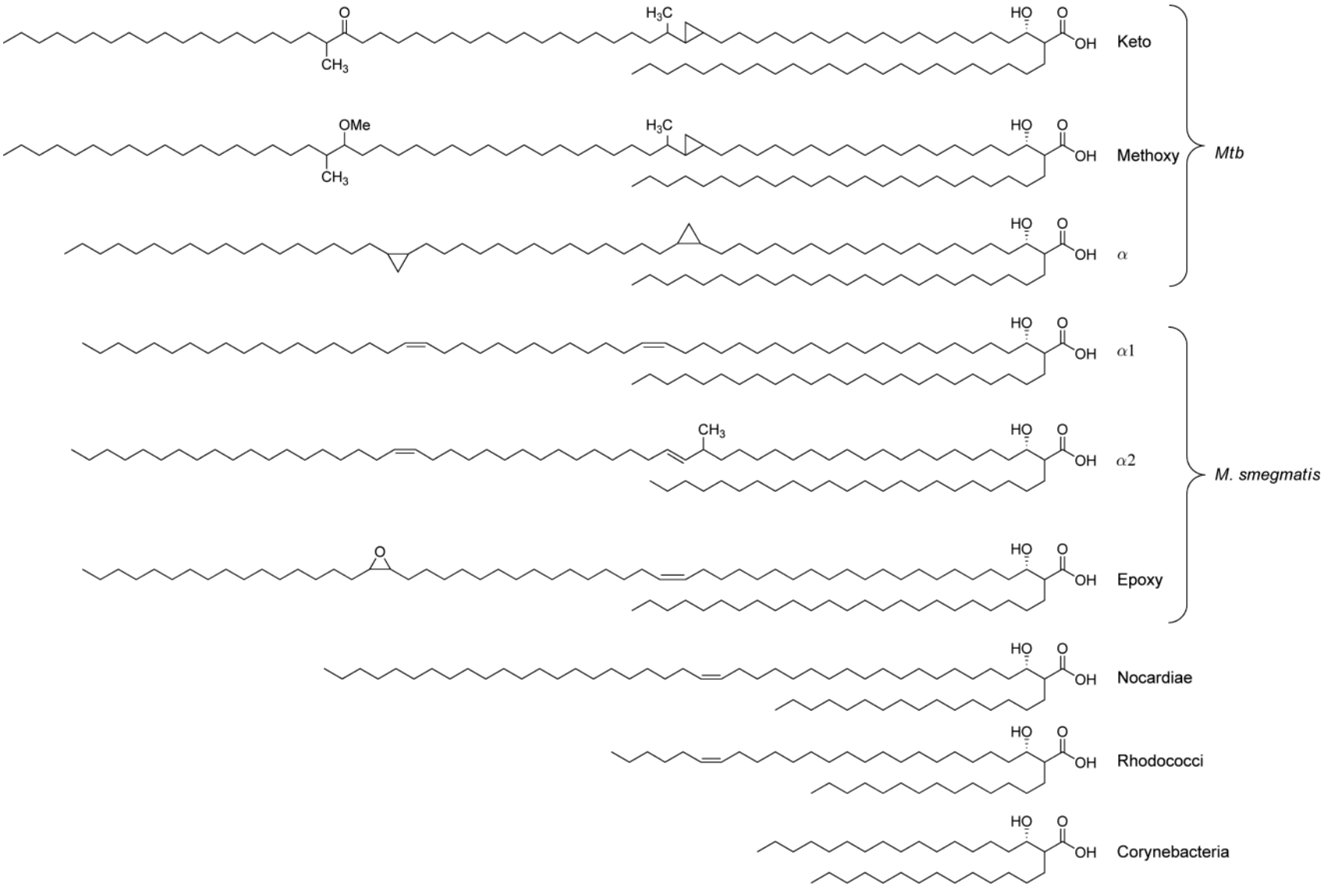


Figure 2 Different classes of mycolic acids from *Mtb* and *Mycobacterium smegmatis* and examples from other species.

1. Chain length – longer chain lengths increase toxicity, granulomagenicity, pathogenicity, and the phase transition temperature of the cell wall.
2. Oxygenated moieties – for example, the keto- and methoxy-mycolic acids characteristic of virulent strains and the epoxymycolates and wax esters that characterize some nontuberculous mycobacteria (NTM).
3. Cis- or trans-cyclopropanations, and *cis*- or *trans*-alkenes some with adjacent methyl branches.

The analysis of cell wall-bound mycolates requires treatment of whole cells, or delipidated cells, with tetrabutyl ammonium hydroxide, which saponifies the bound mycolyl esters. The mixture of mycolates so obtained can be converted into methyl esters and purified by silica gel column chromatography. To resolve the cyclopropane-containing mycolates from their alkene precursors, argentation chromatography, which uses silver nitrate-adsorbed silica gel, is employed. The alkenyl mycolic acids elute more slowly than the cyclopropyl forms because of a strong noncovalent interaction of silver ion with double bonds. Further structural analysis can be conducted by mass spectrometry (MS) and ^{1}H NMR spectroscopy. Initially, analysis of saponified mycolates by MS proved difficult because of pyrolysis of the β-hydroxy acids. However, the free β-hydroxy of the methyl esters can be trimethylsilylated, which prevents fragmentation and allows the components to be further separated and analyzed by gas chromatography (GC)/MS. The innovation of silylation[17] allowed for fine structural analysis of the intact mycolic acid by MS and revealed insights into mycobacterial cell wall structure and lipid biosynthesis.

The species-specific distribution patterns of mycolic acids have been reviewed previously[18] and will not be summarized here. But it is worth noting that technological advances have expanded our understanding of the distribution of structures in the mycolata. Some notable examples include the application of matrix-assisted laser desorption/ionization (MALDI) analysis[19–21] that allowed Fujita *et al.* to solve the structure of a variety of mycolic acids in TDMs (glycolipids that are esters of mycolic acids see below) from different actinomycetes. The mass distribution of the TDMs they analyzed ranged from 2.6 to almost 3 kDa. This technique allowed them to easily compare the lipid lengths, keto versus methoxy variability, and the degree of cyclopropanation from a variety of *Mtb*, *Mycobacterium bovis*, and *Mycobacterium kansasii* strains and presages a reinvention of chemotaxonomy[22] among the mycolata using increasingly high-resolution MS methodologies.

Of recent note, MALDI time-of-flight (MALDI-TOF) MS analysis of armadillo nostrils allowed the first comprehensive identification of *M. leprae*'s glycolipids. Previous methods lacked the sensitivity to detect the components of this recalcitrant pathogen that cannot be cultured *ex vivo*. TDM and trehalose-6,6′-monomycolate (TMM) were present and contained both α (C_{78}) and keto (C_{81}–C_{83}) mycolates. The keto-mycolates were slightly shorter in *M. leprae* than in *Mtb* and in contrast to *Mtb*, which synthesizes a wide variety of mycolates (*cis*- or *trans*-alkenes, and *cis*- or *trans*-cyclopropane-containing moieties), *M. leprae* contains only C_{81}–C_{83} *cis*-monoenoic keto-mycolates. Recent evidence showed that the *trans*-cyclopropanes play a role in virulence and the fine structure of these mycolates may explain the reduced pathogenic properties of *M. leprae* compared to *Mtb*.[23]

The mycolic acids of *Mtb* contain three general classes: α, methoxy, and keto whereas the mycolates of other mycolata can contain additional functionalities and lengths as illustrated in **Figure 2**.

The total length of the mero chain ($\sim C_{50}$–C_{65}) in mycobacteria differs between subclasses and follows the order $\alpha <$ methoxy $<$ keto. Conversely, the absolute quantity of each subclass found in the cell wall follows the order $\alpha \geq$ methoxy $>>$ keto. The α-mycolates may contain *cis*-alkenes, α-methyl-*trans*-alkenes, *cis*-cyclopropanes, or α-methyl-*trans*-cyclopropanes (**Figure 2**). The positions are labeled proximal and distal, an indication of their position with respect to the (*R*,*R*)-α-alkyl-β-hydroxy acid. The proximal sites in all cases (α, methoxy, and keto) contain *cis*- or *trans*-alkenes, or *cis*- or *trans*-cyclopropanes. In the case of the oxygenated forms, the proximal site is a *cis*- or α-methyl-*trans*-cyclopropane with an α-methyl keto or an α-methyl methoxy at the distal site.

1.04.2.2 Biosynthesis of the Fatty Acid Core

Different types of enzyme systems are employed for fatty acid synthesis across the five kingdoms of life. With the exception of plants, eukaryotic cells employ fatty acid synthase type I (FAS I), a large multifunctional enzyme that performs all of the basic chemistry. In contrast, plants and prokaryotic cells employ a type 2 system

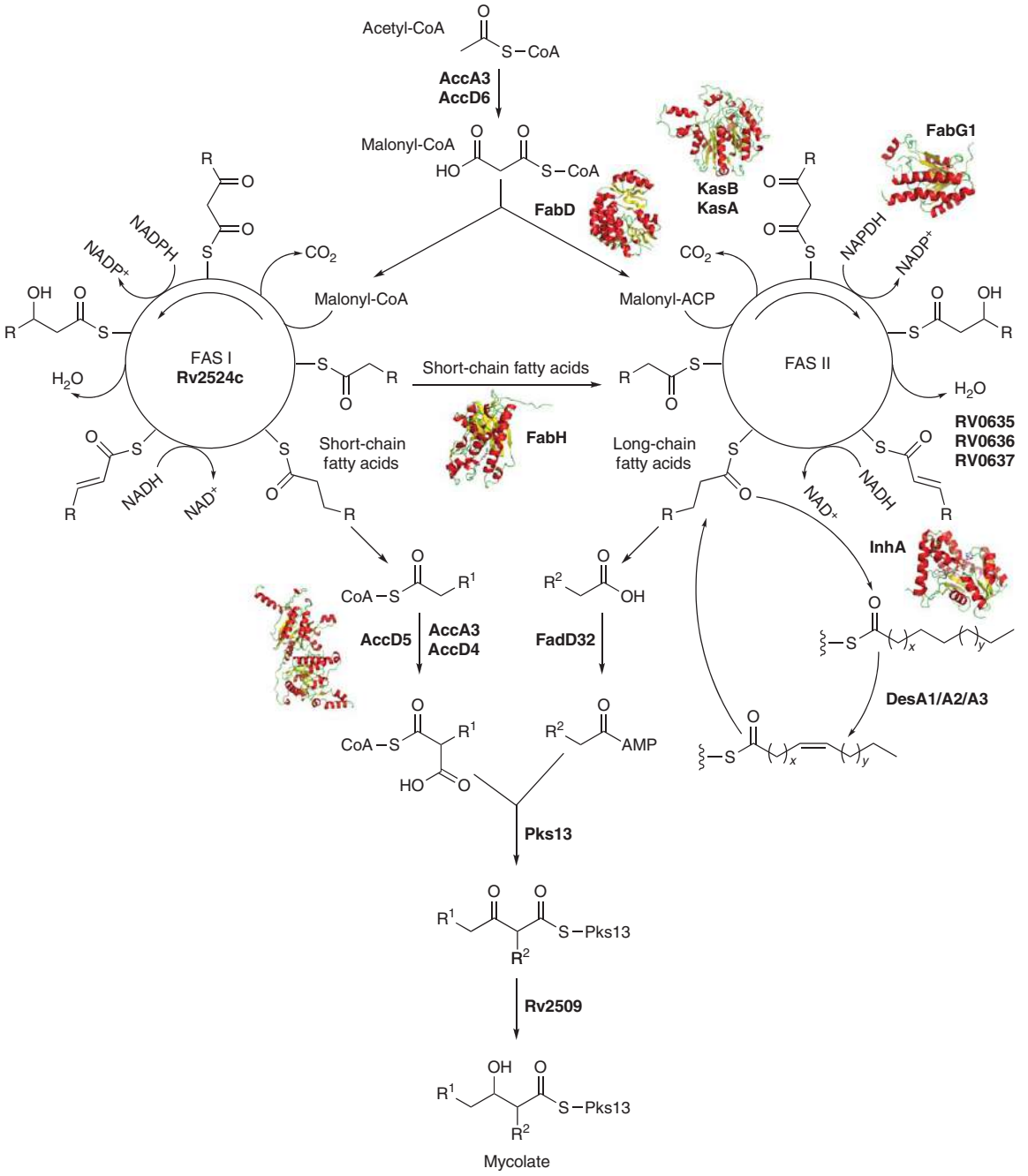


Figure 3 Mycolic acids biosynthesis showing FAS I polyketide synthesis of short-chain fatty acids and FAS II biosynthesis of long-chain fatty acids followed by the formation of the mycolate; includes renderings of PDB 1HZP, 1UZL, 1ZA0, 1ZID, 2BZR, and 2QC3.

(FAS II), that encodes the same activities but each on a different protein.[18] This is illustrated in **Figure 3**. The two core fatty acyl chains of mycolic acids (mero- and α-branches) are derived from a combination of FAS I and FAS II systems. In both systems, the synthesis follows a four-step iterative process that incorporates two carbons at a time catalyzed by the following enzyme domains: Claisen condensation between a growing chain and malonyl-coenzyme A (CoA) with concomitant decarboxylation of the β-keto ester (β-ketoacyl synthase, KAS); reduction of the β-keto ester (β-ketoacyl reductase, KR); dehydration (β-hydroxyacyl dehydratase, DH); and

enoyl reductase (reduction of α,β unsaturated enone, ER). Mycobacteria present an intriguing hybrid since they contain both FAS I and FAS II systems. FAS I produces fatty acids of 16–24 carbons bound to CoA. These can be extended, in turn, or condensed with other fatty acids by a separate FAS II system to provide meromycolates, mycolic acid α-branch precursors, and intact mycolic acids. The details of these two fascinating systems have been reviewed many times in the recent literature[24–27] and will not be exhaustively covered here as the basic synthetic steps do not directly influence the host–pathogen interaction. The basic importance of these early steps in mycolic acid biosynthesis is highlighted by the variety of drugs that target steps in this pathway (**Table 1**). Although the mechanisms of these are not discussed in detail, the interested reader is directed toward several comprehensive reviews.[28–30]

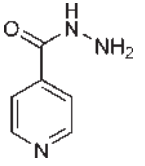
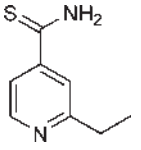
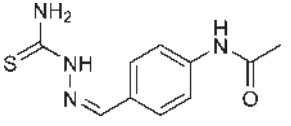
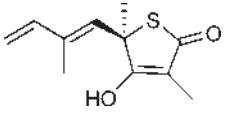
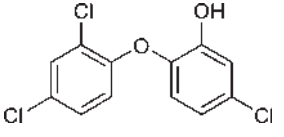
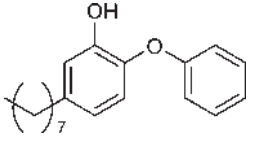
In vitro studies have shown that the methylation modifications (see below) occur on acyl carrier protein (ACP)-bound long-chain fatty acids, the precursor to the full-length meromycolate chain.[49] The final step in mycolic acid synthesis is the condensation of the α-branch with the meromycolate by a number of recently characterized enzymes (**Figure 3**). The mero acid is converted to the acyl adenylate by FadD32 (Rv3801c),[50] an acyl-adenosine monophosphate (AMP) ligase, and it is then loaded onto the condensase, pks13 (Rv3800c).[51] The identity of *pks13* had been assigned based on its corynebacterial homologue. Unlike mycobacteria, corynebacteria are still viable without cell wall-associated mycolates, and the knockout strains exhibit buildup of both mero- and α-branch intermediates. The *pks13* gene was proximal to FadD32 and a pair of acetyl-CoA carboxylases AccD4 (Rv3799) and AccD5 (Rv3280). The α-branch moiety emerges from the FAS I system as an acyl-CoA adduct. This species then undergoes carboxylation by the heterodimer consisting of AccD4 and AccD5 and which utilizes AccA3 (Rv3285) as the biotin carboxyl carrier.[50] This activated substrate then condenses with the mero chain. The ketoreductase has also been identified and characterized (Rv2509) through its homologue in corynebacteria (NCgl2385).[52]

1.04.2.3 Modifications to the Meromycolate Core Structure

Escherichia coli and other bacteria must modify shorter chain fatty acids, introduce unsaturation, and convert between cis and trans forms to modulate fluidity. The mycolic acids of virulent strains of mycobacteria are decorated with a wide variety of oxygenated functionalities, cyclopropanes, and *cis*- and α-methyl-*trans*-alkenes.[53] Mycolic acids can contain several degrees of unsaturation or the aforementioned functionality that derives from unsaturated precursors. This functionalization might occur after or during[49] the synthesis of an intact meromycolic acid containing at least two *cis*-alkenes on its mero chain. The origin of these two olefins could be a permutation of the normal fatty acid biosynthetic pathway. However, in *Mtb*, it appears more likely that a trio of enzymes – two soluble (DesA1 (Rv0824c) and DesA2 (Rv1094)), and one membrane-bound (DesA3 (Rv3229c)) – desaturate the meromycolyl precursors. These proteins are diiron desaturases that convert stearic acid into Δ^9 oleic acid, a precursor to membrane lipids like tuberculostearic acid (9-methyl stearic acid). DesA3 was found to be the target of the antituberculosis antibiotic isoxyl (4,4′-diisoamyloxydiphenylthiourea).[48] Rv3230c was identified as the oxidoreductase partner for DesA3.[54] These desaturases are closely associated with a number of genes involved in the mycolic acid biosynthesis pathway. Although a crystal structure for one of these has been solved (DesA2), important biochemical details as to the timing and positioning of the introduction of the double bonds remain to be clarified.[55] The cyclopropanation and oxygenated functionalities are added by a family of highly homologous, *S*-adenosyl-L-methionine (SAM)-dependent methyltransferases called cyclopropane mycolic acid synthases. These genes are present in *Mtb*, *M. bovis*, and other pathogenic strains, and code for proteins that convert mycolic acid alkenes into the cyclopropane-containing and oxygenated mycolates. Enzymes that synthesize *cis*-cyclopropanes are widely distributed in bacteria and plants, whereas *trans*-cyclopropane-synthesizing enzymes are rare, but present in *Mtb*. Mammalian fatty acids do not contain cyclopropanes, and the linkage of cyclopropanation with pathogen self-defense and virulence has generated much interest in these pathways.

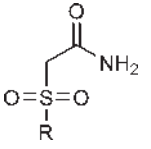
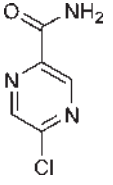
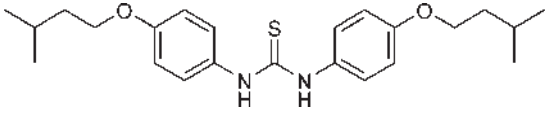
Identification of the biosynthetic machinery responsible for producing the cyclopropanated mycolates was not a simple task. Studies were limited by the inability to synthesize the true substrates for these reactions and the past difficulties in making genetic modifications to *Mtb*. Much of the initial work relied on introducing genes into *Mycobacterium smegmatis*, complemented by later work in *Mtb*, and performing a total lipid analyses to determine the effects on mycolic acid structure.

Table 1 Inhibitors of mycolic acid synthesis in *Mtb*

Inhibitor	*Structure*	*Process inhibited*	*Reference(s)*
Isoniazid (INH)		Reacts with KatG (catalase peroxidase) and NAD to form a covalent adduct that binds to InhA	31–34
Ethionamide		Prodrug, oxidized by EthA to sulfinic acid, inhibits InhA	31, 35, 36
Thiacetazone		Prodrug activated by EthA, inhibits cyclopropanation	37–39
Thiolactomycin		Reversibly inhibits FAS II	40
Triclosan		Inhibits FabI and InhA	41–43
5-Alkyl diphenyl ethers		FabI (200-fold more active than triclosan)	44

(*Continued*)

Table 1 (Continued)

Inhibitor	*Structure*	*Process inhibited*	*Reference(s)*
Sulfonylacetamides	 R = hydrophobic side chain	β-Ketoacyl synthase (KAS)	45
Pyrazinamide derivatives (5-chloropyrazinamide)		May inhibit FAS I pathway; target unclear	46, 47
Isoxyl		Inhibits DesA3	48

A cluster of four highly homologous genes with SAM-dependent methyltransferase domains was identified in a series of initial studies based on laborious screening of mycolates prepared from individual transformants by argentation thin-layer chromatography (TLC). The *Mtb* gene *cma1* (Rv3392c) was sufficient to convert the distal alkene to a cyclopropane in α-mycolates of *M. smegmatis*. A second *Mtb* gene, *cma2* (Rv0503c), identified through its homologue in *M. leprae*, was found to synthesize the proximal cyclopropane of the α-mycolates. Expression of both *cma1* and *cma2* produced a mycolate in *M. smegmatis* with both cyclopropanes nearly identical to the major species produced in *Mtb*. Isolation of epoxy mycolates with proximal cyclopropanes showed that *cma2* was promiscuous at the distal position (at least toward the mycolates in *M. smegmatis)* (**Figure 4**).[56,57]

Using similar heterologous expression techniques, *mma1* (Rv0645c), *mma2* (Rv0644c), *mma3* (Rv0643c), and *mma4* (Rv0642c) revealed the biosynthetic machinery responsible for the synthesis of oxygenated mycolic acids. A mechanism that explains the observed structures and the enzymes involved was proposed based on the formation of a common cationic intermediate by methylation of the *cis*-alkene precursors (**Figure 5**). This intermediate can react in three different ways, depending on the enzyme and the position of its active site residues, which would become clear in later studies. The methylene can add directly to the double bond, yielding the *cis*-cyclopropane. Alternatively, a suitably positioned base could deprotonate the carbon at the α-position relative to the double bond (A), or the cation could be quenched by H_2O to yield the hydroxyl (B). The high-resolution structures of four of these proteins are available from *Mtb*[58,59] and support the proposed mechanism. The presence of a carbonate ion at the active site that could function as a general base is also consistent with the carbonate dependence of the cyclopropane FAS from *E. coli*.[60] The hydroxy intermediate was formed at trace levels, could be isolated, and was proposed to be a common intermediate to both keto- and methoxy-mycolates. Coexpression of *mma3* and *mma4* demonstrated that *mma3* catalyzed the O-methylation responsible for producing methoxy-mycolates (C). Before this discovery, it was assumed that methoxy-mycolates were formed by the reduction of keto-mycolates (D). Without a reductase in the cluster, it is clear that keto-mycolates are not a precursor for methoxy-mycolate formation. To date, the enzymatic oxidase that putatively converts the hydroxyl to the keto form has not been discovered.[61]

The information provided by expression in a heterologous host was informative, but not conclusive since *M. smegmatis* produces its own mycolic acids with chain lengths and fine structures different from those seen in *Mtb*. The overexpression of the enzyme, MMAS-1 (*mma1*), in *Mtb* generated an increase in the presence of proximally modified, *trans*-alkene and *trans*-cyclopropane-containing oxygenated mycolates. There was also an increase in the ratio of keto- to methoxy-mycolates, so the formation of either the proximal *trans*-alkene or the cyclopropane initiates a pathway specific for oxygenated mycolate synthesis. This overproducing strain grew more slowly, and with higher cell wall fluidity and permeability, demonstrating a physical consequence of oxygenated mycolate upregulation.[62,63]

One of the insights gained by exploring mycolate content in *M. smegmatis* was the discovery that the *trans*-methyltransferases of *Mtb* and *M. smegmatis* exhibit selectivity for different positions of the double bond. It was initially unclear as to why only *cis*-cyclopropanes could be synthesized by the transformation of *M. smegmatis* with *Mtb cma1* and *cma2*, as it was assumed that *cma2* was generating the proximal *cis*-cyclopropane in both species. Deletion of *cma2* from *Mtb* provided oxygenated mycolates devoid of proximal *trans*-cyclopropanes.[64] Thus, the conundrum was that the *cma2* gene appeared incapable of synthesizing proximal *trans*-cyclopropane in *M. smegmatis* when that was its primary function in *Mtb*. Even more surprising was that the precursor, proximal α-methyl-*trans*-alkene, was present. A closer, more detailed look at the fine structure of mycolic acids using ^{13}C NMR with mycolic acids produced from 1- and 2-^{13}C-labeled acetate and [^{13}C-methyl]-L-methionine in *M. smegmatis* revealed the source of the discrepancy.[53] The methyltransferase in *M. smegmatis* produces a proximal α-methyl-*trans*-alkene with the methyl substituent closer to the mycolic acid C-terminus, and consequently this structure was not a substrate for *cma2*. The methyltransferase of *Mtb* produces the alternative, α-methyl-*trans*-alkene, as previously noted in **Figure 2** (compare the position of the methyl group in the TB *trans*-cyclopropyl mycolates with the methyl group in the *M. smegmatis* α2 series). The key observation was that the location of the methyl adduct is precisely on an odd- (*M. smegmatis*) or an even-numbered (*Mtb*) carbon with respect to the C-terminus, a result of the different reaction regioselectivity.

Concurrently, Daffé and colleagues isolated the same cluster of four methyltransferase genes (*cmaA–cmaD*) present in the TB vaccination strain *M. bovis* Bacillus Calmette–Guérin (BCG).[65] Introduction of this cluster into *M. smegmatis* produced three new, cyclopropanated mycolates that corresponded to hydroxy, keto, and

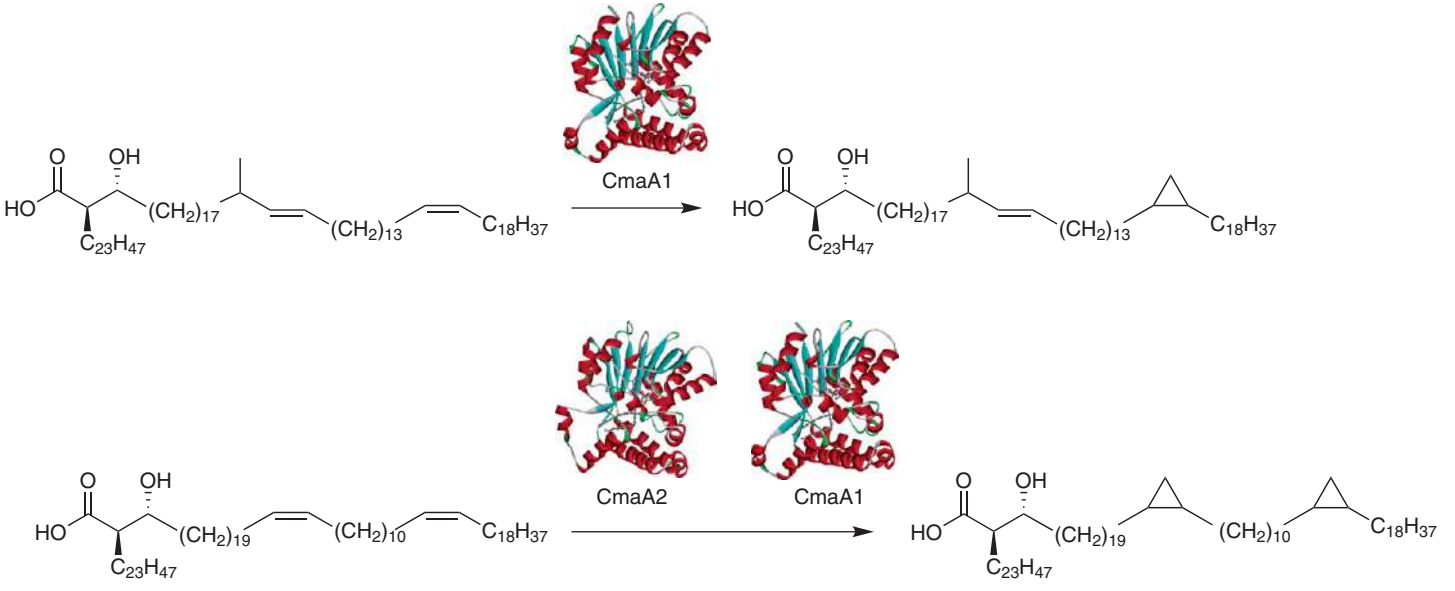


Figure 4 CmaA1 and CmaA2 are responsible for the cyclopropanation of mycolic acid alkenes; includes renderings of PDB 1KPG and 1KPI.

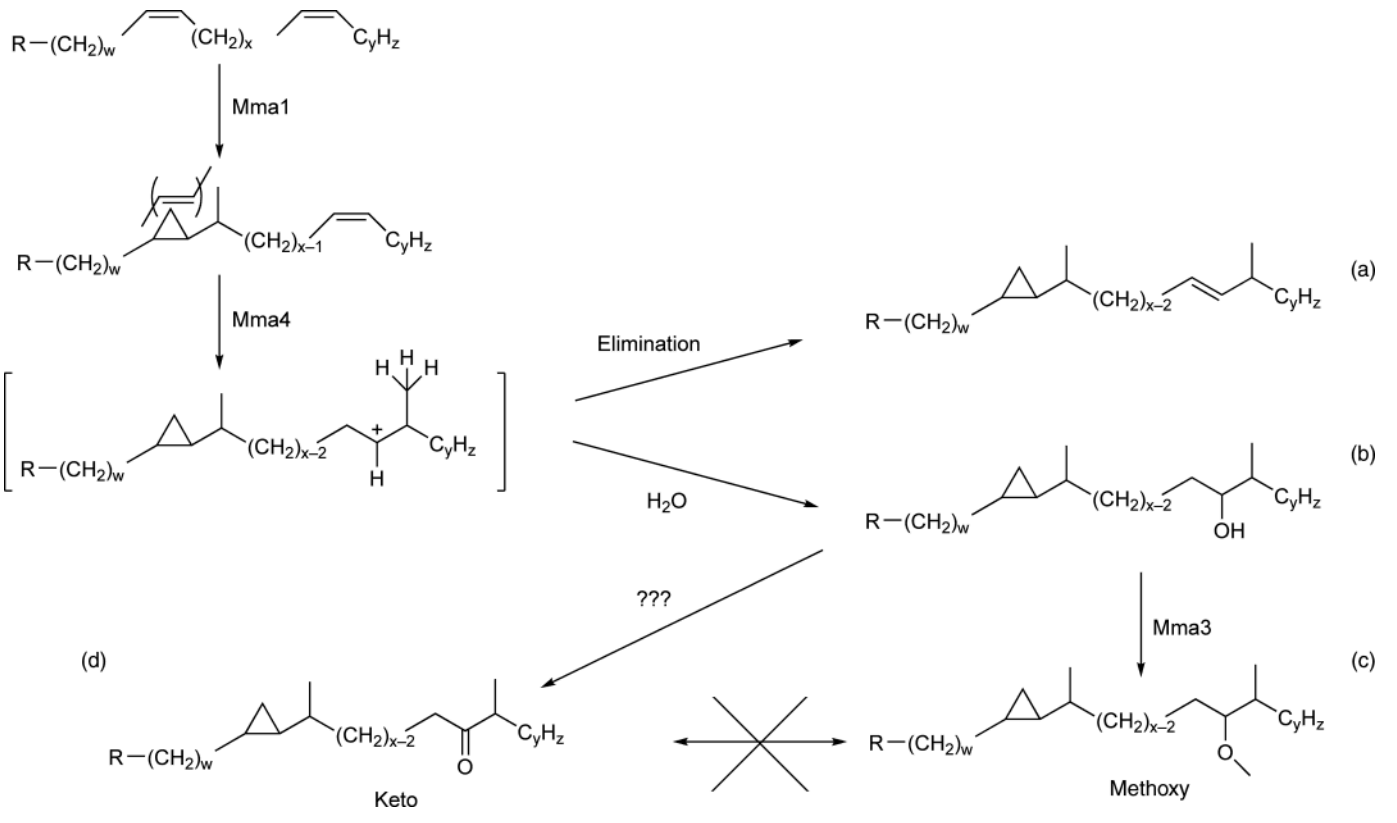


Figure 5 Synthesis of the virulence-associated oxygenated mycolates of *Mtb*. Mma1 transforms *cis*- to the proximal *trans*-alkene/cyclopropane. Methylation of the distal site by Mma4 produces a common intermediate that undergoes elimination to form (a), or is trapped with H_2O to form the hydroxy (b). B can be methylated by Mma3 to provide (c) or is processed by an unknown pathway to give (d).

methoxy forms. The chain lengths of keto and hydroxy forms were identical suggesting that the hydroxy-mycolate was a putative intermediate for the production of keto forms. Later studies demonstrated that *Mtb* also produced hydroxymycolic acid.[66] A deletion mutant in *Mtb*, Δ*cmaA* (*mma4*), did not produce oxygenated mycolates at all.[67] This mutant showed that the decrease in permeability to hydrophilic substrates was attenuated in the ability to cause infection in mice, confirming the results obtained from the heterologous expression studies. It has been proposed that the oxygenated mycolates, especially the keto forms, might provide hydrogen bond acceptor sites for the anchoring and transfer of polar material into the cell wall. Thus, the oxygenated mycolates might have roles in modulating cell wall fluidity and in the flow of hydrophilic substrates into the cell.

The study of the enzymes that modify mycolic acids was advanced again as better genetic tools for manipulating *Mtb* became available. Jacobs and coworkers employed newly designed transposon mutagenesis techniques to disrupt *Mtb* genes and screen a library for mutants deficient in the ability to form cords (a macroscopic property in which the bacteria arrange themselves in rope-like aggregates).[68] They isolated a mutant containing a transposon in a gene called *pcaA* (Rv0470c) with high homology to the previously described *cma1* and *cma2*. The gene *pcaA* was expressed at equivalent levels in all phases of growth, and its deletion from BCG or *Mtb* inhibited the proximal cyclopropane modification on α-mycolic acids, resulting in the synthesis of proximal, unsaturated α-mycolate with a distal cyclopropane (**Figure 6**(**a**)). This also caused a dramatic and unexplained increase in keto- and methoxy-mycolates. The Δ*pcaA* mutant has shown a lack of cording phenotype and an inability to sustain persistent infection. There are complex and intriguing sets of explanations contained in these reports.[68,69]

Notably, these results suggested a redundant biochemical function for the enzymes encoded by *pcaA* and *cmaA2* based on the observed outcome of overexpression of *cmaA2* in *M. smegmatis* and the observed function upon deletion of *pcaA* in *Mtb*. Disruption of *cmaA2* in *Mtb* resulted in the accumulation of unsaturated keto- and methoxy-mycolates, revealing its actual role as a *trans*-cyclopropane-synthesizing enzyme for the oxygenated mycolate series (**Figure 6(b)**). This result illustrates the confusing relationship between product structure and substrate structure from enzymes in this family. In fact, these results support the notion of a common cationic intermediate whose exact positioning at the active site (dependent on the substrate-binding pose) ultimately determines its chemical fate.[64]

MmaA2 is a *cis*-cyclopropane-synthesizing enzyme that acts on the distal end of mycolic acid of α-mycolates, an activity previously thought to be exclusive to CmaA1. Although CmaA1 also possesses this activity, apparently it is not active in *Mtb* since the *cmaA1* knockout does not exhibit an altered mycolate profile. Overexpression of *mmaA2* in *M. smegmatis* revealed that it can also add cyclopropanes to the proximal site. The Δ*mmaA2* revealed a deficiency in distal *cis*-cyclopropanes for α-mycolates and a deficiency in proximal *cis*-cycloproanes for methoxy-mycolates. Thus, Mma2 is the first enzyme found to cyclopropanate multiple classes of mycolates. These three genes (*pcaA*, *cmaA2*, and *mmaA2*) therefore appear to work on substrates that have already undergone distal modification (**Figure 6(c)**).[70]

A number of interesting conclusions emerged from these studies. It was evident that multiple duplications and divergent evolution of one set of genes occurred with minor changes in active site residues (or perhaps even residues far from the active site that alter the precise binding of individual substrates at the active site resulting in differences in the product formed). From this common intermediate, the cell regulates the functionalization of mycolic acids and the fluidity and permeability of its cell wall. In contrast to many biological systems where the substrate undergoes modification and optimization, the enzymes in this case have been optimized. There are eight homologous enzymes in this family in *Mtb* and our understanding of their roles and relationships is far from complete. The importance of these enzymes in producing the virulence-associated cyclopropanated, methoxy, and keto forms offered promise for their exploitation as drug targets. But before this can be fully exploited, the biological roles and interconnections require considerable clarification.

1.04.2.4 Structural Effects of Mycolic Acids on the Cell Envelope

It stands to reason that the *cis*-cyclopropane modifications are an adaptation in pathogenic strains, which replace the alkene to protect the molecule (and subsequently the bacterial surface) from oxidative damage. In early work, the distal cyclopropane was implicated in resistance to hydrogen peroxide by virtue of enhanced

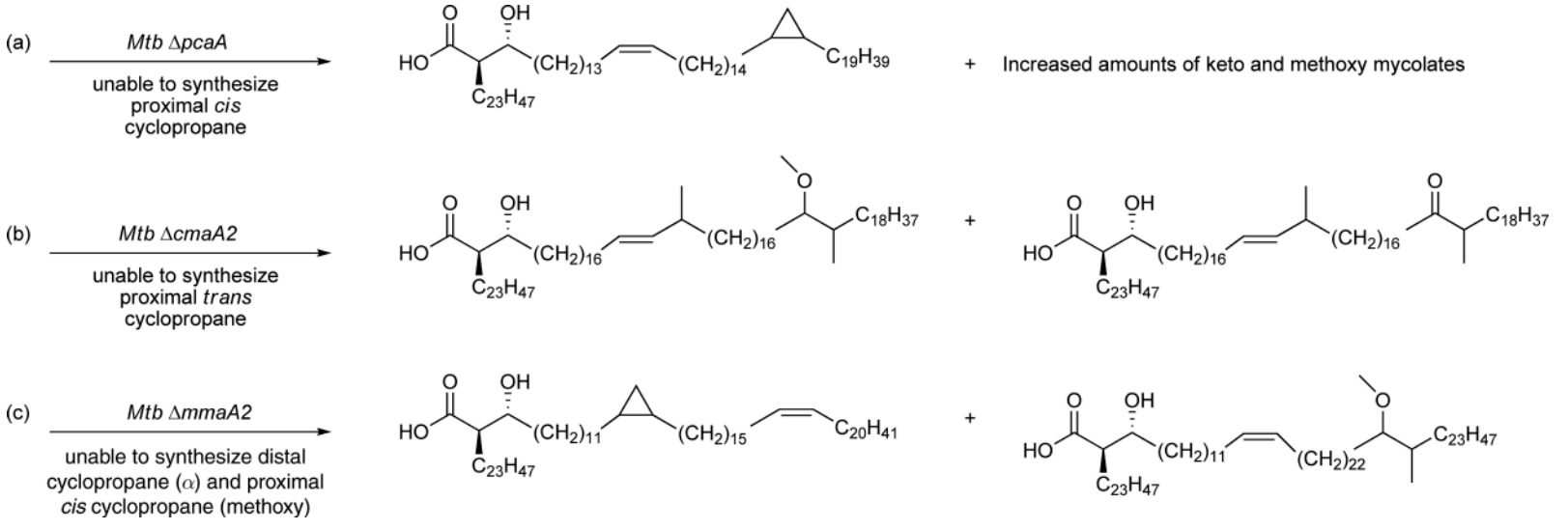


Figure 6 Altered mycolate profiles of (a) Δ*pcaA*, (b) Δ*cmaA2*, and (c) Δ*mmaA2* mycolic acid methyltransferase knockout mutants.

resistance to peroxide in *M. smegmatis* where the distal cyclopropane is abnormally installed.[57] However, an analogous role for the proximal cyclopropane was not apparent when *pcaA* was knocked out in virulent *Mtb*.[56,68]

The phase transition temperature of the *Mtb* cell wall is extremely high (60–70 °C). The cell wall has often been likened to a ball of wax, and such a strikingly high transition temperature implies that the cell wall is effectively a solid at room temperature, creating a formidable permeability barrier. Corynebacteria (C_{32}–C_{36} mycolic acids) exhibit much lower transition temperatures and examination of a diverse array of organisms illustrated that phase transition temperature is proportional to mycolic acid chain length. The introduction of cyclopropanes also has an effect on the transition temperature of the cell wall as measured by differential scanning calorimetry. Compared to the cis-cyclopropanation, trans-cyclopropanation does not induce a large kink into the mycolate chain. The *trans*-mycolates cause predictable decreases in membrane fluidity, as measured by increases in thermal transition temperature (a measure of the temperature at which a solid wax changes to liquid). In contrast, the *cis*-cyclopropanes cause a decrease in transition temperature and an increase in fluidity. For *Mtb*, only proximal *trans*-cyclopropanes are present in the oxygenated mycolate series. The *trans*-cyclopropanes and oxygenated forms are present only in virulent strains, most likely to control cell membrane fluidity during rapid temperature increases caused by host inflammatory responses. *Mtb* has a very specific temperature range in which it can grow. Thus, it is not surprising that in a number of species, including *M. smegmatis* and *M. avium*, increasing the temperature of a culture causes an increase in the amount of proximal *trans*-mycolates produced.[56,71] The exact locations of the double bonds are not usually determined, but it is reasonable to suggest that the number of alkenes controls the fluidity of the cell wall, and some mycolates contain up to three alkenes.[72] The cyclopropanated components of the lipids are particularly effective at providing extra membrane stability. Introduction of *Mtb*-type mycolates into the walls of *M. smegmatis* by cloning into that organism a gene from *Mtb*, which converts double bonds to cyclopropyl groups raises the 'melting point' by 3 °C.[56,68]

Overexpression of the protein MMAS-3 in BCG and *Mtb* converted all of the hydroxyl intermediates to methoxy-mycolates; as a consequence, no keto-mycolates were synthesized. These strains did not grow well in macrophages and were more permeable to some hydrophilic antibiotics, but less permeable to glucose. Keto-mycolates, therefore, are necessary for infection and intracellular survival, and the production and ratio of keto to methoxy can be modulated by control of the *mma3* gene.[63]

Interpretation of all of these phenotypic data are complicated by several factors: the impact of mycolate modification on the function of the porin MspA (as well as other transport systems that have yet to be identified); the ability of the host immune response to recognize the specific modifications of various forms; the impact of the modifications on residence time of loosely associated lipids that may also have immunologic consequences; and the variation in such modification seen between strain clades and even over time during infection. It is worth emphasizing that even a simple comparison of the fine structure of mycolates from 10 recent clinical isolates showed significant differences in the proportion of oxygenated and nonoxygenated mycolates as well as difference in the extent of cis- to trans-cyclopropanation.[62,73] Given the strong phenotypes observed when these are experimentally manipulated, these results suggest that such variation may have important clinical implications.

1.04.2.4.1 Thiacetazone and its impact on cyclopropane synthesis

Thiacetazone is an inexpensive antitubercular antibiotic that was formerly used in combination with isoniazid in resource-poor settings in Africa and South America (**Table 1**). It is a prodrug, a bacteriostatic agent, and its mechanism of action involves inhibition of mycolic acid cyclopropane synthesis. Using *Mycobacterium marinum* and *M. bovis* BCG as models, treatment with thiacetazone caused decreases in the cyclopropanation of cell wall mycolates as measured by traditional TLC methods (argentation and normal silica gel chromatography) and a high-resolution magic angle spinning experiment of whole cells. The isolation of total lipids revealed two prominent components: a dienoic α-mycolate and an unsaturated keto-mycolate. The precise cyclopropane synthase being inhibited is not clear, but the overexpression of PcaA, CmaA2, and MmaA2 independently in either *M. marinum* or *M. bovis* BCG diminished the effects of thiacetazone treatment. A series of thiacetazone analogues including one that proved to be slightly more active than thiacetazone (minimum inhibitory concentration (MIC) 0.05 $\mu g\,ml^{-1}$ vs 0.1 $\mu g\,ml^{-1}$) were also tested in this study.[38]

In another study aimed at locating the target of thiacetazone, spontaneous mutants to thiacetazone in BCG contained a mutation in *mmaA4*. These mutants could no longer synthesize oxygenated mycolates and were rendered resistant to thiacetazone. Complementation and overexpression of *mmaA4* restored thiacetazone susceptibility. It is suggested that although thiacetazone did not inhibit *mmaA4*, the gene promotes its activity in some way possibly by facilitating its uptake.[74]

1.04.2.5 Immunopathogenic Effects of Alterations in Mycolic Acid Structure

A thoughtful discussion of the immunologic effects of alterations in mycolic acid structure is impossible without considering the additional complexity of the two major locales of these molecules. In one locale (the cell envelope), these molecules are covalently anchored to the bacterial calyx and provide the milieu through which molecules must transit on their way into or out of the periplasmic space, as well as the environment within which proteins such as the porin must function. In addition, mycolates are found covalently attached to trehalose in the form of trehalose dimycolate (TDM), a molecule that is actively secreted and has almost magical abilities to induce pathology and directly interact with the host immune response. It is impossible to know *a priori*, particularly in complex models of disease, which niche contributes to the biological readout in a given experiment since both variables are changed simultaneously. As an example, a change in proportion of keto-mycolates in the envelope might inhibit the secretion of a factor necessary for downregulating the immune response of the host. The result of the biological assay would suggest that keto-mycolates were proinflammatory, or if examined in a murine infection system, the mice might die earlier and keto-mycolates would be dubbed 'virulence factors'.[72] Many such phenotypes have been reported and these are summarized in **Table 2**.

Because of this complexity, it should be considered very speculative to interpret the results obtained from whole organisms with genetically altered mycolate structures. Therefore, in this section, the observed phenotypes of altering the mycolate subclasses that have been reported without expounding on the possible mechanisms involved are just summarized. In the section on TDMs, the complexities of these interactions further from the perspective of understanding the nuances of the interactions of purified TDM with the host immune system, removing the second variable involving the complex host system are explored.

1.04.2.5.1 Variability of mycolic acid structures in the BCG vaccine

To illustrate the complexities introduced by the variation in mycolic acid structure, consider the major vaccine strain used to immunize for TB. *Mycobacterium bovis* BCG is an attenuated form of *M. bovis* currently used as a vaccine for *Mtb* throughout the world with the exception of the United States. The initial strains were prepared by serial culture, *in vitro*, to a point where the bacterium ceased to cause active infection in animals. There is much controversy concerning the effectiveness of BCG vaccination programs because of conflicting results in large-scale clinical trials. It is effective at controlling TB meningitis and disseminated forms of TB in children, but its utility in preventing pulmonary TB in adults is questionable. This renders the vaccine of little value in endemic areas. A major confounding issue is that many of the strains have been cultivated by serial passage, and might be overattenuated and incapable of producing adequate immunity. Many strains of BCG contain a mutation in *mma3*, which renders these strains unable to convert hydroxy-mycolates to the methoxy forms, and as a result only the keto- and α-mycolates are produced. The lack of methoxy-mycolates may play a role in altering the host immune response; therefore, this variation in mycolic acid structure could have had an influence on the clinical trials depending on the exact characteristics of the strain used for vaccination. Careful tracing of this mutation allowed reconstruction of the various existing vaccine strains phylogeny, and this has been substantiated by more extensive genomic studies with the conclusion being that the early methoxy-mycolate-producing strains may have been much more protective than the strains in current usage globally.[75] However, methoxy-mycolates are not the only difference between BCG and pathogenic strains.[76]

Table 2 Phenotypic consequences of alteration in mycolic acid fine structure

Alteration	*Phenotype*	*Reference*
MmaA1 overexpression	Increase in ratio of keto to methoxy Increase in proximal *trans*-mycolate Increase in cell wall fluidity and permeability	62
MmaA3 overexpression	Increase in methoxy-mycolates Keto-mycolates absent Impaired growth in macrophage	63
cmaA1, *cmaA2* (in *Mycobacterium smegmatis*)	Increase in cyclopropanated mycolates Increase in cell wall transition temperature by 3 °C	56
cmaA1 (in *M. smegmatis*)	Cyclopropanates distal site Resistance to H_2O_2	57
Δ*mmaA4*	Proximal *trans*-cyclopropane absent Oxygenated mycolates absent Decreased permeability Attenuated in mice	67
	Deficient in oxygenated mycolates Increase in IL-12p40, TNF-α Attenuated in mice TDM is hyperinflammatory	77
Δ*pcaA*	Lack of cording colony morphology Increase in proximal unsaturated α-mycolic acid Initial infection unaffected Unable to persist	68
	Decrease in methoxy increase in keto Decrease in TNF-α, IL-6 (hypoinflammatory)	69
cmaA2	Increases in IFN-γ, TNF-α, and granuloma size	78

1.04.3 Trehalose-Based Glycolipids

1.04.3.1 Overview

Trehalose monomycolate and dimycolate (TMM and TDM), the mycolic acid esters of trehalose, are widely distributed among both pathogenic and nonpathogenic strains of the mycolata, notably in the acid-fast mycobacteria. These glycolipids are the most abundant, noncovalently associated cell wall lipids, which provide these organisms with a waxy, impermeable outer membrane and the characteristic of acid-fastness. TMM and TDM serve a variety of purposes during infection and are purportedly secreted during infection of macrophages. TDM and TMM are thought to be oriented with their mycolates toward the cell cytoplasm, presenting the disaccharide moiety to the external environment of the bacterium as shown in **Figure 7**. Because of its importance to disease progression, some of characteristics of TDM, focusing on its discovery, characterization, biosynthesis, and immunomodulatory behavior as well as some of the implications of TDM research in tuberculosis antibiotic and vaccine development are highlighted.

1.04.3.2 Structural Features of TDM

TDM was coined 'cord factor' by Hubert Bloch who observed that pathogenic strains of *Mtb* grew in tightly wound bundles of snake-like cords. The crude, extractable, noncovalently cell wall-associated lipids that bestow this property were characterized by hydrolytic degradation, acetylation, reductive cleavage, and combustion analysis, yielding the structure 6,6′-dimycolyl α,α'-D-trehalose as shown in **Figure 8**.[79–81]

The carbohydrate moiety of TDM and TMM is trehalose, a nonreducing dissacharide, consisting of two glucose molecules with an α,α' linkage. TDM is symmetrically modified on the 6,6′ hydroxyl groups with mycolic acid esters. Trehalose is an important sugar for many of *Mtb*'s glycolipids and is also found in sulfolipid-1 (SL-1), diacyl trehalose (DAT), and triacyl trehalose (TAT) (see below). To analyze TDM and other cell wall-associated lipids, cells are harvested, dissolved in methanol and aqueous brine, and sequentially

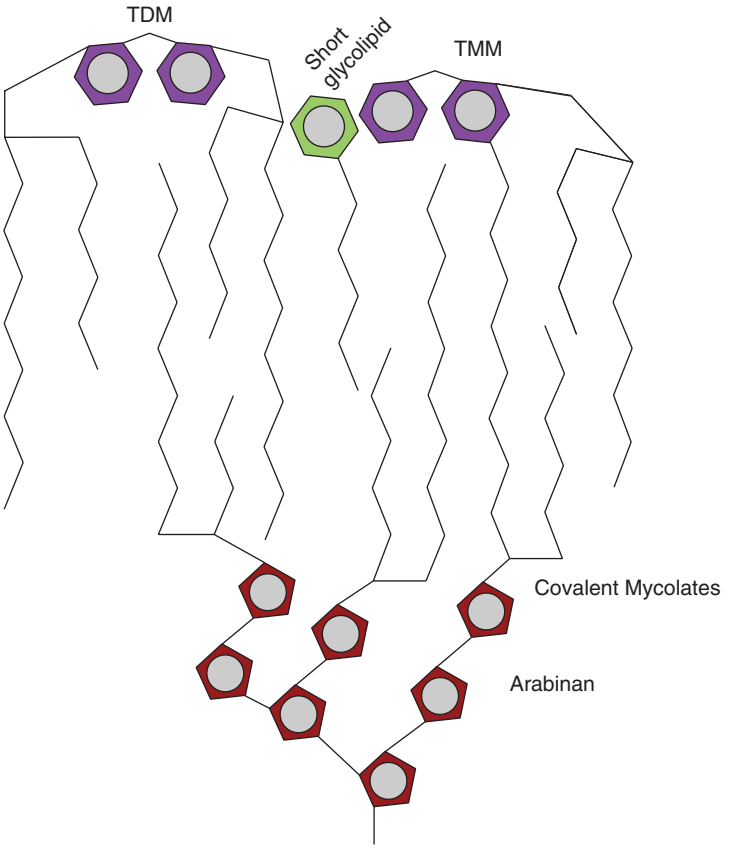


Figure 7 TDM is intercalated with the covalently attached mycolates of the mycobacterial cell wall and its hydrophilic trehalose exposed to the external environment.

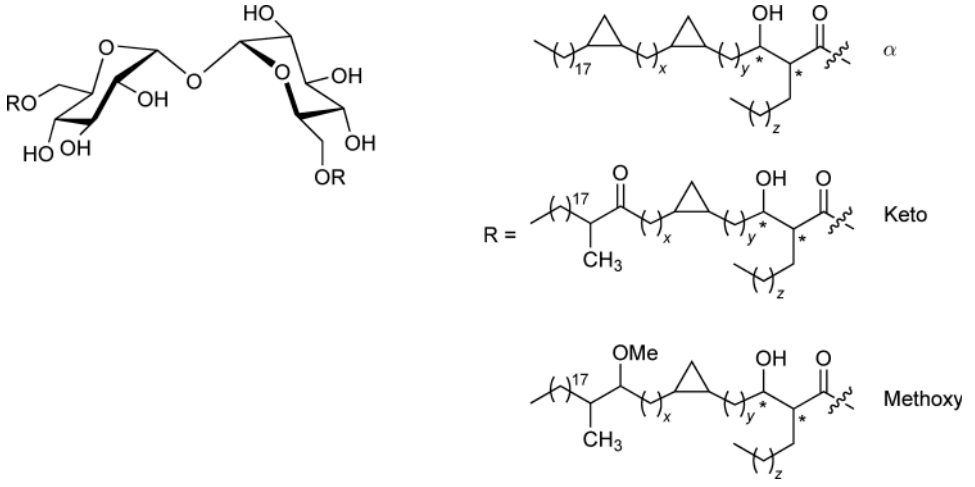


Figure 8 The general structure of 6,6′dimycolyl-α,α'-D-trehalose (TDM). The trehalose moiety can be coupled to α-, keto-, or methoxy-mycolic acids. $x = 10–20$; $y = 17–20$; $z = 25–30$; * indicates *R*-stereochemistry.

extracted with petroleum ether to remove apolar lipids (trehalose mycolates, phthiocerol dimycocerosate (PDIM), PGL, and acyl-trehalose). Chloroform:methanol (2:1) is then used to remove the relatively polar lipids (phosphatidyl-*myo*-inositol mannoside (PIM), lipoarabinomannan (LAM), and GPL). The TDMs contain a diverse mixture of mycolic acid components that can be separated from other cell wall components by preparative TLC and then saponified to remove the trehalose. The remaining fatty acids are then methylated to provide the carboxylated methyl esters. As an alternative to saponification, MALDI-TOF MS has proven useful for characterizing heterogenous mixtures of TDM[19,21] as whole glycolipids. Prior to MALDI, analysis by electron ionization (EI)/MS and fast atom bombardment (FAB)/MS[82] and GC/MS proved helpful for characterization. However, these methods were insufficient for determining the mass of intact TDMs, which is typically greater than 2.6 kDa.

1.04.3.2.1 Synthesis

Natural TDMs contain an inseparable mixture of mycolic acids of varying lengths and degrees of cyclopropanation. Because of this heterogeneity, synthetic methods have been used to obtain pure samples of some nonmycobacterial TDMs. *Mtb* TDM has not been synthesized through fully chemical methods because the chiral cyclopropanated and oxygenated mycolates are difficult to access. *Mtb* TDMs have not been reproduced through fully synthetic processes, but have been prepared from natural extracts by semisynthesis.[83] TDMs from related bacteria[84] and various analogues of TMM and TDM[85–87] have been prepared by total synthesis. Pure, synthetic TDMs are useful tools to probe structure–function relationships and have also been used as adjuvants for the development of vaccines against *Mtb*.[85,88,89]

Semisynthetic strategies capitalize on the ability to purify native mycolic acids from *Mtb* and then couple them to benzyl- or silyl-protected trehalose by common esterification techniques.[88,90] The mycolyl β-hydroxyl is an additional complication and must be protected to prevent dehydration. The 2,2′-, 3,3′-, and 4,4′-protected trehaloses can be obtained by utilizing either selective silyl protection[91] or selective mesylation.[92] Once the 6- and 6′-hydroxyl groups are freed, a variety of esters can be coupled to form TDM analogues (**Figure 9**).

Synthesis has allowed researchers to probe the important features of TDM. The 6,6′-trehalose diester of 2-eicosyl-3-hydroxy-tetracosanoic (behenylbehenic) acid, which has a similar structure to TDM, has been produced semisynthetically by Toubiana *et al.*,[83] as well as the straight-chain TDM analogue. The 2*R*,3*R* stereochemistry of the mycolic acids in the trehalose dicorynomycolates (TDCMs) of corynebacteria can be manipulated chemically and all four stereoisomers of TDCM have immunoadjuvant activity. Natural *R,R* TDCM and the synthetic *S,S* TDCM showed the most significant activities.[88]

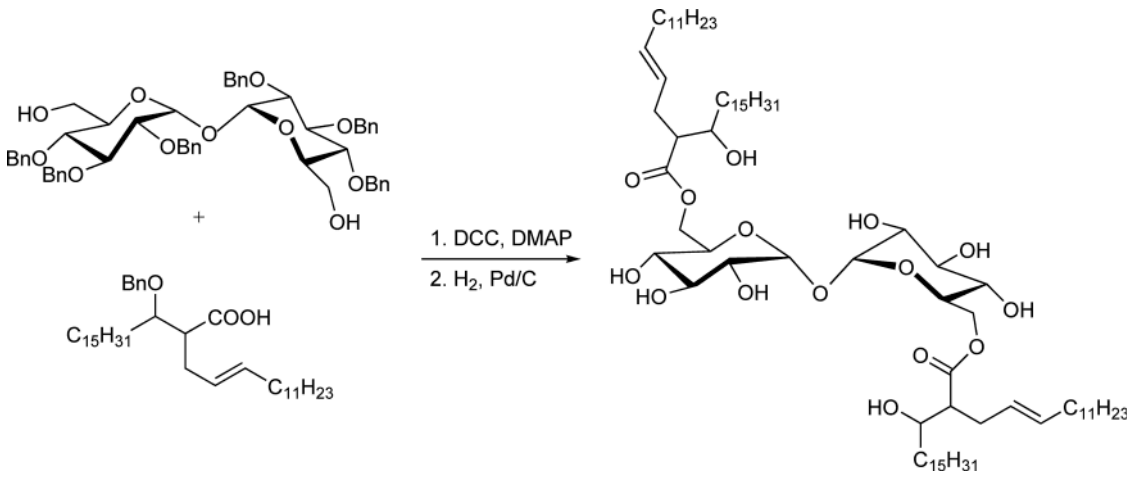


Figure 9 Synthesis of a corynebacteria TDM analogue.[88] The lipid was synthesized starting from a simple expoxide. The final DCC coupling and deprotection yielded the native TDM as well as three stereoisomers.

1.04.3.2.2 Biophysical studies of TDM properties

A peculiar feature of TDM is its ability to play dichotomous roles in a biological milieu. These are the result of its two aggregate forms: the first form consists of micelles with trehalose on the surface that are nontoxic, even at high concentrations whereas the second form consists of surface-adsorbed monolayers of TDM that expose their hydrophobic tails to the external environment and exhibit extreme toxicity to cells. These structures have been confirmed by scanning tunneling microscopy (STM), and have been linked to the cord-like growth patterns of mycobacteria.[93]

1.04.3.3 Biosynthesis

Trehalose is synthesized by *Mtb* through three independent pathways:[94,95] from UDP-glucose and glucose-6-phosphate through the OtsA and OtsB pathway (trehalose-6-phosphate synthase and phosphatase); from interconversion of maltose;[96] or from degradation of glycogen.[97] Of these three pathways, only OtsA/OtsB appears to be essential for trehalose biosynthesis.[98] A triple knockout mutant of *M. smegmatis* for all three trehalose synthesis pathways could survive without supplementary trehalose.[99]

In the biosynthetic model based on studies by Takayama and coworkers (**Figure 10**),[26] a fully formed mycolic acid is transferred from the ACP domain of polyketide synthase 13 (pks13) to D-mannosyl-1-phosphoheptaprenol, yielding 6-*O*-mycolyl-β-D-mannosyl-1-phosphoheptaprenol (Myc-PL).[100] This step is catalyzed by the putative mycolyl transferase I. The prenol tail of the Myc-PL causes it to migrate and dock with the proposed ATP-binding cassette (ABC) transporter, after which a second reaction transfers Myc-PL, through putative mycolyl transferase II to trehalose-6-phosphate (synthesized by OtsA) to form TMM-phosphate. The membrane-associated phosphatase for trehalose-6-phosphate cleaves the phosphate and the TMM is then immediately exported through the ABC transporter. The specific role of this ABC transporter has not been validated, but current data corroborates this model well.[101,102]

This coupling constructs TMM and exports it to the cell wall. Counterintuitively, TDM is not produced by a second acylation with Myc-PL. Instead, three extracellular proteins, antigen 85A, 85B, and 85C (fibronectin-binding proteins (FBPs)), perform a transesterification reaction to generate one TDM molecule and one free trehalose out of two TMM units (**Figure 11**).

1.04.3.3.1 The role of antigen 85 in TDM biosynthesis

Antigen 85A, 85B, and 85C are the most abundant extracellular *Mtb* proteins, making up as much as 41% of the culture supernatant protein content.[103] They are expressed in a ratio of 3:2:1 (85B, 85A, and 85C).[104] Genetic

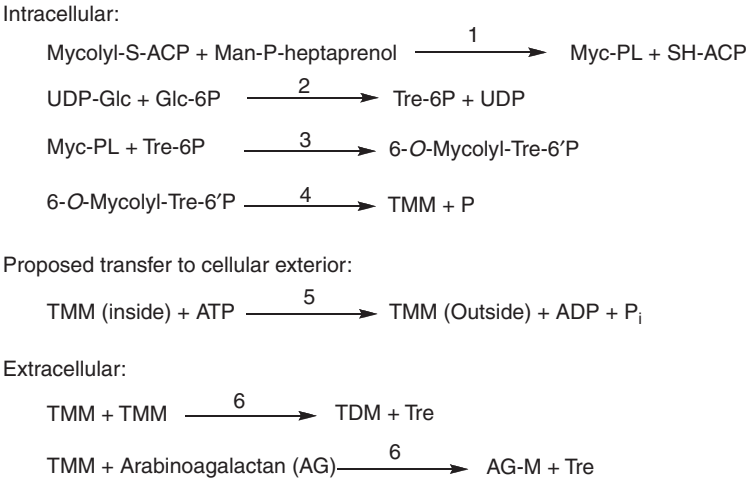


Figure 10 Proposed biosynthesis of TDM. (1) Putative mycolyl transferase I; (2) trehalose-6 phosphate synthase (OtsA); (3) putative mycolyl transferase II; (4) trehalose-6 phosphate phosphatase (OtsB); (5) proposed ABC transporter; and (6) antigens 85A, B, and C.

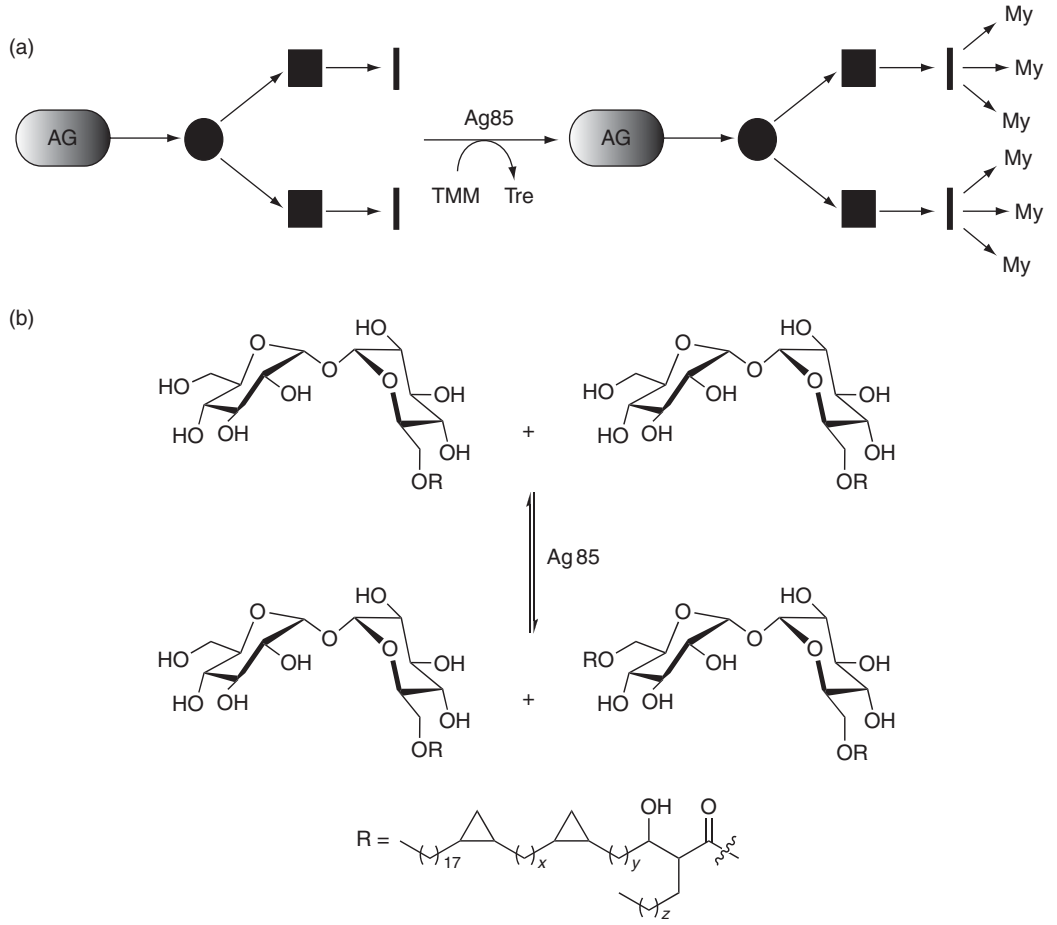


Figure 11 (a) Antigen 85-catalyzed mycolylation of arabinogalactan; AG = arabinogalactan, ● = 3,5-α-D-Araf, ■ = 2-α-D-Araf, My = mycolates. (b) Antigen 85-catalyzed transesterification; $x = 10–20$, $y = 17–20$, $z = 25–30$.

inactivation of 85C causes a 40% decrease in covalent mycolic acid content of the cell wall.[105] Multiple knockouts of 85A, 85B, and 85C have not been reported, presumably because they are not viable. The three antigen 85 *Mtb* proteins have high sequence homology to TMM transesterases from *M. leprae*, *M. bovis*, *M. avium*, and corynebacteria. Corynebacteria contain only one TMM transesterase called PS1. PS1 is the best studied of this class of enzymes and has provided much insight into the *Mtb* antigen 85 complex.

Antigen 85 can also work in reverse, acylating free trehalose at the expense of TDM to form two molecules of TMM. There is evidence that antigen 85 catalyzes transesterification between TMM and the arabinogalactan, suggesting a possible role in the anchoring of mycolates to the arabinogalactan.[106,107] Despite this, *Mtb* knockouts of 85A and 85B had normal mycolic acid content.

1.04.3.3.2 Crystal structure

The crystal structures of antigen 85A, 85B, and 85C have been solved (**Figure 12**).[108–110] These antigens share high sequence and structural homology, characterized by an α,β-hydrolase fold and a hydrophobic fibronectin-binding domain. These three structures support a transesterification mechanism analogous to the serine hydrolases, in which the formation of a covalent Ser-mycolic acid enzyme intermediate is followed by attack from the 6′-hydroxyl of the trehalose.[108] The active sites of 85A, 85B, and 85C are highly conserved, featuring a His–Asp–Ser catalytic triad, a hydrophobic tunnel for the mycolic acid, two trehalose-binding sites, and a key phenylalanine residue that is thought to regulate a hydrophobic channel between the two binding sites. This

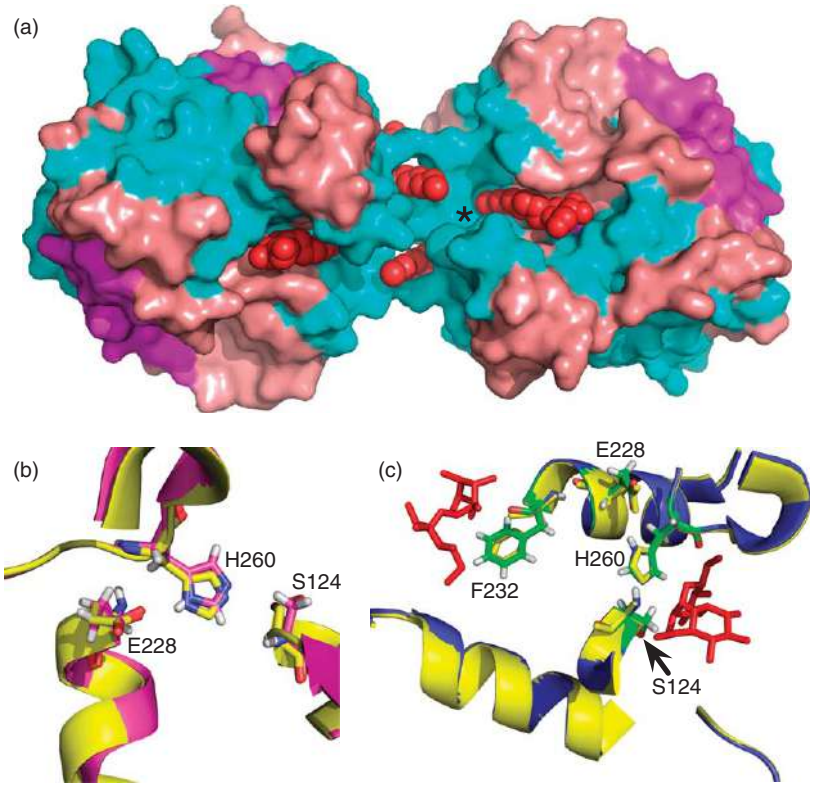


Figure 12 (a) Crystal structure of antigen 85C bound to *S*-octylglycose (PDB 1VAS). The structure shown is a dimer of two 85C proteins each bound to two substrates. The proposed hydrophobic channel, through which a second TMM could transverse the active site, is labeled with an asterisk (*). (b) Active site overlay of 85B and 85C shows very strong structural homology (PDB 1F0N and 1DQZ). (c) Multiple overlayed images of 85B bound to the trehalose active site shows very little motion. Phe230 moves slightly to make room for the second trehalose to enter the active site.

second channel may support a 'scooting mechanism', whereby the second TMM substrate can flip from its binding site to the active site without exposing the mycolate to the solvent.

1.04.3.3.3 Enzymology

These enzymes were first described in 1982 as a cell-free extract from *M. smegmatis* that was capable of synthesizing TDM.[86] This enzymatic activity was used as a basis for purification,[111] and demonstrated that TDM was derived from TMM. Antigen 85 activity is measured in a ^{14}C-trehalose exchange assay, where TDM is mixed with ^{14}C-trehalose and the enzyme as depicted in **Figure 13**.[112] Turnover is measured by scintillation counting of the lipid extract, monitoring for ^{14}C-trehalose incorporation into the lipids. Alternatively, autoradiographic TLC analysis can be used to calculate rough k_{cat} values. Using this assay, antigen 85C was found to be 8 times more active than 85A and 85B. However, the substrate specificity and a full kinetic analysis of these enzymes have not been explored.

High-throughput screening of inhibitors is not possible with the cumbersome ^{14}C-trehalose assay. Ronning and coworkers[113] have devised a coupled assay, utilizing a *p*-nitrophenol-glucose substrate, antigen 85, and

$$\text{TMM} + {}^{14}\text{C-trehalose} \underset{}{\overset{\text{Ag 85}}{\rightleftharpoons}} {}^{14}\text{C-TMM} + \text{trehalose}$$

$$^{14}\text{C-TMM} + \text{TMM} \overset{\text{Ag 85}}{\rightleftharpoons} {}^{14}\text{C-TDM} + \text{trehalose}$$

Figure 13 ^{14}C-Trehalose antigen 85 (Ag85) activity assay.

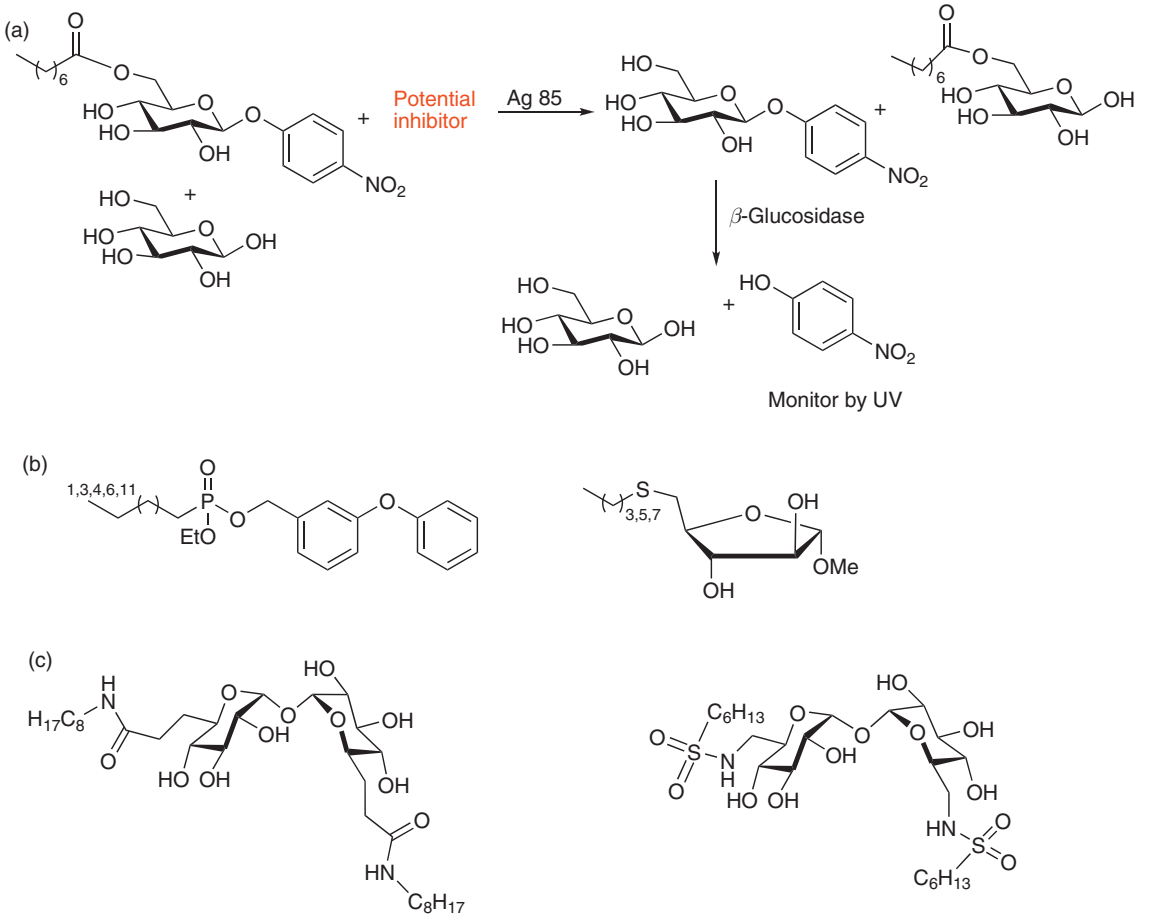


Figure 14 (a) A novel antigen 85-coupled activity assay based on the hydrolysis of *p*-nitrophenol-glucose by β-glucosidase. The release of *p*-nitrophenol can be monitored quantitatively by ultraviolet (UV) absorption. Representative work toward inhibitors of antigen 85 using nontrehalose-based (b) and trehalose-based (c) scaffolds.

glucosidase (**Figure 14**). This assay has potential for high-throughput screening applications and has helped to prove that antigen 85 transfers mycolates to arabinogalactan.[114]

1.04.3.3.4 Inhibitor development

Because of TDM's role in pathogenesis, inhibitors of the antigen 85 complex can provide leads for antitubercular antibiotics. Studies with antigen 85 mutants have suggested that inhibiting this protein would increase susceptibility of the bacterium to other drugs. Antigen 85A mutants displayed increased sensitivity to frontline TB drugs and other broad-spectrum antibiotics.[115]

The designed inhibitors of antigen 85 (**Figures 14(b) and 14(c)**) have utilized the trehalose scaffold[116–118] with modifications made to the reactive C-6 such as azides,[112] amines,[116] alkyl amides or sulfonamides,[117,118] and phosphonates.[119,120] These phosphonates and sulfonamides mimic the tetrahedral intermediate. The presence of larger lipophilic groups correlates with increased inhibition, but this causes problems with solubility in the assay.

Recent work of methyl 5-*S*-alkyl-5-thio-D-arabinofuranosides has suggested potential low inhibitory activity against antigen 85. Solubility issues with the compounds prevent direct assay against antigen 85. However, MIC, modeling, and indirect evidence corroborate inhibition.[121]

The antigen 85 crystal structures reveal two trehalose-binding sites separated by a hydrophobic channel. Anderson *et al.*[109] have proposed a novel series of inhibitors consisting of two trehalose molecules linked with a

hydrophobic linker. The fundamental problem in designing these inhibitors is the lack of an effective assay to directly measure antigen 85 inhibition. MIC and other growth inhibition assays are useful, but run the risk of highlighting off-target effects.

1.04.3.4 The Immunopathology of TDM

Granulomas induced by mycobacterial TDMs are of two types: foreign-body-derived (nonspecific immunity) and hypersensitivity-derived (mediated by T cell responses for specific antigens).[122] The foreign-body-derived response is a result of TDM's toxic properties, potentiated by host lipids and derived from TDM's ability to disrupt cell membranes. Upon inhalation of *Mtb*, the bacterium reaches the surface of the lung and releases TDM and other antigens that interact with host lipids, macrophages, and proximal tissue. TDM is toxic to these host cells. The aggregate debris does not clear well, possesses adhesive properties, and the resulting recruitment of more immune cells gives rise to the formation of a granuloma to enclose the debris. This takes the form of a lesion with a core of infected and necrotized cells and infiltration of macrophages, surrounded by damaged host tissue.

Once an infection site is nucleated, tissue necrosis begins and the bacteria replicate and become engulfed by macrophages. Once inside the macrophage, the molecule is no longer toxic. TDM is responsible for the bacterium's survival in the intact macrophage. This is likely due to its incorporation into subcellular membranes. The influence of TDM allows a stalemate to be reached between the pathogen and host defenses.

The foreign body granuloma type is remarkably similar to those produced by quartz (silicon dioxide). Silicon dioxide exists in several mineral forms, some of which are nontoxic and some of which are extremely toxic to cells and can even potentiate *Mtb* infection. The mechanism of this toxicity, observed toward macrophages, is a consequence of crystalline surfaces binding to, and subsequently rupturing cell membranes. TDM exerts toxicity similar to silicon dioxide particles in mouse lungs.[123]

Once the adaptive immune processes begin, subsequent exposure to antigen results in hypersensitivity, the result of a T cell reponse to the strongly antigenic components of *Mtb*. Macrophages can be activated by antigen and also by antigen-specific T cells, intensifying the response and the damage caused by it, probably a result of our evolved ability to defend against environmental mycobacteria.[124] *Mtb* creates a chronic inflammatory condition that is suppressed once the granuloma is formed, and reactivated when the granuloma reopens or with newly acquired infection and exposure to antigen.

The host's response and pathogenesis of *Mtb* infection require the involvement of innate and adaptive components of the immune response.[125] The innate response is characterized by the nonspecific interception of foreign material by complement proteins and phagocytotic cells. Phagocytes and antigen-presenting cells (APCs) capture foreign material, release the proinflammatory mediator tumor necrosis factor-α (TNF-α), and digest peptides and lipids. These digested materials are then presented through MHC II molecules for peptides, or cluster of differentiation 1 (CD1) molecules for lipids, to T cell receptors. Binding of naïve T cells to antigens stimulates T cell maturation and inception of the Th1 response, which includes secretion of interferon-γ (IFN-γ). The roles of TNF-α and IFN-γ involve complex signaling pathways that are beyond the scope of this review. Briefly, TNF-α is the global proinflammatory mediator secreted by macrophages upon encountering foreign particles. It is responsible for the initiation of the inflammatory response and further recruitment of immune cells. IFN-γ is secreted by T cells and is responsible for activating macrophages at the site of infection, stimulating the release of cytotoxins like nitric oxide (NO) to destroy foreign material, upregulating expression of MHC II and CD1 antigen-presenting molecules, and recruiting more cells to amplify the processes of adaptive immunity. Once begun, the actions of cytokines and chemokines direct lymphocytes, and initiate their differentiation into the mediators of cellular immunity. *Mtb* antigens result in the development of cell-mediated hypersensitivity and development of a necrotic lesion. This hypersensitivity (delayed-type hypersensitivity, DTH, type IV), the basis for purified protein-derivative (PPD) skin testing, causes a much more dramatic immune response upon secondary exposure to *Mtb* antigens. For the unfortunate host this has two results. The first is a chronic inflammatory condition because of the slow secretion of *Mtb* antigens from granulomatous lesions. The second is a greater propensity to form granulomas as a result of the dramatic, hypersensitive response that ensues when antigens are mechanically dispersed throughout the lung by coughing or obtained from other infected individuals.[124] TDM was shown to induce granulomas both in immunized mice (hypersensitivity) and in athymic and unimmunized mice (foreign body).[122] The degree of

response was much greater in immunized mice suggesting that both responses are active but that the hypersensitive response is much more important for maintenance of infection and disease pathology.

The immunopathology of *Mtb* infection is driven by TDM that can single-handedly induce granuloma formation without the assistance of protein antigens. The formation of a granuloma has ambiguous results. To the host's benefit, it helps to prevent the spread of infection; therefore, disseminated infection is rare and occurs mostly in HIV-infected and other immunocompromised individuals who lack a functioning adaptive immune system. To the host's disadvantage, the granuloma creates a walled-off, locally isolated environment that suppresses diagnostic markers, presents a barrier to diffusion of drugs, and prevents the immune system from clearing the entire infection. This provides a niche for the survival of latent, metabolically inactive bacteria. Subsequent rupture of these lesions provides an excellent means for the infection to spread to a new host. This represents a complex evolutionary strategy on the part of the mycobacterium that tends toward symbiosis in a healthy host while maintaining the potential for transmitting progeny onward. This phenomenon underlies the stable 'latent' infection present in fully one-third of the global population.

For the purpose of elucidating the cellular and immunological responses to TDM, the purified subclasses are typically emulsified in a solution of 0.2% Tween 80 and 3% Freund's incomplete adjuvant in phosphate-buffered saline. Lipids can then be assayed *in vitro* or *in vivo* as oil droplets in a water–oil–water emulsion.[126] Alternatively, lipids can be incorporated onto adsorbent microparticles[127–129] and then assayed so the lipids do not form aggregates or precipitate. Animals administered TDM in this way experience an enlargement in the size of the lungs, liver, and spleen. These organs harbor tissue-specific macrophages with measurably different reactions to the mycobacterial components. Much of the research in the past 10 years has focused on elucidation of the immunological responses to TDM, and determination of the subclasses of mycolic acid involved in virulence. *In vitro* and *in vivo* immunological assays of TDM have proven difficult because of several factors. It is difficult to draw conclusions from *in vitro* assays, such as those in cultured macrophages and T cells,[123,130] because they are done in specific cell lines and exclude other potentially important players. *In vivo* assays are complicated by the complexity of the host and the different pathologies that *Mtb* infection manifests in model organisms as compared to humans, particularly with respect to the granuloma.

1.04.3.4.1 The structure–activity relationship of granuloma formation

TDMs isolated from many different species have been examined to determine their potential to induce granulomas in mice. Such studies have shown that TDMs from pathogenic species have larger effects than those from saprophytic species.[131] The α-branch does not impart much diversity to the series, nor does it differentiate pathogenic and nonpathogenic strains. Conversely, the mero branch, its chain length, and its functionalization, distinguishes pathogenic strains and determines organ-specific responses.

Given the observation that TDM has context-specific effects that depend on the fine structure of the attached mycolic acids, it seems reasonable to question how well the response translates to different host animals. Although easiest to implement, the mouse does not provide an adequate model system for human tuberculosis as mice do not develop the caseous, necrotic lesions characteristic of a human granuloma.[132] There are significant differences between the human and murine immune systems because each has evolved to withstand different pathogens and environmental stimuli. Mice are not naturally infected with *Mtb* and instead suffer from infection with the 'vole bacillus' *Mycobacterium microti* with its own unique complement of mycolic acid-bearing TDMs. There is experimental evidence to support this contention.[133–135] In contrast to the mouse, tuberculosis infection of rabbits does result in the formation of cavitary lesions and a disease model more comparable to that of a human. Systemic injection of TDM in this model produces granulomas in the livers and lungs of rabbits, and significantly decreases spleen and thymus weight 7 days after administration. The histopathology is more pronounced in rabbits than in mice showcasing the greater utility of the rabbit as an experimental disease model.

In efforts to further define the microbial components responsible for granuloma formation, Sugawara *et al.*[136] examined the granuloma-generating ability of aerosolized glycolipid components in guinea pigs. Until this point, the biological role of TDM had only been investigated by systemic administration. Particularly, this experiment directly addressed the aerosol-transmission mechanism. Guinea pigs were subjected to aerosolized preparations of purified TDM, a heat-treated BCG Pasteur strain, an autoclaved BCG Pasteur strain, and methoxy-, keto-, or α-mycolyl methyl esters. TDM, the methyl ester of keto-mycolate, and heat-treated BCG

induced granuloma formation whereas the autoclaved BCG and methoxy- and α-mycolates did not. In all cases, the pulmonary granulomas contained epitheloid macrophages and lymphocytes (akin to a foreign body reaction) but lacked the central necrotized tissue present in human granulomas that is a consequence of the hypersensitivity reaction. Thus, the mycolic acid subclass, molecular composition of TDM, and its milieu affect the resulting toxic and granulomatogenic properties.

1.04.3.4.2 Complement binding

Complement proteins comprise a portion of the innate immune system in the serum, which binds to foreign organisms or antigenic debris. The proteins contain thioesters that are exposed upon binding to foreign particles. This can result in acylation and labeling of foreign material (opsonization) for engulfment by phagocytes through their complement receptors. Alternatively, the thioesters can hydrolyze, resulting in a cascade that activates more components of the complement system further resulting in the formation of a complex that attacks an invading cell's membrane.[125]

Mtb and *M. leprae* bind to complement C3. It is hypothesized that the binding of cell wall-associated TDM to complement enables the bacterium to enter phagocytes by a pathway that bypasses the oxidative burst normally accompanying phagocytosis. Whether or not TDM and TMM specifically bind to complement C3 is controversial with evidence both for[137] and against[138] this taking place. Other mycolyl glycosides (glucose, fructose, and mannose) do not bind C3.[137]

Activation of complement C3 by TDM can also lead to activation and cleavage of C5.[139] The products, C5a and C5b, are chemoattractants for leukocytes, and components of the membrane cleavage complex. Administration of TDM to C5 and C5a receptor-deficient mice results in a much greater, dysregulated inflammatory response. The interaction of TDM with the complement system suppresses the inflammatory response, potentially enabling *Mtb* cells to reach their intracellular home.

1.04.3.4.3 The 'cytokine storm' response to TDM

Mtb secretes several lipids that can modulate and induce production of cytokines by host immune cells. The cytokine response to TDM has been widely investigated.[140–142] Most importantly, TDM induces host cells to produce TNF-α, interleukin (IL)-12, and IFN-γ. These cytokines recruit T cells and initiate the process of granuloma formation. A complete discussion of these cytokines is beyond the scope of this text and much of it is still in debate. An attempt to briefly cite some key examples to illustrate how *Mtb* modulates host cytokine production will be made.

TDM plays a predominant role in inducing proinflammatory cytokine production. Delipidation of *Mtb* leads to reduced cytokine production, particularly IL-6, IL-12, and TNF-α, and decreased bacterial viability in the host.[127,143,144] Administration of TDM itself does not induce cell-mediated adaptive responses. TDM must be constantly regenerated to proceed to granuloma development. $CD4^+$ cells isolated from TDM-immunized mice produced significant amounts of IFN-γ and IL-2 when exposed to TDM-pulsed macrophages *in vitro*.

TNF-α is the global proinflammatory mediator secreted by macrophages upon encountering foreign particles. It is responsible for the initiation of the inflammatory response and further recruitment of immune cells, making it crucial for macrophage activation and granuloma formation. TNF-α levels strongly correlate with changes in organ index, a measure of granuloma development. Notably, resting macrophages fail to kill *Mtb*, whereas activated macrophages can inhibit the growth of the bacteria. TNF-α-deficient mice[145] did not experience any increase in lung index, indicating little to no granuloma formation and underscoring crucial ability of TNF-α to mediate granuloma formation. Furthermore, TNF-α plays a role in regulating chemokine expression in new and established granulomas.[146]

IFN-γ has been implicated as one of the most important cytokines in the immune response to *Mtb*. Individuals unable to produce IFN-γ have enhanced susceptibility to *Mtb*. IFN-γ is thought to be responsible for regulating the activation of resting macrophages, and stimulating antimycobacterial mechanisms. There is also significant evidence that IFN-γ downregulates the inflammatory cascade upon activation of macrophages. This leads to decreased production of inflammatory cytokines including IL-12, IL-18, and TNF-α. IFN-γ knockout mice experience a proinflammatory cascade and fatal pathology upon infection with *Mtb*.[147] These results suggest that IFN-γ plays an important role in stabilizing and sustaining the granuloma. Conversely, Takimoto *et al.*[148] proposed that initial granuloma formation is IFN-γ-independent, and instead proceeds through a TNF-α pathway.

1.04.3.4.4 NO production

Ultimately, production of TNF-α and IFN-γ by T cells and natural killer (NK) cells activates macrophages to produce NO. NO destroys phagocytosed material, has been implicated as a response mechanism to TDM exposure, and is a contributor to the tissue damage that is characteristic of the necrotic portions of granulomas.[140] The delay of phagosomal maturation by *Mtb* is reversed in the presence of NO, leading to increased bacterial killing.[149]

1.04.3.4.5 TDM arrests phagosome maturation

Mycobacteria arrest phagosome maturation at an early stage, preventing acidification. Delipidated mycobacteria cannot do this and have decreased survival in macrophages,[143] pointing to TDM as a potential culprit. After 24 h inside macrophages, *Mtb* bacilli contain up to 50% more TDM than at initial infection. Vacuoles containing the bacteria sustain many of the characteristics of the early endosomal system, and are not particularly hostile with respect to either pH or hydrolytic behavior. Modulation of the phagosome to arrest at pH 5.8 is thought to be mediated by bacterial glycolipids, such as TDM[144] and extracellular proteins; TDM is particularly crucial to this inhibition and the survival of bacteria in macrophages. In model systems, TDM inhibits fusion of phospholipid vesicles.[150,151] This inhibition of phagosomal maturation allows the bacteria to survive within host cells. Beads coated with TDM can also arrest the phagosomes in an early, nonacidified stage at pH 6.4.

1.04.3.4.6 Putting it all together – A model of granuloma formation

The inflammatory cascade initiated by TDM is known as the cytokine/chemokine storm. This frenzy of activity does not continue unabated during *Mtb* infection. Instead, IFN-γ, which is also produced during cellular immune response, downregulates the proinflammatory response. At this point, a stable granuloma has been established (**Figure 15**). There is a vast literature, dating back to the 1960s on how mycobacterial cell wall lipids induce granulomas.[152–159] Characteristics of a stable granuloma include 'foamy macrophages' and giant cells as well as regions of necrosis, the presence of T cells, B cells, vascularization, hypoxia, and extracellular mycobacteria trapped within the granuloma.[160,161]

1.04.3.4.7 Angiogenesis: VEGF formation and granuloma stability

Vascular endothelial growth factor (VEGF) regulates the process of neovascularization,[162] supports the development of inflammatory responses, and is required for the progression of chronic disease.[163] VEGF-induced vascularization is an essential part of the chronic inflammation that is characteristic of granulomas.[164] VEGF is produced by activated macrophages[165] and increased production has been linked to TDM. To confirm this, purified TDM from *Mtb* demonstrably induced vascularization when injected into rat corneas.[166]

1.04.3.4.8 Antitumor properties

TDM-induced antitumor activity has been observed since the 1970s.[83,167] TDM activates macrophages and the secretion of TNF-α, and NO is cytotoxic to cancer cells.[168] The recruitment of NK cells[169] may also contribute to the chemotherapeutic behavior of TDM. However, TDM is also extremely toxic and unselective, so synthetic work or nonmycobacterial TDMs might provide TDM analogues that have antitumor effects with reduced toxicity.[170]

1.04.3.4.9 Biological properties of antigen 85

The antigen 85 proteins invoke a diverse set of immunological responses, including humoral immunity, IgG, and IgA.[171] Antigen 85-derived peptides are presented by the MHC II complex and induce T cell differentiation and modulate the production of cytokine-producing T cells and the production of IFN-γ.[172] The redundancy of the three isoforms of the antigen 85 proteins may enable the bacterium to evade the immune system of the host. There is good evidence that the genes for these proteins are expressed under different conditions and that the bacterium adjusts antigen 85 expression in response to the environment.

The antigen 85 proteins also bind fibronectin. Crystallographic studies have suggested the presence of a fibronectin domain,[108] although the exact location of the domain remains unclear.[173] It has been demonstrated through binding studies that fibronectin binds to antigen 85 from BCG and that this binding is quite strong.[174,175] Thus, antigen 85 proteins could adhere to cell surfaces and be presented to T lymphocytes.

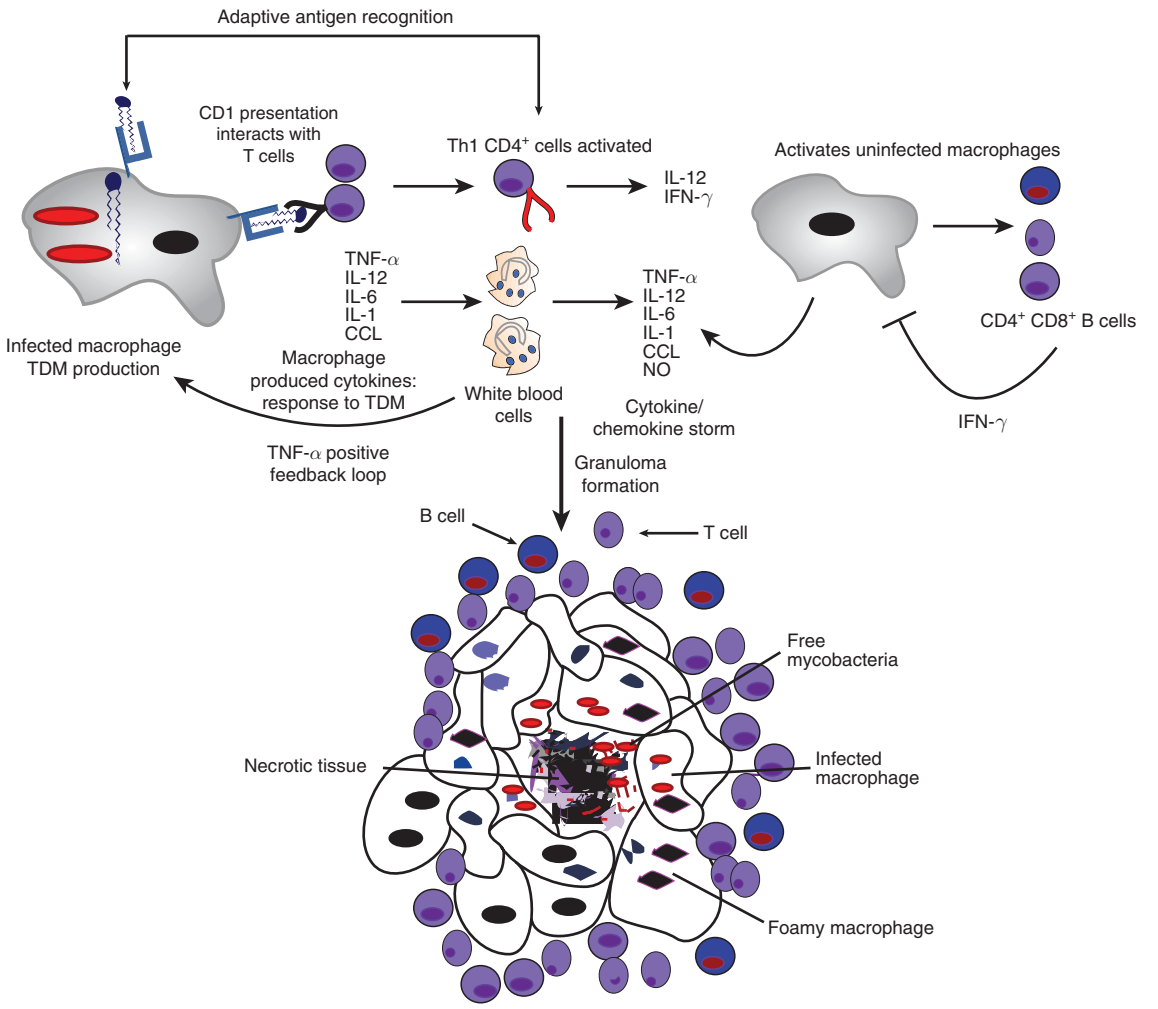


Figure 15 Granuloma formation as caused by TDM and related glycolipids. TDM induces infected macrophages to produce an array of cytokines, including TNF-α, IL-12, IL-1, CCL, and other chemokines (cytokine/chemokine storm). These cytokines and chemokines lead to the recruitment of lymphocytes, including NK cells and $CD4^+$ and $CD8^+$ T cells and B cells. Cytokines also activate uninfected macrophages, leading to expanded TNF-α production. Additionally, GMM, free mycolates, and SL-1 are presented through CD1 receptors to $CD4^+$ T cells, causing adaptive antigen recognition. Ultimately, IFN-γ is released, leading to a quieting effect on this storm and stabilized granuloma formation.

Pasula *et al.*[176] demonstrated that fibronectin facilitates attachment of *Mtb* to macrophages. Release of antigen 85 could also prevent fibronectin from binding to the bacteria, by binding the fibronectin to the antigen 85 before it reaches the bacteria. Binding to fibronectin could also encourage fibronectin-mediated phagocytosis.

1.04.3.4.10 Vaccine potential of TDM and antigen 85

TDM is intriguing as a potential vaccine against tuberculosis, alone or in combination with the BCG vaccine to improve its efficacy. TDM is a potent adjuvant, raising a strong response by both the cellular and humoral immune systems. Just recently, it was discovered that a synthetic analogue of glycerol monomycolate (MMG) with 32 carbons (C_{32}) exhibited comparable levels of immunostimulatory activities to the natural MMG.[177] This work might suggest a role for less toxic synthetic, mycolic acid-based adjuvants.

The antigen 85 proteins are also important candidates as additives to an *Mtb* vaccine. They are one of the major antigens in the humoral immune response to *Mtb*. As such, vaccination with these proteins, either alone or in conjunction with the BCG vaccine, has been proposed as a means to raise immunity against *Mtb*. Protective

immunity induced by vaccination with antigen 85 has even been demonstrated in the guinea pig model system.[178,179] A fusion protein of antigen 85B and ESAT-6 conferred a high level of protection against *Mtb* in the guinea pig aerosol model of infection.[180] ESAT-6 is a 6 kDa early-secreted antigenic target that is a common additive in *Mtb* vaccine development. Preliminary results in mice suggest that vaccination with the Hsp65 protein[181] or with the new BCG-type ESAT vaccine[182] are also possible routes toward an effective TB vaccine.

These and numerous other results suggest that including the antigen 85 proteins as well as TDM and other glycolipids into a vaccine mixture of killed *Mtb* can raise a stronger immune response than from killed bacteria alone. Vaccine trials are being conducted using BCG boosted with a modified vaccinia Ankara that expresses antigen 85 and ESAT-6.[183]

1.04.3.4.11 Humoral responses to trehalose dimycolates

Although the lipophilic nature of mycobacterial antigens like TDM and PDIM was thought to be incapable of recognition by the humoral immune system, antibodies to TDM and other lipids have been observed. Anti-TDM antibody has been observed in both active and latent cases of TB. The anti-TDM IgG antibodies produced in the sera of patients infected with either *Mtb* or *M. avium* were reliably specific for the TDM produced by each respective species, implying that the antibodies were specific for the lipid tail, and not the trehalose. Thus, they presumably differentiate the mycolic acid subclasses expressed by each species.[184,185]

Interestingly, the anti-TDM IgG reacted most strongly with methoxy-mycolates and much less with the keto- and α-mycolates. The antibodies were not cross-reactive to SL-1, which contains trehalose esterified to phthiocerol lipids, no cardiomycolates (C_{44}–C_{46}) or palmitic acid (C_{16}). This confirms that the mycolates are a specific antigen for the antibody in spite of their lipophilicity. These findings were confirmed in a rabbit and mouse model as well, with the additional evidence that monoclonal antibody to TDM suppressed granuloma formation in these animal models. In contrast to protein-based detection systems, the detection of anti-TDM IgG by enzyme-linked immunosorbent assay (ELISA) has been shown to be much more reliable and has the potential to differentiate smear positive and smear negative cases, since the antibody titer for specific lipids varies with disease severity. The titer of anti-TDM IgG has been shown to decrease after chemotherapy and return to normal after 3–4 months of therapy, correlating with the level of acid-fast bacilli present and giving a quantitative measure of disease state.[184,185]

1.04.3.5 CD1 – A Mammalian Lipid Recognition System

The discovery of CD1, a transmembrane glycoprotein, marked a major leap in the field of immune recognition. CD1 molecules present self-derived and microbial-derived glycolipids on the surfaces of APCs, such as peripheral blood monocytes (PBMCs), macrophages, dendritic cells (DCs), and B cells, to restricted T cells in the same fashion as major histocompatibility complex I (MHC I) molecules. These restricted T cells produce IFN-γ, Th1, and Th2 cytokines much like their MHC-restricted T cell counterparts. They also initiate cytotoxic effector mechanisms like the production of perforin and granulysin, which lyse infected APCs. There are two classes of CD1 in humans, Group I (CD1a, CD1b, and CD1c) and Group II (CD1d and CD1e), but only one type, Group II (CD1d) in mice. CD1 Group I binds exclusively to the T cell receptors of a specific subset of cells (called $CD4^-8^-$ $\alpha\beta^+$ or $\gamma\delta$ T cells), and Group II binds to natural killer T cells (NK T).[186,187]

The five human subtypes of CD1, (a–e), differ in structure, function, and the types of immune cells that express them. CD1a–d molecules bind to a β-microglobulin protein expressed on the cell surface in the same manner as MHC I. Each subtype is specific for the type of phagocytic compartment they interact with and for the types of lipid molecules to which they bind. It has been suggested that CD1e occupies an intermediate role as a lipid chaperone. In contrast to MHC or human leukocyte antigen (HLA) molecules, which display significant allelic diversity within their respective populations, the CD1 molecules display little variation among individuals, suggesting recent evolution. To date, CD1a, CD1b, and CD1c have been implicated in the presentation of different microbial-derived lipids. Although the prospect of using CD1-presented antigens for vaccines is enticing, it should be emphasized that CD1-restricted T cells make up a vanishingly small percentage of total lymphocytes. Thus, the true impact of CD1 presentation in controlling mycobacterial infections is unclear. The fact that the mammalian immune system possesses a unique method for surveying lipids is an intriguing development and offers insight into the mechanisms of cell-mediated immunity.[188,189]

1.04.3.5.1 CD1 subtypes and lipid partners

CD1a is expressed on the surface of Langerhans cells, a subset of DCs specific to the skin. Because *M. leprae* causes predominantly epidermal infections, the role of these APCs in lipid presentation was examined. CD1a does not present LAM, mycolic acid derivatives, or any previously examined mycobacterial antigens to restricted T cells. It was found to present *M. leprae*-derived MAGP to restricted T cells from leprosy patients. The precise molecule involved was suggested to be arabinoyl-mycolic acid. Of recent note, the CD1a molecules bind dideoxymycobactin (DDM), a precursor to the *Mtb* siderophore mycobactin, and present it to restricted T cells. The role of this interaction in host defense is not yet clear, and a more comprehensive discussion of mycobactins is covered in a later section.[190,191]

CD1b binds the hydrophobic portions of free mycolates, monoglycosylated mycolates, SL_{1278}, and diacyl glycerol-based phospholipids such as PIM and LAM.[192] An initial series of experiments was conducted with TDM and glucose monomycolate (GMM) from *Mycobacterium phlei* and *Rhodococcus equi*.[193] Their results indicated that lipids are processed by monocytes in the same way as peptides in that CD1b presentation requires phagocytosis followed by lipid processing in acidic endosomal compartments, association with a CD1 isoform, and translocation of these lipid–CD1 complexes across the APC cell membrane. Neutralizing the endosome with chloroquine or inhibiting membrane transfer by fixation with glutaraldehyde abolished T cell recognition. Interestingly, GMM and free mycolates underwent processing and presentation, but TDM did not. This might be due to its size and also because the endosomal compartments do not appear to digest glycolipids in the same manner as proteins. It might seem surprising that the most prevalent mycobacterial lipid is not presented by CD1 molecules. However, it is not always the case that the lipids need to be presented for restricted T cells to bind to CD1 and exert their effects. **Figure 16** illustrates the presentation of GMM to a CD1b-restricted T cell.[194]

Inquiries into the nature of lipid binding revealed CD1b's large hydrophobic groove that binds the lipid portion of a mycolic acid containing up to C_{80}. This presents the hydrophilic carbohydrate to the waiting T cell

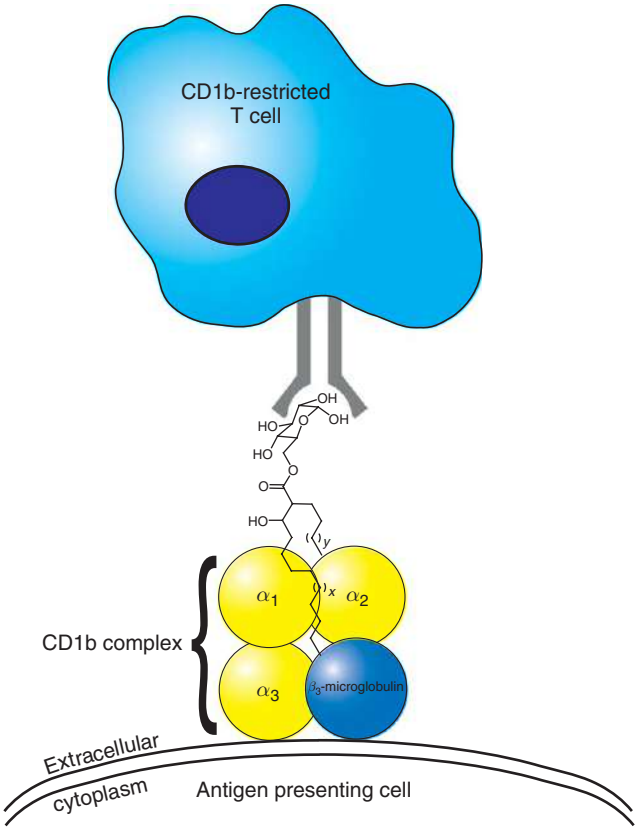


Figure 16 CD1b complex presenting GMM to a CD1b-restricted T cell.

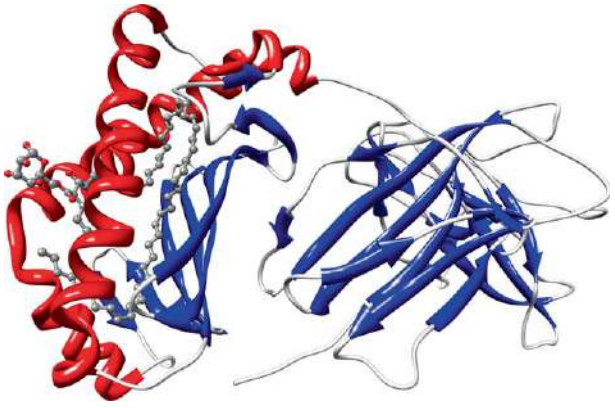


Figure 17 Crystal structure of CD1b bound to GMM (PDB 1UQS).[198] Rendered using Chimera.[199]

receptors. Alterations to the mycolic acid portion did not affect the binding of LDN5 T cells as the GMMs isolated from *M. bovis* BCG, *M. smegmatis*, *M. phlei*, and *Mycobacterium fortuitum* bound equally, as did a synthetic C_{32} GMM. This derivative lacked the chain length and pattern of cyclopropanation, unsaturation, and oxygenation of the virulent species. A crystal structure of CD1b bound to GMM from *Nocardia farcinica* (C_{58}) has been obtained and is shown in **Figure 17**. Using this structure, *Mtb* GMM and the various mycolate subclasses were modeled inside the binding groove. None of the lipid substituents appeared to effect lipid loading and could be accommodated without causing significant structural distortions in the binding site. More importantly, the subclass substituents on the lipids do not appear to affect the way in which the T cell receptors bind.[194]

In contrast to the hydrophobic tail, alterations to the β-hydroxy glycosyl ester had serious consequences for the LDN5 T cell receptor binding. Removal of the mycolic acid α-branch, the β-hydroxyl, or modifying the β-hydroxyl with tetrabutyl dimethyl silyl abolished LDN5 binding. LDN5 cells also did not bind to trehalose, glycerol, or arabinosyl monomycolates. Mannose and galactose monomycolates exhibited negligible binding, thus the T cell receptor is specific for one stereoisomer of glucose.[193] These results are of interest since *Mtb* GMM synthesis is temperature dependent. GMM is synthesized at 30 °C whereas TDM is synthesized at 37 °C. The lower temperature is more characteristic of a ventilated lung, and the production of GMM is stimulated by abundant, host-derived glucose. This would imply that CD1b recognition of GMM is a specific host response to the production of a virulence-associated glycolipid. Another recent report has suggested that rifampicin (RIF), a first-line tuberculosis antibiotic, increases CD1b expression on PBMCs.[195,196]

Diacyl sulfoglycolipid SL_{1278} is also a CD1b-presented antigen. T cells that recognize this compound secrete IFN-γ and can kill *Mtb*-infected macrophages. The sulfate residue and both acyl moieties are essential for this activity; its properties are discussed in detail in the Section 1.10.3.7 on SL-1.[197]

CD1c binds to a previously unidentified *Mtb* phosphoglycolipid called mycoketide. It is present in *Mtb* and *M. avium* and contains saturated alkyl chains with multiple methyl branches. In contrast to mycocerosic and phthioceranic acids, the methyl branches occur on every fifth carbon. Without prior knowledge of the biosynthetic apparatus, it was assumed that these lipids were derived from isoprenyl diphosphate, a five-carbon synthon. A noticeable flaw in this hypothesis was the presence of an additional, non-five-carbon multiple chain at the end. A recent study determined that these lipids are the products of a polyketide synthase (pks12)-mediated assembly of alternating two- and three-carbon fragments. These phosphoglycolipids are fully saturated analogues of the mannosyl-phosphopeptide fragment of Myc-PL, the hypothesized transfer agent of mycolic acid from the cytoplasm to the cell wall. These unsaturated lipids anchor membrane-bound precursors in peptidoglycan and arabinogalactan synthesis and are also unique among CD1-binding lipids in that they contain a single lipid chain. Their purpose as CD1c ligands is not known, but the binding of restricted T cells has proven specific for the presence of methyl branches with the *S*-configuration, the β-mannose-phosphate linkage, and the C_{30}–C_{34} chain length.[200–203]

The murine immune system possesses only one CD1 isoform, CD1d. Human CD1d presents mammalian glycosylphosphatidylinositols and α-galactosyl ceramides, a class of marine natural product-derived glycolipid.

The adjuvant properties of α-galactosyl ceramides have generated recent interest for their potential in anticancer therapeutics. Mouse CD1d is not homologous to human CD1b, the isoform that presents mycobacterial lipids and so its function in lipid presentation in the murine host is not analogous. A number of studies have pointed out the expression of CD1d and the importance of NK T cells in controlling mycobacterial infection,[204–206] but these do not necessarily reflect the absolute contribution of CD1. Several studies observed the effect of the absence of CD1d using anti-CD1 monoclonal antibodies or a $CD1d^{-/-}$ knockout mouse on the response to *Mtb* infection. The single role of CD1d in lipid presentation might not be clear, but the importance of CD1d in an immunomodulatory role was reinforced. $CD1d^{-/-}$ mice experience a more intense, unregulated inflammatory response. Systemic administration of TDM to the $CD1d^{-/-}$ knockout mouse generated a much more severe inflammatory response.[207] Activating NK T cells by alternative methods helps to protect mice against *Mtb* infection, and TDM upregulates the expression of both MHC II and CD1d on mouse macrophages.

1.04.3.6 Di-, Tri-, and Polyacyl Trehaloses

The interest in non-TDM cell wall-associated lipids focused on explaining the diversity, complexity, and biological functions of this nontrivial fraction cell that includes PDIM, PGL, SL-1, and related variants: DAT, TAT, and polyacyl trehalose (PAT).[208] Considering that GPLs and PDIMs are confined to certain species, especially virulent species, suggested a possible role for DAT, TAT, and PAT in modulating host defenses and bacterial survival. At the time of their discovery, the lack of modern genetic methods rendered functional studies inconclusive, and early efforts instead aimed at identifying species-specific signatures detectable by common immunological methods. Although PCR-based nucleotide amplification technologies for diagnosing infectious diseases are common in the developed world, the developing world lacks access to these systems in spite of a greater disease burden. There is an enormous demand for clinically adapted, serological methods. DAT, TAT, and SL-1 have all been investigated for this purpose because they contain smaller fatty acyl chains compared to mycolic acids, and they are bound to trehalose. These molecules are less toxic and contain a carbohydrate portion that can be used to generate lipid-specific antibodies for use in diagnostics.[209–211]

A major glycolipid isolated from the *Mtb* Canetti strain contains a trehalose core with five acyl substituents 2, 2′, 3′, 4, and 6′. Four of these substituents consist of multi-methyl-branched, C_{27} unsaturated phthienoic acids and an additional stearic acid. Minor amounts of the hydroxyphthienoic acid were also present. PATs generally contain dextrorotatory C_{14}–C_{28} fatty acids with one to three methyl branches. *Mycobacterium fortuitum* contains a series of DATs and TATs containing predominantly a monomethyl-branched C_{18} fatty acid (**Figure 18**).[212,213]

The biological role of DAT and PAT has been the subject of few reports. DATs from *M. fortuitum* have been shown to inhibit T cell proliferation.[214,215] Administration of purified DATs to PBMCs resulted in decreased secretion of IL-2, IL-12, and TNF-α in a dose-dependent manner. A transposon mutant in a polyketide synthase gene (*msl3*) cannot synthesize the dimethyl-branched (mycolipenic) and trimethyl-branched (mycosanoic) acids, which are incorporated into DAT and PAT.[216,217] The Δ*msl3* strain exhibited altered growth and released more protein and carbohydrate into the culture supernatant, suggesting a role for DAT and PAT in modulating the cell surface. The synthases corresponding to *msl* and SL-1 synthesis genes were recently shown to be positively regulated by PhoP, a two-component transcriptional regulator that may play a role in the response to nutrient starvation. A PhoP knockout mutant displayed dramatic attenuation in its ability to infect mice and in macrophage survival, most likely a consequence of its lack of DAT, PAT, and SL-1.[218]

1.04.3.7 Sulfolipids

The pioneering work of Gardner Middlebrook determined the presence of an *Mtb* cell wall-associated sulfolipid-1 (SL-1, **Figure 19**). SL-1 exhibits both acidic and lipophilic properties. SL-1 is produced in comparable quantity to TDM and it is present in the virulent strains *Mtb* H37Rv and *M. bovis* BCG, whereas it is noticeably absent from attenuated strains of *M. bovis* and *Mtb* H37Ra.[219] It is also absent from most strains of rhodococci and nocardiae[219] with few exceptions.[220]

The chemical structure of SL-1 deduced by Mayer Goren was 2,3,6,6′-tetraacyl-α,α-trehalose 2′-sulfate (**Figure 19**). It is slightly more polar than TDM and can be purified on silica gel. SL-1 contains a central trehalose with 6,6′-phthioceranic acid esters, a 2-stearic or palmitic acid ester, a 3-hydroxyphthioceranic acid

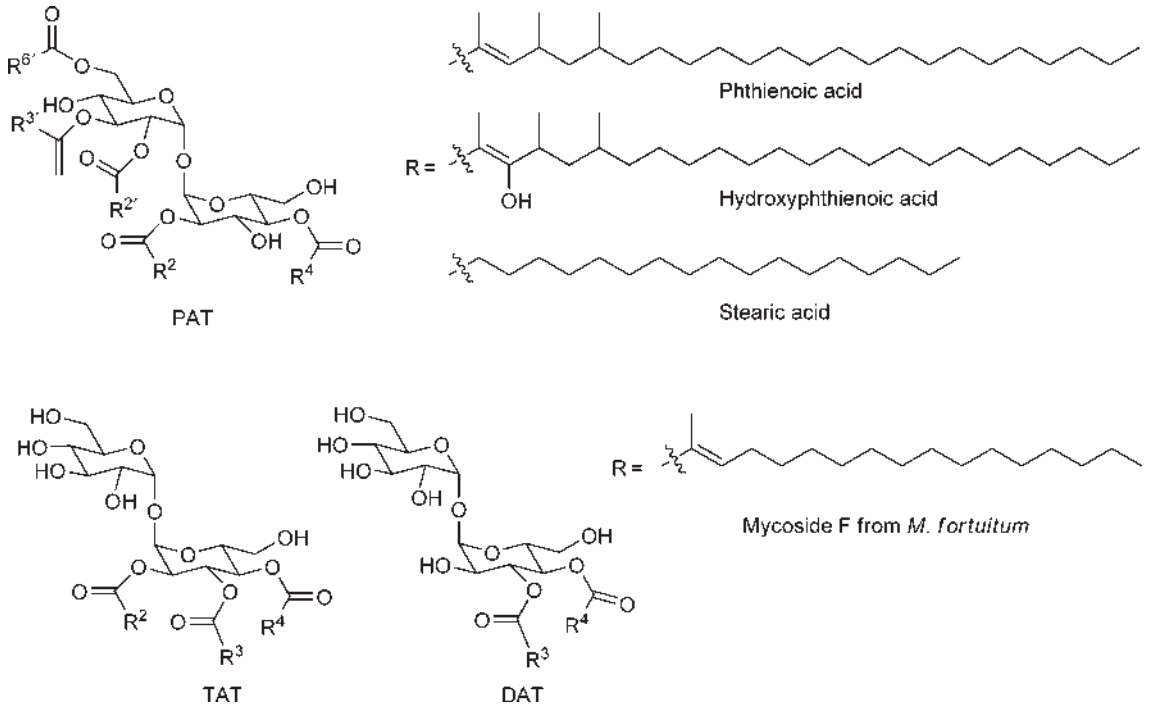


Figure 18 The structure of DAT, TAT, and PAT.

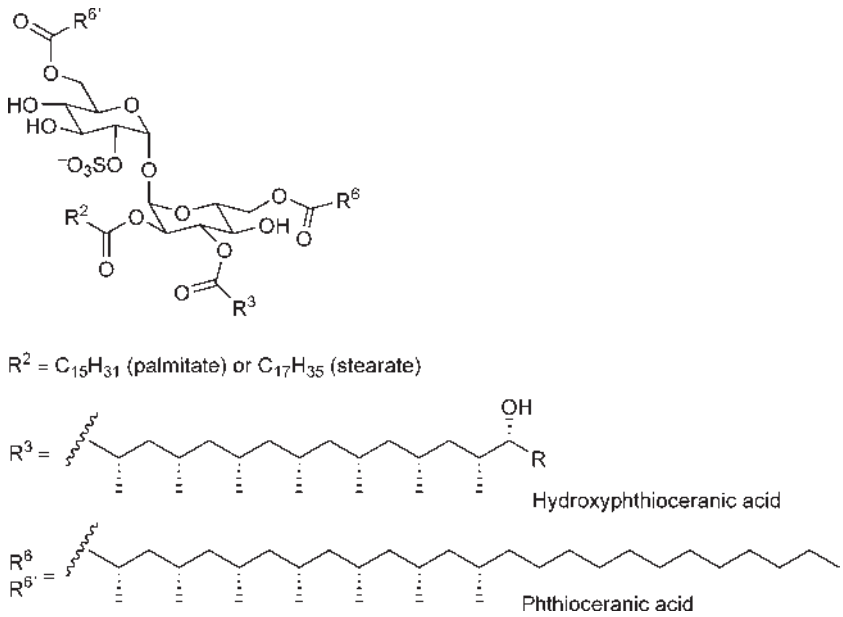


Figure 19 The structure of SL-1.

ester, and a 2′-sulfate. *Mtb* sulfolipid is a mixture of related compounds, in which SL-1 is the most prominent. The interest in the properties of SL-1 derived from its abundance among *Mtb*'s apolar cell wall-associated lipids. It has also been implicated in the prevention of phagosome–lysosome fusion, a mechanism for survival inside the macrophage.[221] This is a result of its ability to incorporate into, and affect the membrane properties of intracellular compartments. SL-1 bears some architectural similarity to TDM, but it does not contain mycolic

acids. Instead it contains the polyketide synthesis-derived hydroxyphthioceranic and phthioceranic acids that are similar to those contained in PDIM.

In contrast to TDM, SL-1 does not exhibit profound toxicity, nor does it exhibit granuloma-forming capabilities in lung, liver, or spleen. SL-1 does not induce significant production of proinflammatory cytokines, but can stimulate neutrophils and monocytes. SL-1 does not induce upregulation of MHC II or CD1d1 expression on murine macrophages or the secretion of IFN-γ that is customary with TDM, but its role in human adaptive immune responses is an area of recent concern.[204,222–224] Coadministration of SL-1 with TDM inhibits the inflammatory effects of TDM.[225] Recently, it has been proposed that the suppression of some of TDM's proinflammatory effects such as induction of TNF-α production leads to increased pathogen survival and successful infection.

1.04.3.7.1 Sulfolipid biosynthesis

Much of the work highlighted here focuses on delineating the biosynthetic machinery for SL-1 with the aim of determining its role in pathogenesis, immune modulation, and ultimately to survey these networks for potential antimicrobial drug targets. The biosynthetic machinery for SL-1 synthesis consists of a series of gene clusters: a sulfotransferase (*stf0*), polyketide synthases, acyl transferases (ATs) (*papA1* and *papA2*), and a series of membrane transporters (*mmpL*) that translocate these lipids across the membrane into the cell wall.

The first concerted step in SL-1 synthesis is the transfer of sulfate by Stf0 (Rv0295c) from 3′-phosphoadenosine-5′-phosphosulfate (PAPS) to trehalose to give trehalose-2-sulfate (**Figure 20**).[226] *Mtb* contains seven homologous genes annotated as mycocerosic acid synthase (*mas*) genes. A mutant in one, *pks2*, could not synthesize hepta- and octamethyl phthioceranic acids or hydroxyphthioceranic acids. TLC analysis of the apolar lipids from this mutant confirmed that it was lacking in SL-1. Interestingly, this did not affect the ability of this strain to grow in mice, nor did it abrogate PDIM synthesis, since the synthesis of PDIM's phthioceranic acid-derived lipids has been assigned to other synthases.[227,228]

Two independent groups have provided important evidence regarding the role of the enzymes involved in the early stages of sulfatide biogenesis.[229,230] These groups focused on two genes clustered with *pks2*, Rv3824c (*papA1*) and Rv3820c (*papA2*), and showed that these had AT activity. PapA2 catalyzes the 2-palmitoylation of trehalose-2-sulfate. PapA1 then acylates this product with hydroxyphthioceranic acid to provide SL_{1278}. Mutants in both genes (Δ*papA2* and Δ*papA1*) lacked SL-1 and SL_{1278}, but did not display any growth defects in a mouse model.

Transposon mutagenesis generated five mutants that displayed increased affinity for macrophages. This study was attempted since very little is known about pathogenic factors, which modulate binding and phagocytosis by host cells. Of the five mutants, one was in *pks6* (Rv0405) but the more prominent one was in *fadD23* (Rv3826), a gene of unknown function. A Δ*fadD23* mutant was unable to synthesize SL-1 suggesting its possible role as an AT in subsequent acylations of SL_{1278}.[231]

Another transposon mutagenesis screen determined that *mmpL8*, which encodes a lipid transporter, was essential for virulence. A Δ*mmpL8* mutant exhibited an increased amount of intracellular SL_{1278}, the precursor to SL-1. MmpL8 is responsible for modifying SL_{1278} and translocating it as intact SL-1 across the cell membrane.[232] The mechanism through which this occurs is unknown. The Δ*pks2* mutant, like the Δ*mmpL8* mutant, cannot synthesize intact SL-1, but was not attenuated in mice. This implied that SL-1 is not, in fact, a virulence determinant in the mouse, and that *mmpL8* possesses other more important functions.[233] This lack of attenuation was an intriguing finding considering that SL-1 is absent in these strains, and SL_{1278}, a potent immunogen by virtue of its interaction with human CD1b ligand, is produced in greater quantity. Follow-up studies on the role of SL-1 and SL_{1278} in human infections will be revealing. SL_{1278} might play a role in the human adaptive response that is simply not available to the mouse.[229,230] One such natural history clinical study is underway with patients in South Korea (NCT00341601, www.clinicaltrials.gov).

A number of past reports have described the immunomodulatory effects of purified SL-1, but the dissection of the molecular mechanisms behind this activity is not straightforward. Deleting parts of the biosynthetic machinery results in complete abrogation of SL-1 synthesis and transport. In other words, partially acylated structures cannot be prepared, translocated across the membrane, and tested *in vitro*. There are also some conflicting results. Aside from Δ*mmpL8*, none of the strains exhibit a decrease in virulence in mouse and guinea pig models in spite of their inability to produce SL-1. The strains are capable of growing and persisting in lung

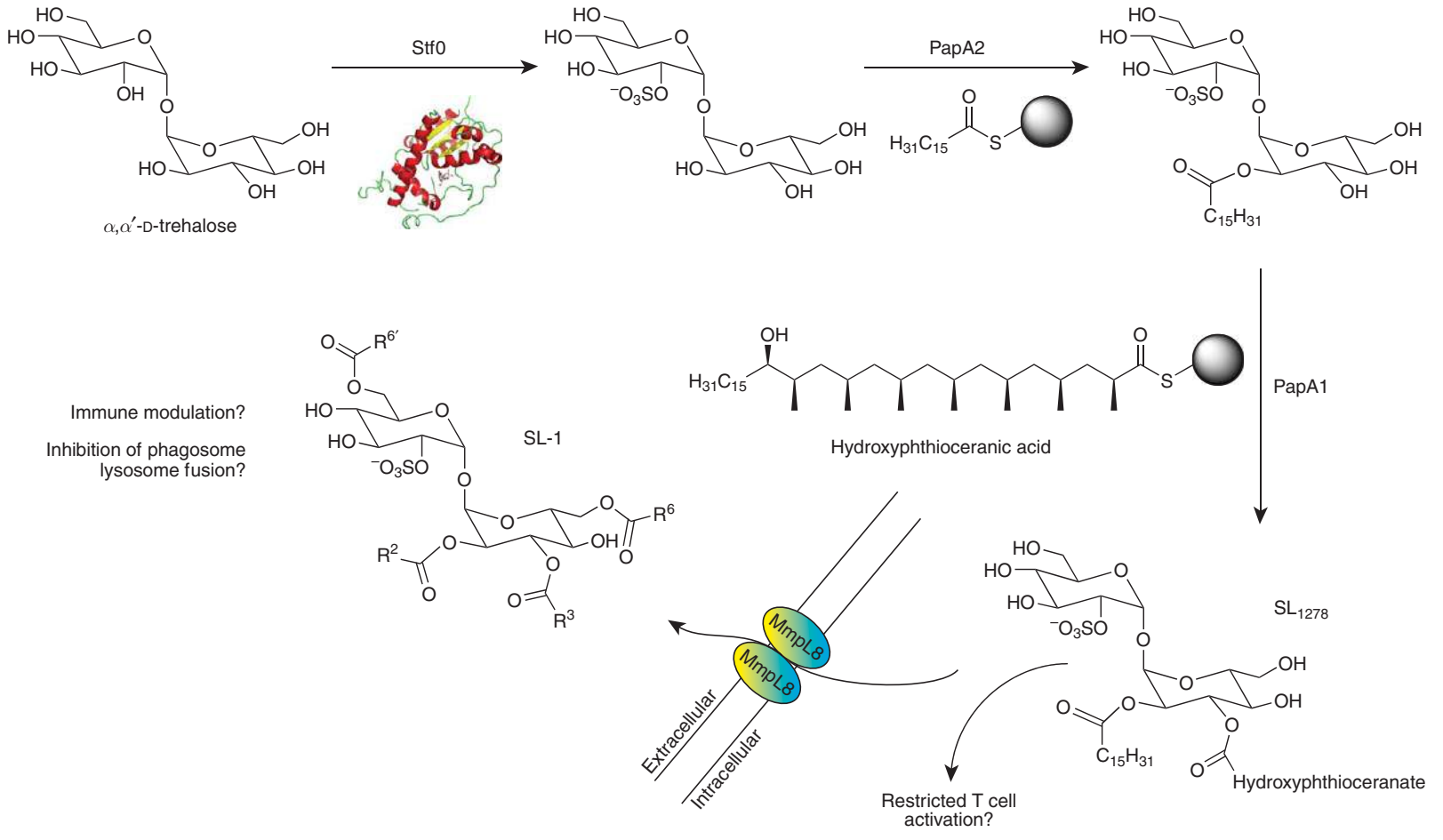


Figure 20 Biosynthesis and secretion of SL-1.

and spleen at a rate comparable to that of the wild type. Furthermore, the cellular immune responses are also comparable. Many previous studies had deduced that SL-1 affects phagosome–lysosome fusion and promotes intracellular survival, but a more detailed study of SL-1's role in these phenomena is warranted.[234]

1.04.4 Phthiocerol-Based Glycolipids

1.04.4.1 Polyketide Synthases in Mycobacteria

The polyketide synthases (PKSs) are a class of large, multifunctional enzymes with similarity to FAS systems. PKSs are widely expressed throughout the biological world. They can be produced as one polypeptide in type I systems, or consist of many discrete enzymes, where they are referred to as type II systems, analogous to the FAS I and II systems, respectively. There are also type III systems in mycobacteria associated with the synthesis of pyrones and chalcones.[235] The PKSs are similar to FAS systems but with some key modifications. PKS modules can contain the same catalytic domains (KAS, KR, DH, ER) but also possess AT and thioesterase (TE) domains. The AT and TE domains assist in relaying the growing lipid chain to other modules, release the products of multimodular iterative synthesis, and transfer them to amines and hydroxyls for the synthesis of modified proteins, carbohydrates, and other natural products. In addition to fatty acids, the PKSs synthesize a diverse array of molecules including polyketides, macrolides, carbohydrates, and chalcones. A number of these compounds possess antibiotic, immunomodulatory, or other intriguing biological properties. The numerous functions of the PKS chemical products testify to the power of microorganisms to produce highly versatile, architecturally complex natural products with a simple set of tools and starting materials.[236]

Phthiocerol dimycocerosates (PDIMs) are one product of the PKS machinery in mycobacteria. *Mtb* contains several annotated PKS homologues that have been listed elsewhere.[237] The most significant difference between PKS enzymes and FAS enzymes is in the substrate specificity and in the number, order, and types of modules present. These factors determine the diversity of products. In *Mtb*, the PKS-like phthiocerol synthases and mycocerosic acid synthases (MAS) utilize methylmalonyl-CoA (MMCoA) in addition to malonyl-CoA (MCoA). Use of both MCoA and MMCoA permits the synthesis of both linear alkanes and those that contain methyl branches, both of which are observed in phthiocerol dimycocerosates. Aside from the PDIMs and related molecules, PKS-derived moieties are present in SL-1 (as hydroxyphthioceranic acid), DATs, and triacyl trehaloses (TATs). More intriguingly, they are synthesized by homologous enzyme systems, suggesting common genetic origin, and this has in fact aided in the investigation of their properties. These represent the majority of PKS-derived cell wall lipids in *Mtb* and represent a significant portion of total cell wall-associated lipids, comparable in quantity to TDM. These lipid waxes, like TDM, can be isolated from the petroleum ether extract of the mycobacterial cell wall and characterized both structurally and immunologically.

The structures of the phthiocerol-based lipids were initially deduced by Noll[79] and Demarteau-Ginsburg *et al.*[238] Later studies by Goren *et al.*,[239] Brennan,[240] and Daffé[241–243] established the presence of these lipids in virulent strains of *Mtb* and *M. leprae*, confirmed their structures, and determined their stereochemistry. The phthiocerol lipids are present in seven pathogenic strains: *Mtb*, *M. bovis*, *M. africanum*, *M. marinum*, *M. ulcerans*, *M. leprae*, and *M. kansasii*; and one nonpathogenic strain, *M. gastri*. The phenolphthiocerol glycolipid (PGL-1) is abundantly produced and is characteristic of virulent strains of *M. leprae* and in some recently studied clinical strains of *Mtb*. Not all strains of *Mtb* and *M. leprae* contain PGL and its parent, phenolphthiocerol, but it is present in *M. bovis* and some other strains, and the reason for this distinction will be discussed below.

1.04.4.2 Phthiocerol Dimycocerosates

The discovery of the biosynthetic machinery required to assemble PDIM commenced with the investigation of mycocerosic acid synthase (mas) in *M. bovis* BCG. This gene occupies a rather large 6330 bp, 2110 amino acid-coding region, which codes for a linearly arranged KAS, AT, DH-ER, ketoreductase (KR), and pantotheine-ACP module as shown in **Figure 21**. These modules were assigned to human FAS domains based on each domain's homology. The region does not contain a thioesterase domain, implying that the synthase performs the role of elongation of the mycocerosic acid, but not the removal of the substrate from the enzyme once the

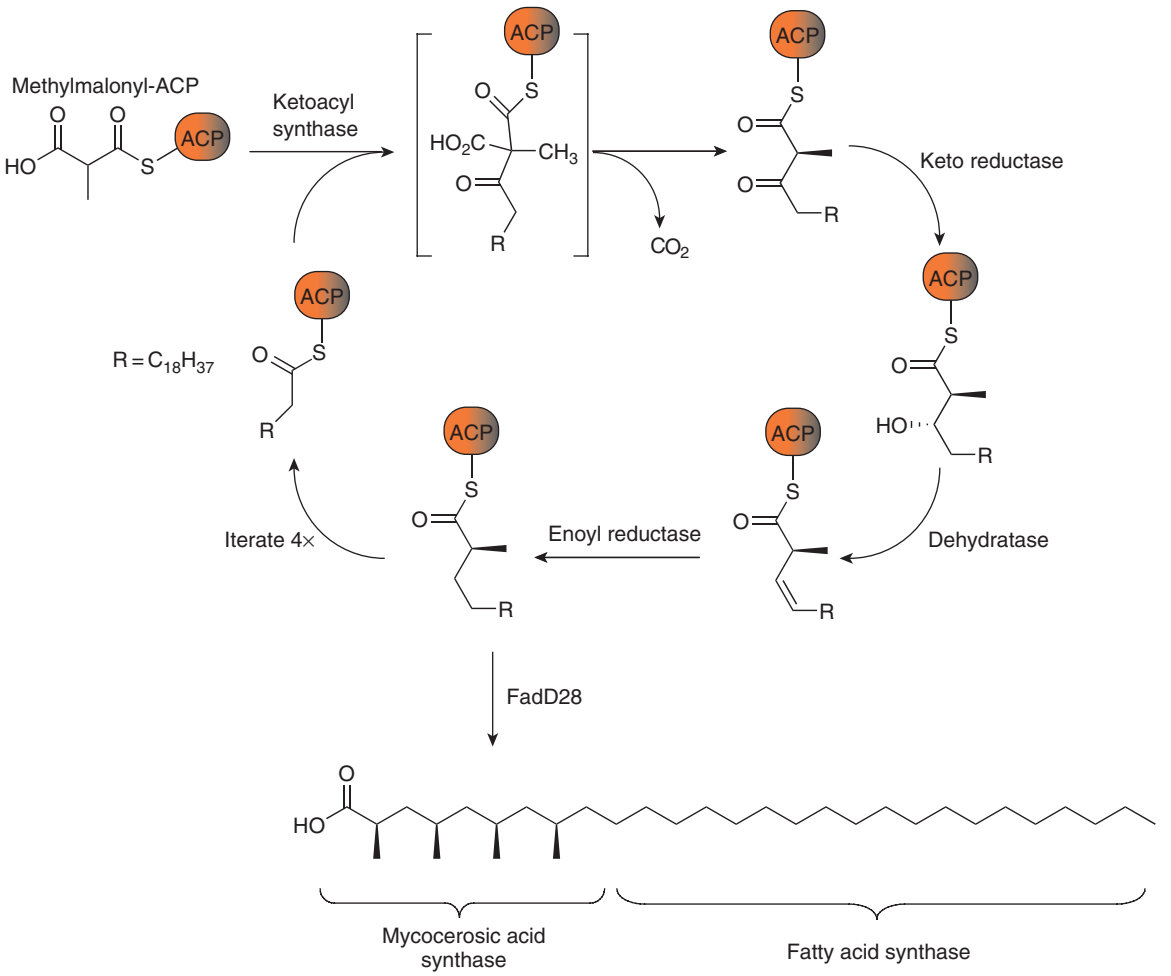


Figure 21 Mycocerosic acid biosynthesis; includes PDB 1TEX.

synthesis is completed. The ability to incorporate both ^{14}C acetate and propionate into the linear- and methyl-branched portions, respectively, suggests that both malonyl-CoA and methylmalonyl-CoA serve as substrates for phthiocerol and mycoceroseric acid synthesis. Two more essential genes located nearby in the genome, *fadD28* or *fadD26* (also called acetyl-CoA synthase, ACS), code for proteins that serve as ATs. The identification of *mas* and the DNA sequence of its domains provided comparative probes to identify the phthiocerol synthases that were present in adjacent open reading frames.[244–246]

Probes for each of the subdomains of MAS (AT, KAS) and FAS identified five sequential genes coding for a type I PKS, phthiocerol synthases 1–5 (*pps*1–5) in an *M. bovis* BCG genomic DNA library. Homologues were identified in *M. leprae* and *Mtb* (*pps*A–E). The first two genes, Pps1 and Pps2, contain KS, AT, and KR. These modules elongate the chain but leave the β-hydroxy intact. The next two, Pps3 and Pps4, contain KS, AT, KR, DH, ER, and ACP, which elongate the chain by four saturated carbon units. Finally, Pps5 contains KS, AT, and a thioesterase (TE); it installs a β-keto acid that decarboxylates to provide phthiodolone, the precursor to phthiocerol. This biosynthetic scheme for phthiocerol synthesis was suggested by Kolattukudy and coworkers,[245] and is depicted below (**Figure 22**).

The ability to analyze *Mtb* genes for function and ultimately to determine genes essential for growth, virulence, and persistence in animal infection models was long halted by the inability to reliably mutate *Mtb* genes by allelic exchange. The low frequency of DNA exchange and the low efficiency of transposon insertion into cells were further complicated by the slow growth rate of the pathogenic strains of mycobacteria. These

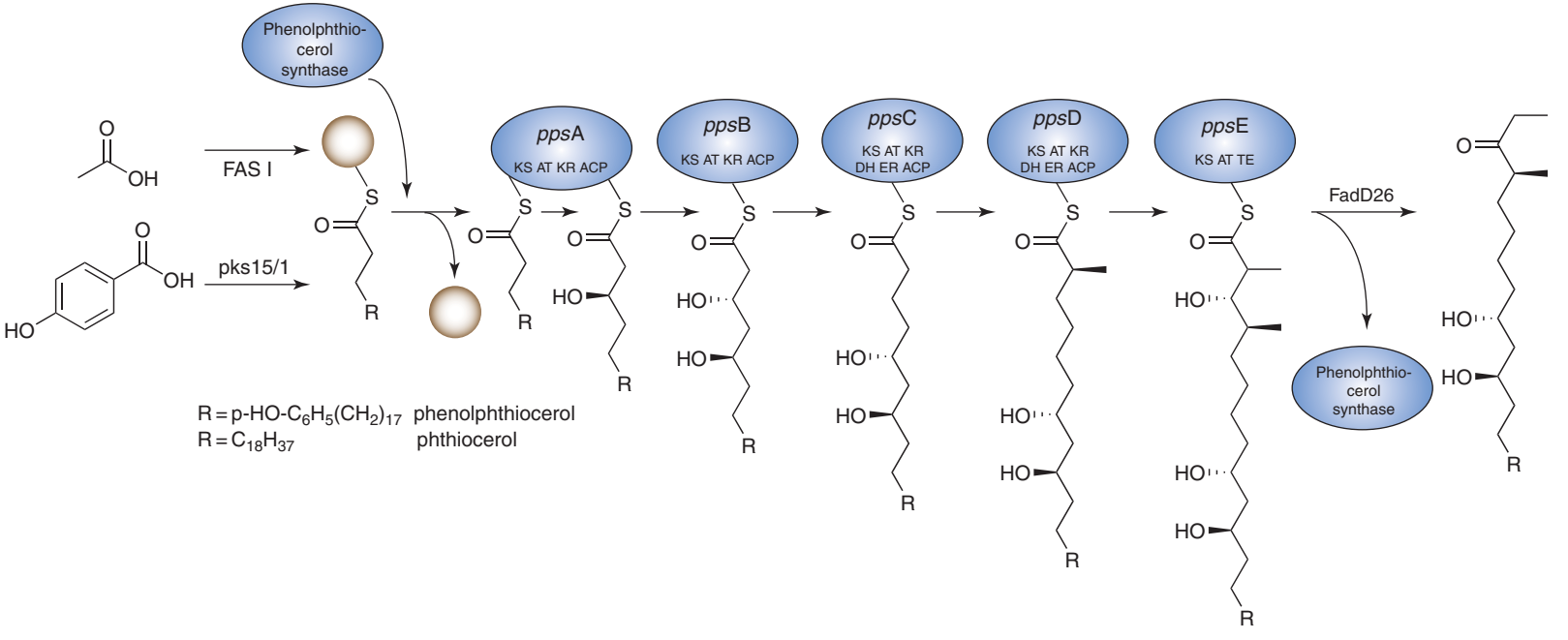


Figure 22 Phenolphthiocerol synthase *pps*A–E.

barriers were overcome by using a replicating plasmid either with a temperature-sensitive promoter or with a suicide vector. The gene of interest could be exchanged with a kanamycin resistance marker and the culture is then selected for insertion in the presence of kanamycin. Counterselection is achieved by growth at elevated temperature, which eliminates the wild-type gene from the cell. These efforts received a further boost from the completion of the *Mtb* genome sequence in 1998[247] and from the development of a Himar1-transposon mutagenesis method that allowed for definition of all essential genes within the *Mtb* genome.[248–250]

The insertion of libraries of labeled transposons into *Mtb* genes illuminated several enzymes associated with PDIM synthesis and also linked PDIMs to *Mtb* virulence. As a note, the effects on PKS-derived lipids could be analyzed by TLC methods used for the analysis of PDIM and other lipids. Using signature-tagged mutagenesis, a method that inserts a unique, inactivated DNA sequence for each gene, transposons were located in three genes, which resulted in attenuated growth of *Mtb* in macrophages. The first transposon was located in a promoter for the phthiocerol synthase *ppsA–E* genes. The second transposon was located in *fadD28*, an AT.[251] The third tranposon was deficient in *mmpL7*, a gene homologous to ActIII-ORF3 produced by *Streptomyces coelicolor*, a transporter that secretes the PKS-derived natural product, actinorhodin.[252]

The phenolphtiocerol synthase and Δ*fadD28* mutants could not synthesize PDIM, and all three mutants displayed altered colony morphology. In *Mtb*, colony structure and cording are linked with virulence, and this is the first time that colony morphology, virulence, and synthesis of a cell wall-associated lipid have been correlated. The cyclopropanation of α-mycolic acids has since been implicated as well.[68] Although it might be premature to assign a macroscopic property of bacterial colonies to the presence of a molecule and certain functional groups, the molecular basis for this phenomenon is emerging.[253]

MmpL7 is a transmembrane molecular transporter involved in the localization of PDIM. The Δ*mmpL7* strain could still synthesize PDIM but was unable to transport it into the cell wall. This lack of PDIM affected growth in the lungs, but not in the liver or spleen. These MmpL transporters have been implicated as a virulence mechanism.[254] Insertional mutants of all 13 of the *mmpL* family revealed that only *mmpL3* was absolutely essential for viability. *MmpL* 4, 7, 8, and 11 were essential for virulence in mouse models of infection. More recently, it has been shown that MmpL7 interacts with PpsE suggesting that the synthesis and transport of PDIM is coupled.[255]

Several subsequent transposon mutagenesis studies further confirmed the role of PDIM synthesis genes in virulence. A search for *Mtb* H37Rv mutants attenuated in the ability to infect mouse lungs identified several genes involved solely in lipid metabolism and membrane transport. These genes included the phthiocerol synthase-associated gene *fadD26* and another *pks* gene, *pks6*. A lipase gene lipF, important for the metabolism of host lipids, and an integral membrane protein (Rv0204c) were also identified. More interestingly, transposons were found in *mmpL7*, and *drrC*, an ABC-type membrane transporter, as well as *mmpL2* and *mmpL4*, two more membrane transporters. The *drrC*, *mmpl7*, and *lipF* mutants displayed attenuated growth in bone marrow macrophages. The role of *drrC*, *mmpL7*, and *fadD26* in mouse infection, in macrophage survival, in PDIM biosynthesis, and controlling cell wall permeability reinforces the importance of PDIM in contributing to infection and virulence.[256,257]

Several other *pks* genes were identified by their proximity to confirmed PDIM biosynthesis genes and were investigated for their relevance. When deletion strains were constructed for these genes, several resulted in attenuated infection in mouse models. These include *pks7*, mycocerosic acid-like synthase 7 (*msl7*),[258] *pks10* (a chalcone synthase),[259] and *pks12*[260] linked to the synthesis of mycoketide.[201] Several of these strains were deficient in PDIM synthesis but the synthesis of other lipids was unaffected. Using previous methods, a series of five proteins of unknown function (Pap1–5) were located in *Mtb*. These were called polyketide-associated proteins and they exhibited homology to the condensation domains of peptide synthetases. Because of their proximity to PKS enzymes in the genome, they were proposed to be the AT domains that catalyze the esterification of phthiocerol to mycocerosic acids. This is illustrated in **Figure 23**. Δ*papA5* did not synthesize PDIM, although both phthiocerol and mycocerosic acid were produced. A recent crystal structure provides evidence for this putative condensase role in PDIM biosynthesis.[237,261,262]

Until recently, little was known at the genetic level about how *Mtb* responds to different environments. Whole-genome sequencing has facilitated these comparisons, using tools like whole-genome microarray chips. *Mycobacterium bovis* BCG contains deletions of regulatory elements. How these deletions impact their ability to survive in macrophages is the key to better understanding bacterial stress responses. Using a cDNA–RNA

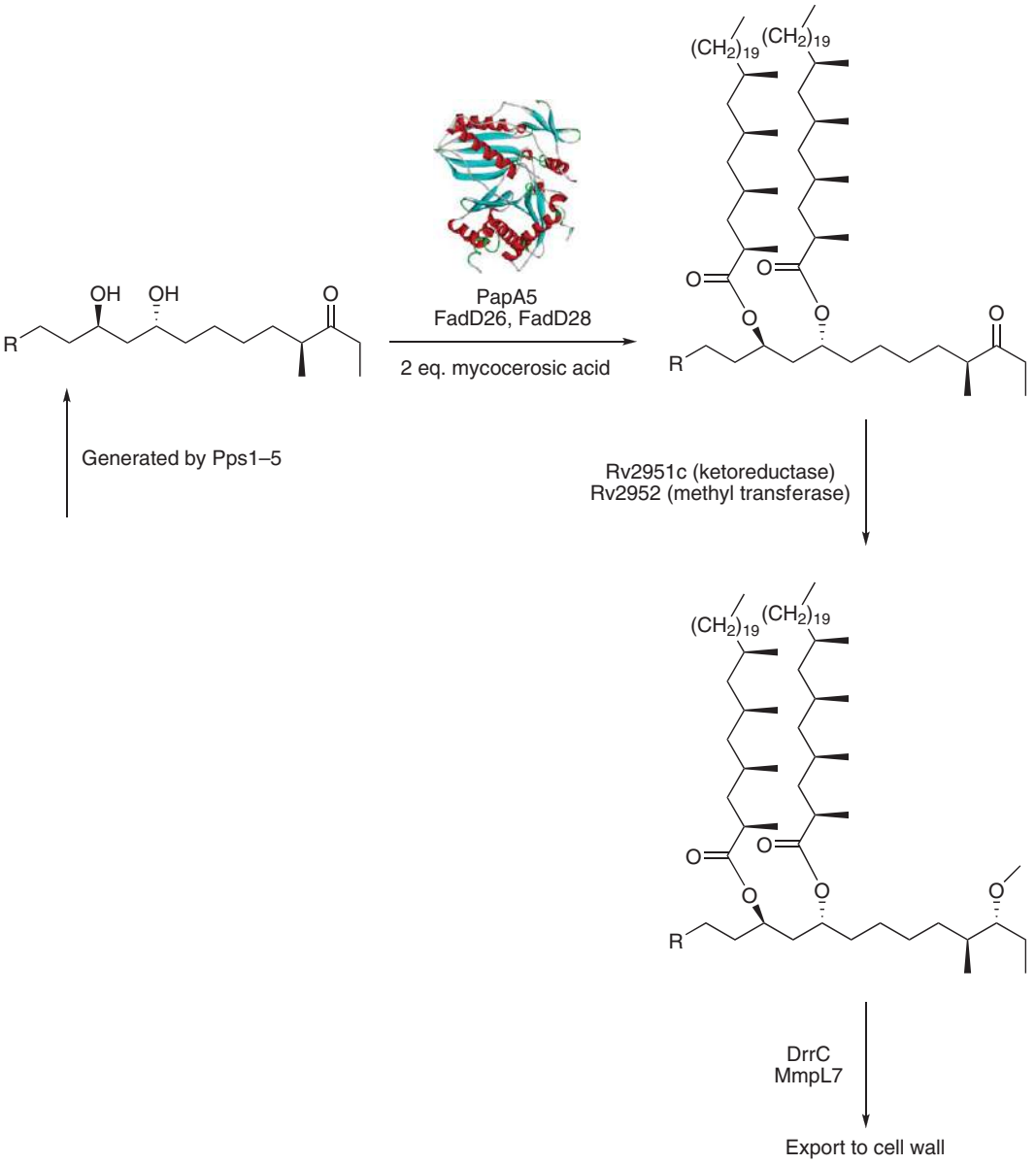


Figure 23 Final steps of PDIM biosynthesis and export to the cell wall; includes PDB 1Q9J.

subtractive hybridization procedure to separate host and bacterial mRNA transcripts,[263] an increased expression of mycocerosic acid synthase in intracellular *M. bovis* BCG was revealed. The upregulated genes in intracellular bacteria versus *in vitro* culture are *mas* (Rv2940c) (2.7-fold increase) and *fadD28* (Rv2941) (2.5-fold increase). This was the first report of the increased synthesis of these enzymes to promote survival in the macrophage. Also increased were *rplE* (ribosomal proteins), *groEL2* (heat shock protein), and *furB* (an iron regulatory protein). The upregulation of these genes implies that PDIM production is important for bacterial survival within macrophages.

1.04.4.3 The Phenolic Glycolipids

The phenolic glycolipids can be thought of as an elaborate alternative to PDIM and the biosynthetic pathways are clearly overlapping but diverge at the point of selection of a starter unit for the relevant PKS. Rather than

waxes like PDIMs, the PGLs are amphipathic and perhaps more than any other molecule produced by pathogenic mycobacteria appears to be designed specifically to interact with the host. Synthetic inaccessibility and the difficulties of working with such a complex lipid have limited the available information. Nonetheless, the pathogenesis of *M. leprae* has been shown to depend on PGL-1 and the presentation of its trisaccharide on the surface of the bacterium. PGL-1 binds to and activates complement C3 through the saccharide hydroxyls, which cleave C3's thioester.[264] Once opsonized, *M. leprae* enters macrophages safely, without oxidative burst, in a process mediated by complement receptors. PGL-1 also binds through its trisaccharide to Schwann cells, a component of the peripheral nervous system. This binding event facilitates invasion of these cells and the development of peripheral neuropathy that characterizes leprosy progression. Removal of the PGL saccharides diminishes both invasion and neuropathy.[265,266] These observations lead to the remarkable conclusion that PGL-1 alone determines the pathology of leprosy to a large extent.

In *Mtb*, PGL-1 has been shown to diminish the immune response, as measured by its effects on levels of proinflammatory cytokines produced in macrophages. Notably, strains of TB that have been proposed to have enhanced virulence properties as well as to be associated with higher levels of drug resistance, such as HN-878 and some W-Beijing family members were found to produce PGL-1.[267] This variability accents the potential that PGL-1 may be responding to evolutionary pressure and altering the virulence dynamic between humans and tuberculosis. Similar to the way polymorphic protein virulence factors respond to pressure from the human immune system, PGL-1 may be a marker for expansion of a formerly less dominant clade of tuberculosis strains. The failure to appreciate the variability of such molecules, and perhaps as relevantly, the lack of tools with which to systematically investigate the variation in such molecules, has led to a common misperception that TB strains are fairly homogenous.

Triglycosylated *p*-hydroxy benzoic acid methyl esters have been isolated in the culture supernatant of all reference strains of *Mtb.* These compounds are truncated forms of PGL-1 and suggest a starting substrate for its synthesis. However, *Mtb* H37Rv and a number of other strains do not express PGL-1. A comparison of PGL-1 nonproducing strains revealed a frameshift mutation in the pks15/1 gene, and complementation with functional pks15/1 restored PGL-1 synthesis. Thus, pks15/1 synthesizes the *p*-hydroxyphenyl alkanoate primer for phenophthiocerol and restores PGL-1 synthesis.[268] Although the presence of triglycosyl methyl benzoate suggested that glycosylation of *p*-hydroxy benzoic acid is the first step in PGL synthesis, there is no evidence to support this timing (**Figure 24**).

Three genes, Rv2957, Rv2958c, and Rv2962c, annotated as putative glycosyl transferases, were individually inactivated to determine their function in PGL-1 glycosylation. The extent of glycosylation was investigated on both *p*-hydroxy methyl benzoate (observed in culture supernatant) and phenolpthiocerol (observed in cell wall lipid fraction). ΔRv2962c did not produce any glycosylated products, suggesting that it catalyzes the first addition of rhamnose, an unusual L-sugar. The ΔRv2958c and ΔRv2957 mutants do not synthesize any of the triglycosylated forms, but produce a significant amount of mycoside B, a PGL derivative containing only one sugar, (2-*O*-methylrhamnose), which is also the major glycolipid of *M. bovis* BCG. Analysis of the *M. bovis* BCG homologue of Rv2958c indicated an identical gene with a frameshift mutation. Transformation of *M. bovis* BCG with functional Rv2958c produced the disaccharide, establishing its function. A similar complementation of Rv2957 did not produce the trisaccharide and so the enzyme that performs this third glycosylation step is still unknown.[269]

Rv2951c and Rv2952 were also investigated, as a consequence of their proximity to these glycosylases. Phthiodolone dimycocerosate, a variant of PDIM, is also present in the virulent strains of mycobacteria like *M. kansasii* and *M. ulcerans*, which contain phthiodolone and not phthiocerol lipids. An oxidoreductase coded by Rv2951c performs the reduction and is succeeded by the methyltransferase activity of Rv2952, completing the biosynthesis of PDIM (see **Figure 23**).[270–272]

1.04.4.3.1 PDIM-less mutant strains of Mtb as vaccine candidates

The fact that an inability to synthesize PDIM reduces bacterial survival in the macrophage and results in attenuated infections has prompted the suggestion that some of these strains might provide better vaccine candidates than BCG. The case of Rv2958 suggests that one reason for the attenuation of the BCG strain is that it has acquired mutations in PDIM and PGL production.[273] Δ*fadD26* and other PDIM-deficient strains are also under investigation.[274,275] Studies have demonstrated that PDIM protects the bacterium from nitric oxide

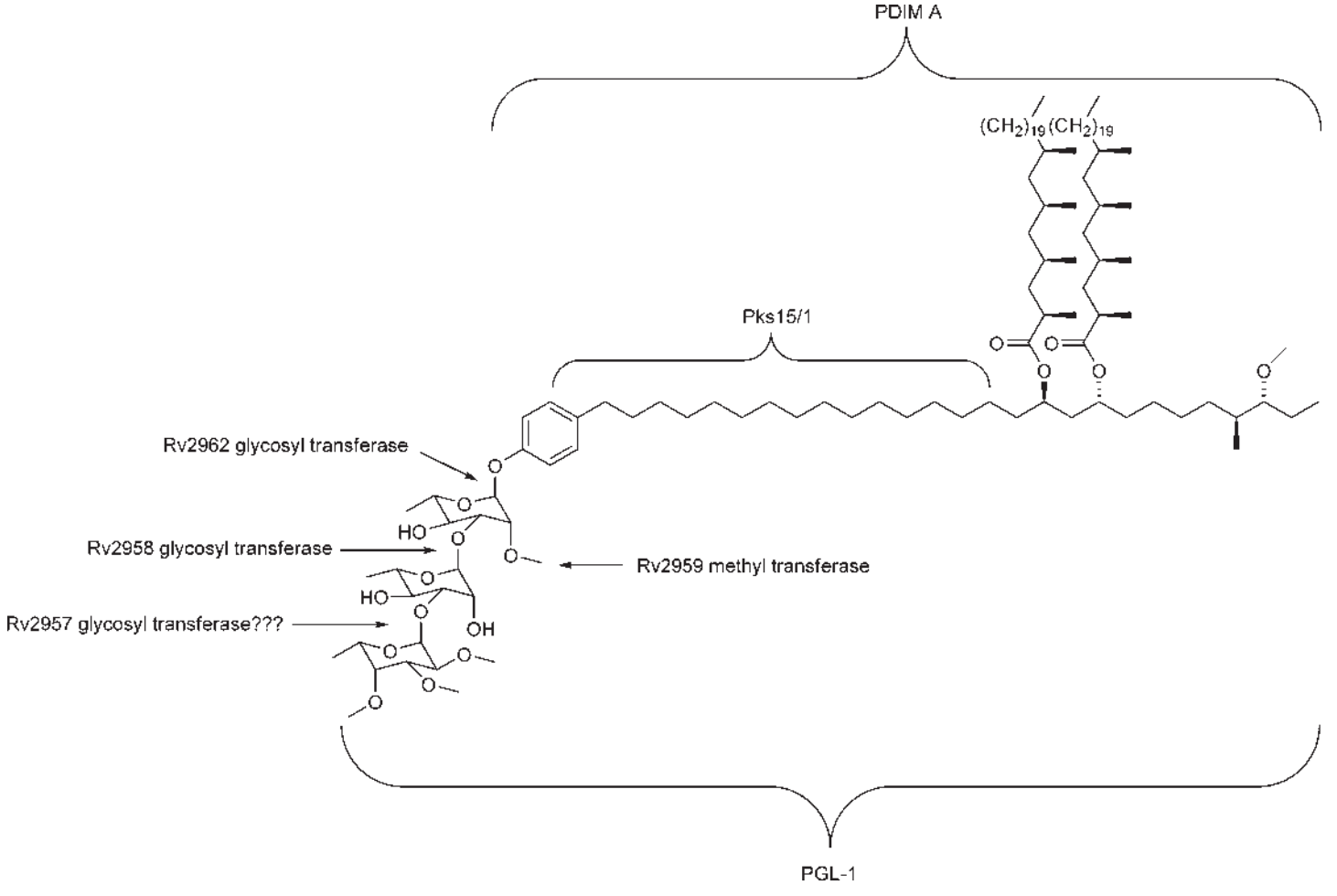


Figure 24 PDIM and PGL biosynthesis.

damage and there is evidence that PDIM modulates early responses of proinflammatory cytokines.[276,277] Understanding the role of PDIM in pathogenesis would facilitate a better understanding of the potential for these strains as vaccines.

1.04.5 Lipoarabinomannan and Phosphatidylinositides

1.04.5.1 Introduction to the Polar Glycolipids

Mycobacteria contain a series of polar, antigenic glycolipids referred to as LAMs. The constituents of these molecules were deduced by degradation studies and characterized by Brennan and coworkers.[278] LAM consists of a phosphatidyl *myo*-inositol (*cis*-1,2,3,5-*trans*-4,6-cyclohexanol) (PI) core connected by phosphate ester to a diacyl glycerol containing tuberculostearic acid and palmitic acid. These fatty acids serve as the molecule's anchor into the cell membrane. The *myo*-inositol is linked to a mannose oligosaccharide and then to an arabinose oligosaccharide. There are truncated forms of LAM that include lipomannans (LM) and phosphatidyl *myo*-inositol manno-oligosaccharides (PIM), which together have been termed 'modulins' because of their ability to modulate immune responses. Because of their polar and heterogenous nature, these compounds were first resolved by gel electrophoresis and purified by anion exchange. LAM has attracted much interest as its properties include the inhibition of macrophage activation and T cell proliferation, and the ability to neutralize oxidative free radicals aimed at the bacterium.[279] The diacyl glycerol portion not surprisingly renders these glycolipids ligands for CD1b and there are restricted T cells, which react with the species-specific carbohydrate portions.[192] There are other cell wall oligosaccharides, including the trehalose containing lipooligosaccharide (LOS),[280] but LAM and its related variants are the most prominent (**Figure 25**). As a result of their complexity, the study of these compounds has been an arduous task and is far from complete. In this section, the structures and some of the recent discoveries regarding the biosynthesis and immunomodulating properties of this intriguing class of glycolipid are briefly highlighted.

The core of LAM, LM, and PIMs is PIM_2 (**Figure 25**). It consists of *myo*-inositol-1-phospho diacyl glycerol. The inositol contains $1 \rightarrow 2$- and $1 \rightarrow 6$-linked mannose residues. There are several variants of PIM_2 that contain from two to four fatty acid-derived acyl chains of tuberculostearic (C_{19}), stearic (C_{18}), or palmitic (C_{16}) acids. Another major component of cell wall glycolipid is PIM_6 (**Figure 25**).[281] In one case, PIM is bound to a D-mannan polysaccharide connected to a D-arabinan polysaccharide. The mannose residues are bound in an $\alpha 1 \rightarrow 6$ fashion, with some *O*-2 (shown) or *O*-3 branching residues, a species-specific signature. The site of attachment of the arabinan is beyond the scope of current detection methods and has not been unambiguously confirmed. The arabinan domain contains arabinofuranose residues in an $\alpha 1 \rightarrow 5$ linkage, comprising a linear backbone. This backbone is linked through $\alpha 1 \rightarrow 3$ manner to two types of arabinan branches. The first is a linear tetrasaccharide (Ara_4) of Ara*f* $1\beta \rightarrow 2$-Ara*f*-$1\alpha \rightarrow 5$-Ara*f*-$1\alpha \rightarrow 5$-Ara*f*-$1\alpha \rightarrow 3$-linear backbone. The second is a bifurcated hexasaccharide (Ara_6) of Ara*f* $1\beta \rightarrow 2$-Ara*f*-$1\alpha \rightarrow 3$ and Ara*f* $1\beta \rightarrow 2$-Ara*f*-$1\alpha \rightarrow 5$-Ara*f*-$1\alpha \rightarrow 5$-Ara*f*-$1\alpha \rightarrow 3$-linear backbone.

Both the Ara_6 and Ara_4 branches are capped in a species-specific manner. There are three general types of capped LAM. The simplest is uncapped araLAM. Phospho-*myo*-inositol-capped LAM is called PILAM, and mannose-capped LAM is called ManLAM. The type of capping is important in modulating the molecule's immune response and will be discussed briefly. *Mtb*, *M. bovis*, *M. leprae*, and *M. avium* contain mono-, di-, and trimannosides in an $1\alpha \rightarrow 2$ linkages bound to the 5-position of the arabinose terminus. These strains contain a heterogenous series of ManLAMs with molecular weights of up to $17\,000 \pm 6000$ Da.[282] Fast growers, *M. smegmatis* and *M. fortuitum*, contain the smaller PILAMs.[283]

1.04.5.2 Biosynthesis of LAM

Previous studies suggest that the construction of LAM starts from inositol (**Figure 26**). SuhB (Rv2701c) catalyzes its phosphorylation to give inositol-1-phosphate.[284] After further processing to PI, PimA (Rv2610c) transfers a mannopyranose from GDP-mannose to PI to provide PIM_1. Although the sequence is not known,

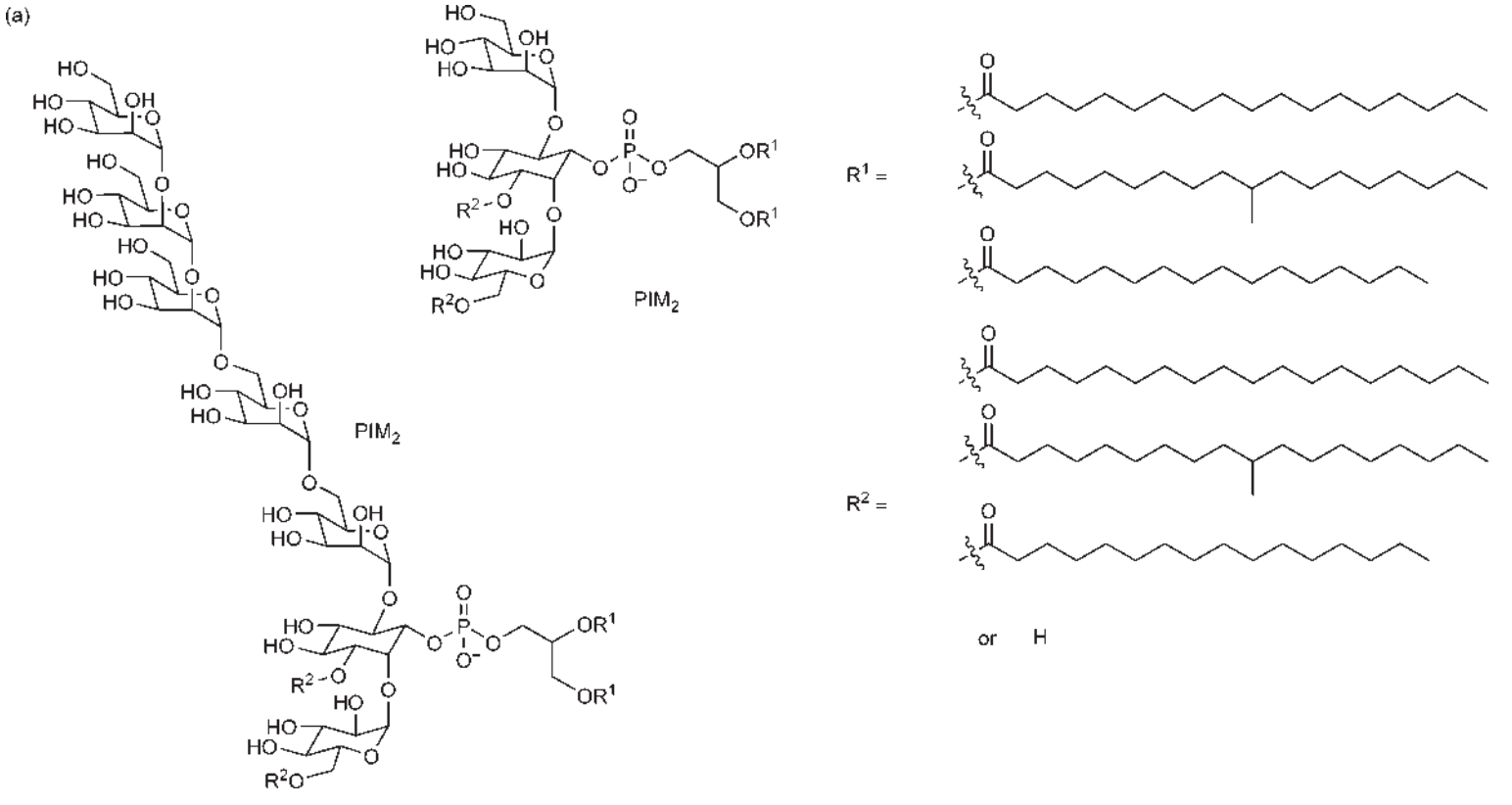


Figure 25 (Continued)

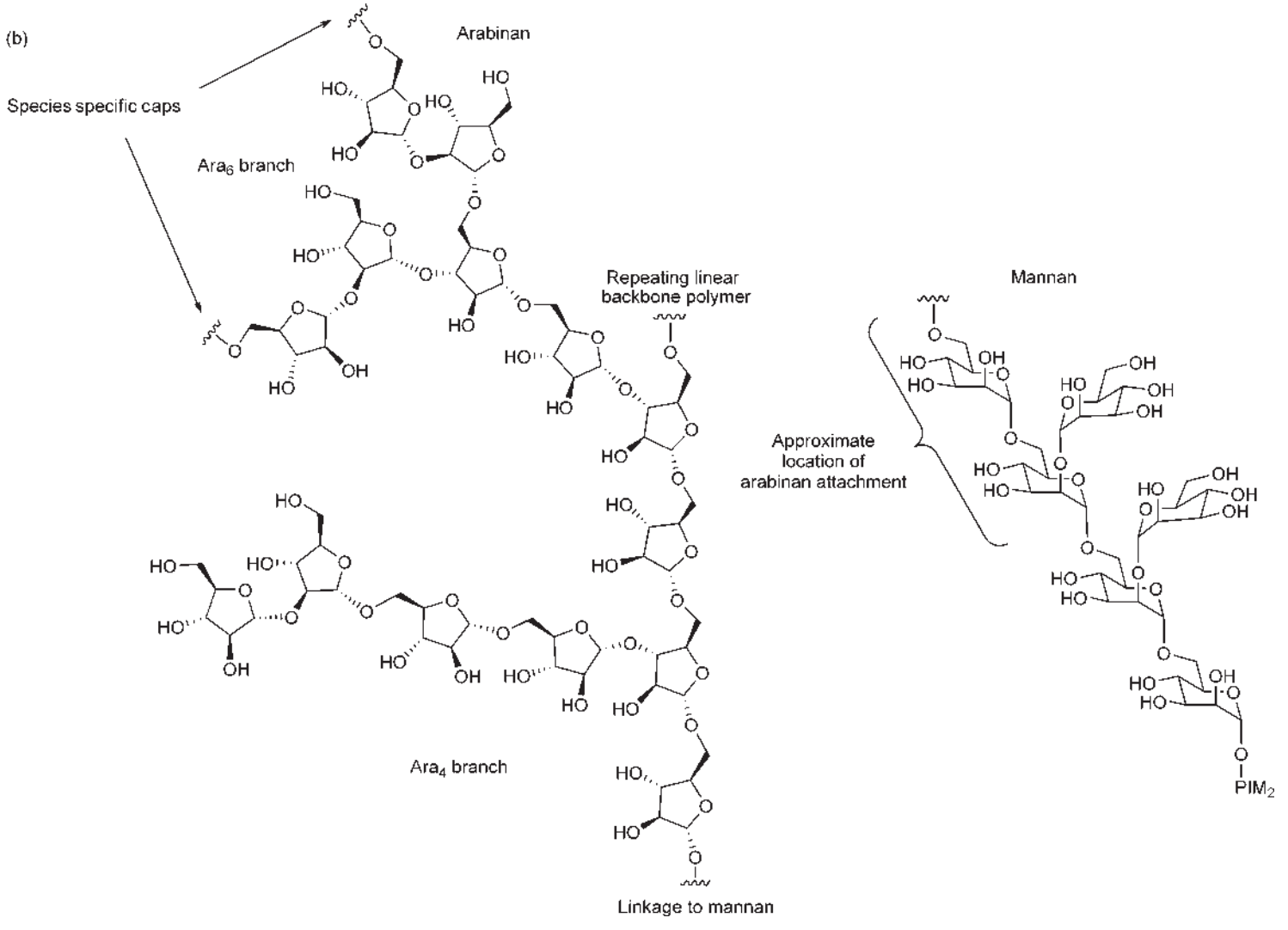


Figure 25 (a) Structures of PIMs, LM, and LAMs. (b) Structures of the mannan and arabinan portions of LM and LAM.

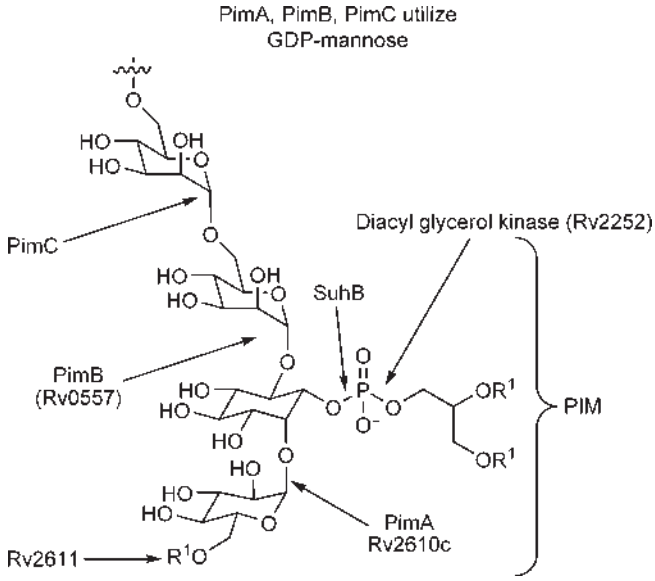


Figure 26 Biosynthesis of LAM core.

the mannose is acetylated at the 6-OH[285] with a fatty acid and PimB (Rv0557) catalyzes the glycosylation of the inositol 2-OH[286] with another unit of GDP-mannose to provide Ac_1-PIM_2. PimC,[287] a possibly redundant enzyme, characterized in *Mtb* CDC1551, catalyzes the attachment of another mannose residue to provide Ac_1-PIM_3. Once the third mannose residue is appended, the attachment of additional mannose residues shifts from a GDP-mannose glycosyltransferase family to a series of enzymes that utilize heptaprenyl (C_{35}) or decaprenyl (C_{50}) mannose phosphate as mannose donors.[288] This change occurs because earlier mannosylations take place in the cytosol with the water-soluble GDP-mannose. Subsequent mannosylations take place in the periplasm or in the cell wall and require the $C_{35/50}$-mannose because it can be translocated across and anchored in the cell membrane. Proof for this changeover derives from the finding that, amphomycin, which binds to polyprenols and inhibits membrane translocases, inhibits the formation of higher-order mannose oligosaccharides.

Although the individual subunit biosynthetic pathways have not been extensively characterized, Rv3257c codes for the phosphomutase, which converts mannose-6-phosphate to mannose-1-phosphate, the entry point to activated mannose building blocks.[289] PimE (Rv1159) and PimF (Rv1500), and Rv2174c are glycosyl transferases, suggested to attach mannose residues to higher-order mannose substrates, in a linear $\alpha 1 \rightarrow 6$ fashion (**Figure 26**). The intermediates and enzymes that precede these higher-order mannose oligosaccharides have not been identified.[290–293] Rv2181 attaches the $\alpha 1 \rightarrow 2$ mannose residues that decorate the linear mannose chain.[294] Although it is implied, it should not be assumed that LM is a biosynthetic precursor for LAM, and where the pathways diverge is unclear.

The interest in the biosynthesis of LAMs related to the fact that one of the first-line antitubercular antibiotics, ethambutol (ETH), inhibits the synthesis of arabinan, affecting both arabinogalactan and LAM biosyntheses. *Mtb* contains three homologous proteins EmbA, EmbB, and EmbC, which are arabinosyl transferases, utilizing β-D-ara*f*-1-monophosphoryldecaprenol as the arabinose donor. EmbA and EmbB synthesize cell wall arabinogalactan whereas EmbC synthesizes LAM. These enzymes are the targets of the antitubercular antibiotic ETH, although the precise mechanism is not clear. Resistance to ETH is usually associated with mutations in these three proteins.[295,296] ETH-resistant strains produce uncapped and truncated forms of arabinan, notably in the linear Ara_4 side chains[297] indicating a defect in arabinan biosynthesis.

1.04.5.3 Immunopathogenesis

The biological properties of LAM are derived from a number of key observations. LAM is secreted and plays a role in the inhibition of phagosome–lysozome fusion, a property specific to the Man–LAM-capped version[298]

of pathogenic mycobacteria. LAM is a CD1b ligand and some of its *in vivo* properties probably derived from the binding of CD1-restricted T cells.[192] An initial profile examined the immunological responses of macrophages to PIM, LM, and LAM. The macrophage cytokine induction profiles for LAM, LM, and PIM include TNF-α, IFN-γ, and several Th1 cytokines. The responses were nearly identical for all three glycolipids, suggesting that the basis for immunological stimulus lies in the phosphatidyl inositol portion of the molecule. PI-Ara_4, obtained from the degradation of a PILAM cap from a fast-growing strain of mycobacteria, showed TNF-α induction in macrophages.[299] The deacetylation of PIM reduced this activity, suggesting that some component of the carbohydrate, phosphate, and fatty acid was essential for immunogenicity.[300]

The expression of different types of polar glycolipid plays an immediate role in the properties of the membrane itself and in the composition of exposed cell surface carbohydrates.[301] The immune response to the arabinan-capping motif is species-specific. Man–LAM does not exhibit apoptotic or IL-12-inducing activity on macrophages, whereas PILAM and LM do. LM by virtue of its exposed mannose core induces a profound inflammatory response, in contrast to PIM and Man–LAM.[302,303] It has been hypothesized that the cell's ability to regulate the LM:LAM ratio might play an important role in modulating the immune response.[304,305]

Some recent clinical evidence implicates a more complex role for LAM's immune-modulating properties. ETH-resistant strains exhibit a decrease in mannose caps from two residues to one per Ara_4 chain and from four residues to two per Ara_6 chain.[306] It is striking that drug resistance might impart not only the ability to evade the action of a particular small molecule, but also the ability to evade the immune system. The production of Man–LAM is not only species-specific but strain-specific as well. Two clinical strains of Mtb, HN885 and HN1154, exhibited decreased affinity for uptake by phagocytic cells, presumably through a mechanism mediated by mannose receptors, a common uptake mechanism used by macrophages.[307]

LAM is an essential component of the *Mtb* membrane, and has been shown to have important immunomodulatory roles and is an important component of granuloma formation and bacterial persistence in the host. It has been proposed as a target for developing new, simple ELISA-based tests for *Mtb*. Many of the details of LAM biosynthesis are still unknown and further research in this area should yield a better understanding of this complex glycolipid and potentially novel antituberular drugs that could have an impact similar to that of ETH.

1.04.6 Exotic Glycolipids and Glycopeptidolipids

1.04.6.1 Glycopeptidolipids and Sliding Biofilms

GPL antigens are a class of polar molecules restricted to the cell surface of atypical mycobacteria including members of the *M. avium* complex (*M. avium*, *M. intracellulare*, *M. paratuberculosis*, and *M. scrofulaceum*) and more recently reported from the opportunistic strains, namely, *M. fortuitum*, *M. abscessus*, *M. chelonae*, *M. xenopi*, and *M. kansasii*. Along with PGL and PDIM, these compounds have been called mycosides but this only refers to their lipophilic nature since the molecules do not contain the multi-methyl-branched mycoceroseric acids of PDIM.[3] *M. avium* is not nearly as common a human pathogen as *Mtb*, *M. leprae*, or *M. ulcerans*, but it is ubiquitous in the environment, living even on plumbing fixtures in a usually benign form. *M. avium* has been a consistent cause of morbidity and mortality among immunocompromised HIV-positive patients. *M. avium* and several other GPL-producing organisms in this group of nontuberculous mycobacteria (NTM) can cause a tuberculosis-like lung infection in patients with a variety of immune disorders, but they are more commonly identified with disseminated infections that are difficult to diagnose and treat.[308]

Unlike other lipophilic molecules on the mycobacterial surface, the GPLs contain a significantly hydrophilic peptide and carbohydrate portion (**Figure 27**). Consequently, these antigens generate a robust and specific antibody response in infected experimental animals and humans. Antibodies produced to GPL have been used to distinguish NTM such as *M. avium* from *Mtb*. *Mycobacterium avium* strains isolated in the clinic can be differentiated into approximately 30 serovars, which differ in the precise carbohydrate composition. A few serovars are overrepresented in HIV patients, suggesting a role of the fine structure of GPLs in immunomodulation and virulence. Virulent strains display an outer capsule composed of GPL that appears to play an important role in intracellular survival and the presence and composition of GPL determines colony morphology and virulence. The serovars display different properties and can be differentiated by immunological reagents or by TLC analysis using chloroform:methanol:water solvent systems.[309,310]

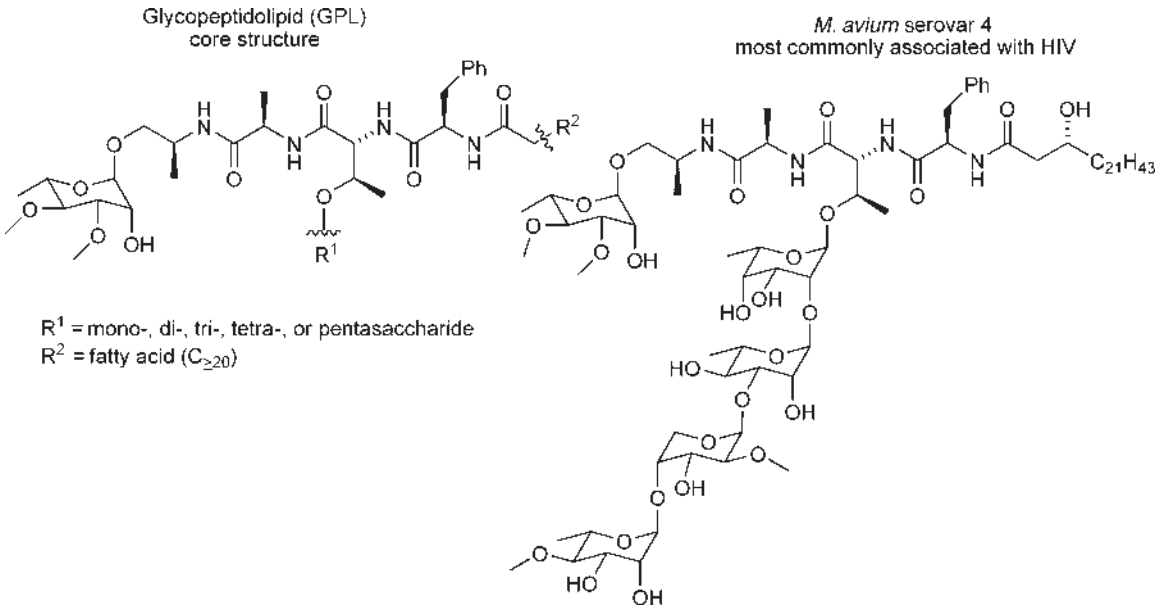


Figure 27 GPL core structure and *Mycobacterium avium* serovar 4.

The GPL molecules contain a common core peptide L-alaninol-D-alanine-D-allo-threonine-D-phenylalanine. The alaninol hydroxyl is usually appended to a 3,4-di-*O*-methyl-α-L-rhamnopyranose. The D-phenylalanine N-terminus is acylated with a fatty acid or a β-hydroxy fatty acid that contains several degrees of unsaturation (C_{26}–C_{34}). The allo-threonine hydroxyl is glycosylated and the number of carbohydrate residues at this attachment point distinguishes most of the serovars since the peptide core is reliably invariant. The α-L-rhamnopyranosyl-(1 → 2)-6-deoxy-α-L-talose disaccharide is common, and in the case of *M. avium* serovar 4, the complete tetrasaccharide 4-*O*-Me-α-L-rhamnopyranosyl-(1 → 4)-2-*O*-Me-α-L-fucopyranosyl-(1 → 3)-α-L-rhamnopyranosyl-(1 → 2)-6-deoxy-α-L-talose is attached (**Figure 27**).[311,312]

The colony appearance of *M. avium* serotypes have been described as smooth opaque, smooth transparent, and rough. In general, bacterial colonies of serovars that contain multiply glycosylated GPLs appear smooth and in prior studies, the individual bacteria were described as encapsulated when viewed under a microscope. The smooth forms contain a highly immunogenic, multiply glycosylated, serovar-specific GPLs (ssGPLs), whereas the rough ones either do not contain GPL or contain a GPL with two or less carbohydrate residues called nonserovar-specific GPLs (nsGPLs). The extent of glycosylation impacts the morphology of colonies and rough forms can be obtained by repeated culture on solid media, during which they spontaneously lose genomic DNA contributing to GPL biosynthesis. Rough and smooth opaque morphotypes derived from a smooth, transparent virulent strain sometimes lose virulence, although the relationship between morphology and virulence is complex.[313,314] The loss of GPL production has also been observed in conversion from smooth to rough forms in *M. abscessus*, an opportunistic pathogen in individuals with cystic fibrosis. The rough forms display an increase in cording, greater virulence, and greater ability to invade monocytes.[315]

It has been hypothesized that the GPLs modulate surface hydrophobicity, and this has received some support since the biological surfactant Tween 80 can restore a smooth phenotype to a rough variant in culture.[316] Two rough colony mutants of *M. avium* have been found that secrete only the peptidolipid core of GPL, but seemed incapable of appending either of the carbohydrate moieties. An important caveat in considering the importance of GPL phenotypes is that the bacterium displays an ability to express a multitude of phenotypes during infection. Morphotypes are culture dependent and so might not be seen in a particular media, but colonies isolated from monoclonal infections have been proven capable of producing all three of the morphotypes. Polyclonal infections have been shown to play a role in increased virulence and drug resistance. A single clone of *M. avium* might achieve this by its ability to display multiple morphotypes with differential

surface characteristics and drug susceptibilities.[317,318] Even isolated GPLs have been shown to have immunomodulatory ability and possibly play a role in preventing phagosome–lysozsome fusion.[319] The precise effects are specific to the type of GPL produced by each serovar.[320]

1.04.6.1.1 Biofilms and sliding

Another virulence-associated characteristic of bacteria drawing much recent attention is the ability to form biofilms. Biofilm formation and the ability to colonize surfaces are phenomena mediated by bacterial populations, which can correlate with increases in drug resistance and resistance to disinfectants and the chlorine treatment of water. *Mycobacterium smegmatis* and *M. avium* colonize surfaces on drinking water fountains, showers, and pipes by forming biofilms. Despite the lack of a flagellum, the bacteria can slide, moving as a coordinated unit on surfaces. This motion is the consequence of a population of cells attempting to spread on a surface while minimizing intercellular friction. Transposon mutagenesis studies produced strains of *M. smegmatis* that could not form a biofilm or slide. These genes include homologues of the *tmtp* (membrane transporters) in *M. avium* and of the *mmpL* genes in *Mtb*, *atf1*, a predicted membrane protein responsible for acetylating the 2-position of the 6-deoxy-α-L-talose residue, and *mps*, the peptide synthetase responsible for synthesizing the peptide core of GPL (see **Figure 28**).[321] Also, *mtf1*, a methyltransferase that methylates the 2- and 3-hydroxyls of the alaninol rhamnose residue was identified.[322] Without these genes, synthesis of GPL in *M. smegmatis* is severely limited, and all of these strains were deficient in surface-associated, multiply glycosylated GPLs.[323] In *M. avium*, disruption of *pstAB*, the nonribosomal peptide synthetase that generates GPL inhibited the bacterium's ability to grow on some surfaces.[324] The ability to form biofilms on steel was also diminished, and this is relevant considering that *M. avium* colonizes surfaces that contact drinking water supplies.[325]

1.04.6.1.2 Biosynthesis of GPLs

The biosynthesis of the carbohydrate moiety of GPL has received scrutiny recently because of observations that a culture of *M. avium* raised under limited glucose or nutrient-starved conditions exhibited increased production of GPLs.[326] These conditions resemble the nutrient-starved environment of the macrophage or granuloma, and the aim of such studies was to understand the nature of the biological adaptation to disease-relevant conditions and to identify new targets for drug intervention. Growth of *M. smegmatis* in glucose starvation causes a reversion to smooth colony forms.[327]

A series of reports have elucidated the *M. avium* genes associated with the biosynthesis of ssGPLs usually by identification of homologues in *M. smegmatis* or by expression of the putative *M. avium* serovar-specific glycosylases in *M. smegmatis* knockout models (see **Figure 29**). The GPLs produced in this heterologous host facilitated assignment of the function of these enzymes, analogous to the study of mycolates. The peptide core is synthesized

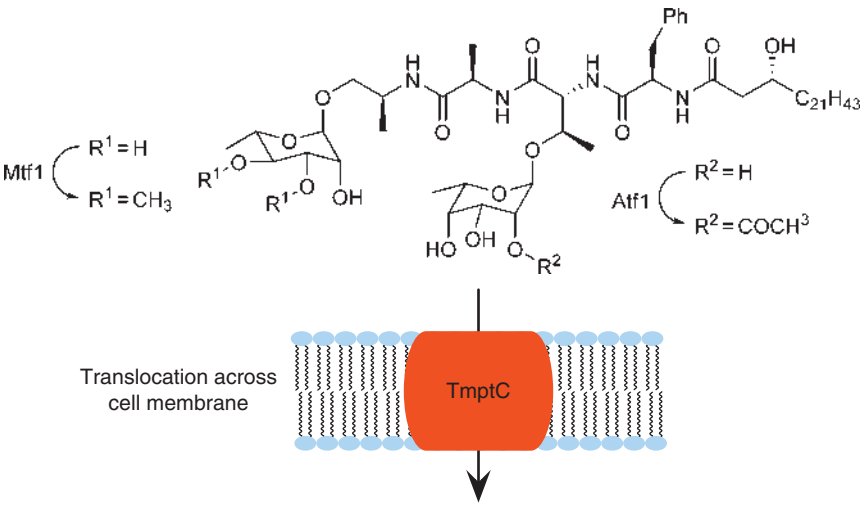


Figure 28 Biosynthesis and secretion of GPLs associated with biofilm formation and sliding in *Mycobacterium smegmatis*.

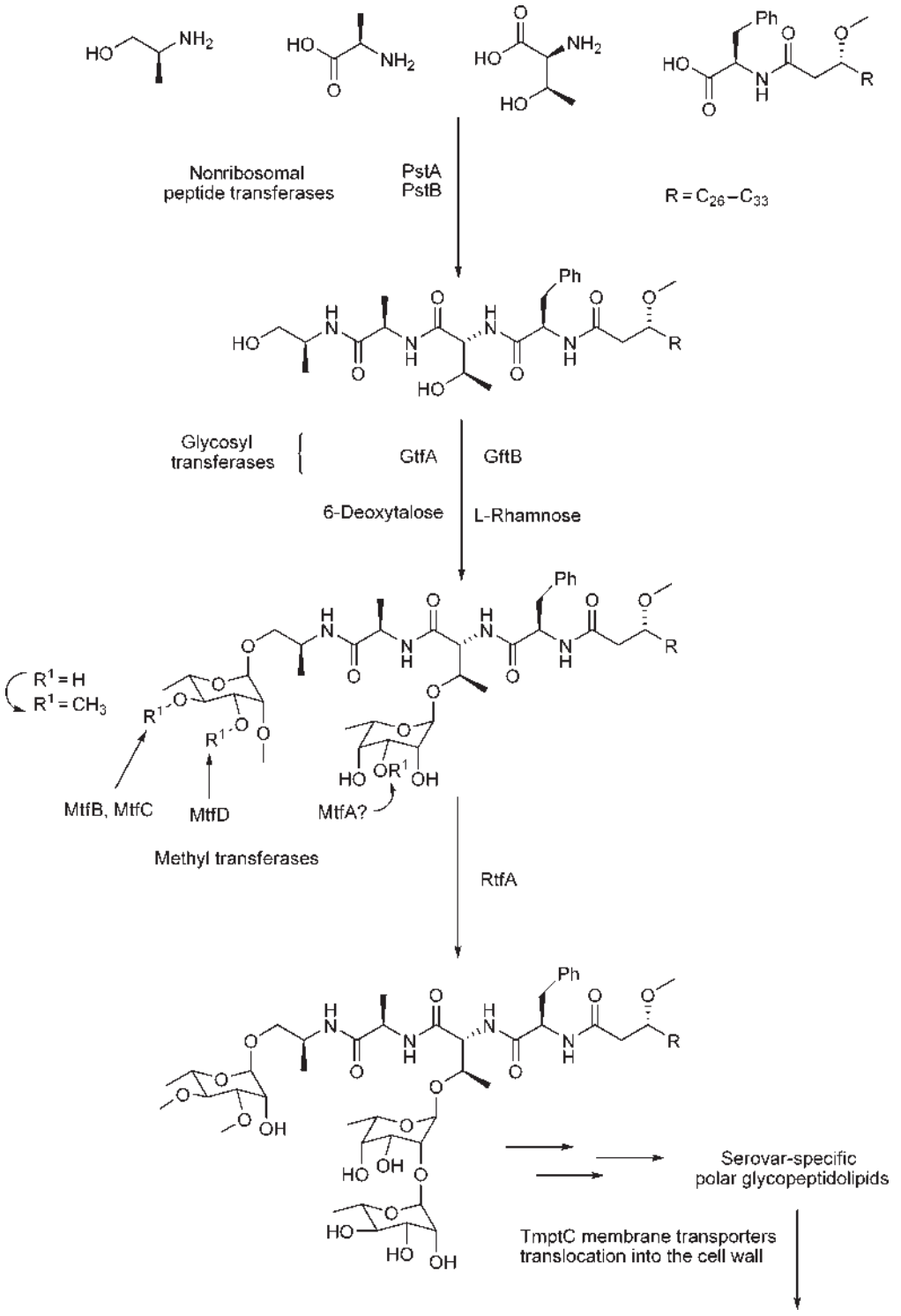


Figure 29 Biosynthesis of *Mycobacterium avium* GPLs.

by *pstA*, *pstB*, identified through its *M. smegmatis* homologue, *mps*.[325,328] The *mps* gene contains four domains, each with amino acid-binding domains, and three that contain amino acid racemase domains to prepare the D-amino acids.

The *M. avium gtfA* and *gtfB* genes correspond to the *gtf1–4* cluster in *M. smegmatis*. GtfA installs the 6-deoxy-talose and GtfB installs the alaninol-rhamnose.[329] The carbohydrate residues are then modified by methyl-transferases. MtfD methylates the rhamnose 3-hydroxy, a prerequisite for further methylation. The rhamnose 4-hydroxy is methylated by MtfC or MtfB, and another protein in the cluster, MtfA, is thought to modify the 3-hydroxy of the 6-deoxy-talose residue.[322,330–332] At this point, the *M. avium* and *M. smegmatis* GPL synthetic pathways diverge. A rhamnosyl transferase (*rtfA*) gene that attaches a rhamnose residue to the 6-deoxy-talose residue has also been confirmed.[333,334] In the course of these studies, allelic exchange techniques for introducing genes into *M. avium* have been developed, which will hopefully provide more insight into the mechanisms of *M. avium* virulence.

1.04.6.2 Mycobactins

The survival and growth of mycobacteria depends on access to iron, a nutrient that is necessary for the function of many essential enzymes and metabolic processes. The presence of oxygen on our planet made insoluble Fe^{3+} a growth-limiting commodity.[335] To overcome this environmental deficiency, fungi and bacteria have evolved high-affinity iron-chelating molecules called siderophores to harvest the iron molecules they require from the environment.[336]

The lipophilic siderophores of mycobacteria called mycobactins were discovered in 1912[337] and isolated in 1953.[338] Subsequent to their discovery, it was envisioned that mycobactin analogues could serve as vehicles for drug delivery to pathogenic mycobacteria.[339] Mycobactins were among the first examples of microbial growth factors, allowing *in vitro* culture of *M. avium* subsp. *paratuberculosis*. After mycobactin was determined to be essential for virulence of *Mtb*,[340] interest in its properties intensified. Mycobactin biosynthesis is now one of the best-characterized nonribosomal polypeptide/polyketide syntheses in nature. Consequently, a host of structurally related siderophores has been discovered in the intervening years, including the water-miscible exochelins of the saprophytic mycobacteria;[341] pyochelin from *Pseudomonas aeruginosa*;[342] the catechol-containing agrobactin[343] from *Agrobacterium tumefaciens*; vibriobactin[344] from *Vibrio cholerae*; anguibactin[345] from *Vibrio anguillarum*; yersiniabactin from yersiniae;[346] the amphipathic marinobactins from marinobacteria;[347] and the nocobactins,[348] nocardichelins,[349] and several other siderophores from nocardiae (see **Table 3**).

1.04.6.2.1 Structure

The core structure of mycobactin (**Figure 30** (**a**)) consists of salicylate linked to an oxazoline ring, followed by a linear lysine residue and a butyrate capped with a seven-membered ring formed out of a cyclized lysine. The linear lysine is acylated at R^1; this is the position of the long alkyl chain in most mycobactins. The butyrate subunit is the most variable section of the molecule between different species and is not derived from a standard amino acid. Both the cyclic and linear lysine residues are N^6-hydroxylated to form the two hydroxamate metal coordination sites. These, together with an oxazoline–phenolate coordination site, serve to make mycobactin a hexadentate metal coordinator with an extremely high ($4 \times 10^{26}\ mol^{-1}\ l$)[350] formation constant for the complex with Fe^{3+}. Each species of mycobacteria has a defined set of R^2–R^6 substituents, whereas the alkyl chain at R^1

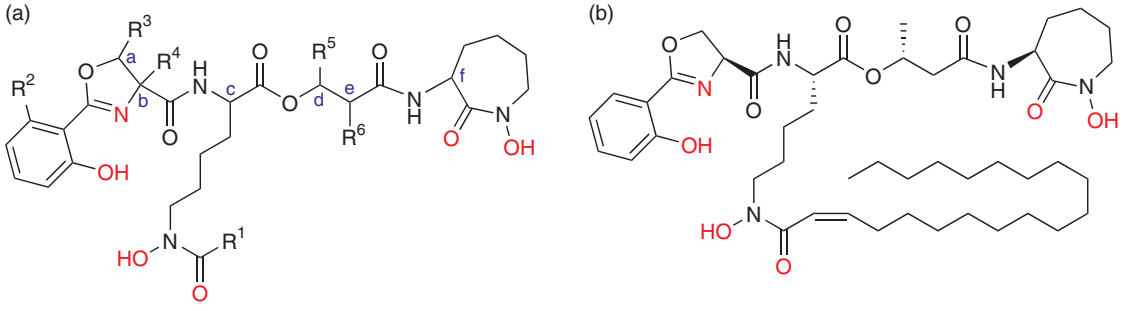


Figure 30 (a) Core structure of mycobactin with iron coordination ligands shown in red and chiral centers in blue. (b) An example of mycobactin T from *Mtb*.

can vary in length (C_{14}–C_{20}) and degree of unsaturation within each species. It is common for the alkyl chain to contain a cis double bond in the α,β-position.[339] **Figure 30(b)** shows an example of a fully substituted molecule of mycobactin T from *Mtb*.

Mycobactin complexes are stable when bound to iron but dissociate in acidic conditions.[351] They are moderately stable to acid, but degrade easily when exposed to base.[339]

The nature of the alkyl tail is critical for the utilization of mycobactins by mycobacteria.[352] Carboxymycobactins (CMBs) have shorter alkyl chains at R^1 (C_2–C_{13}) that terminate in either a carboxylic acid[353] or a methyl ester.[354] CMBs are much more water-soluble than mycobactins, but the distinct functional role of these two forms is still unknown. CMBs were once thought to occur only in pathogenic mycobacteria, but have now been isolated from *M. smegmatis* as well.[355] The siderophores of *M. marinum* and most nocardiae possess alkyl chains at R^5 instead of R^1, and those of nocardiae may have an additional double bond in the oxazoline ring between a and b. The stereochemistries of the chiral centers investigated have thus far conformed to the naturally occurring L-series of amino acids from which mycobactins are derived. The chiral centers d and e result from the action of a polyketide synthase, allowing their chiralities to fluctuate from species to species.

Despite its long, lipophilic tail, mycobactin has a strong amphipathic character. All of the hydrophilic components of the molecule are concentrated in the peptidic head of the molecule, while the lipophilic alkyl tail provides the nonpolar counterpoint. Upon binding iron, many of the hydrophilic ligands are turned inward to coordinate the central iron atom.[374] This causes the molecule to transition into a more compact, less flexible conformation and the external surface to become dominated by the lipophilic character of the alkyl chain. **Figure 31** shows space-filling models of mycobactin T before and after binding Fe^{3+}. Research on related siderophores, the marinobactins, has shown that the desferri form of the siderophore is capable of forming micelles. The marinobactins have a more hydrophilic character overall than mycobactins, but share the attribute of a single C_{12}–C_{16} alkyl chain, making them highly amphipathic. This gives rise to conical molecular shape in which the hydrophilic iron-coordinating ligands are spread out on the surface of the micelle. Upon binding iron, the marinobactins undergo a remarkable phase change to form bilayered vesicles.[347] These vesicles can form aggregates that traffic back to the organism. The length and degree of unsaturation

Table 3 Mycobactins and the siderophores of nocardiae

Species	*Siderophore name*
Mycobacterium aureum	Mycobactin A[351]
Mycobacterium avium subsp. *avium*	Mycobactin Av,[356] carboxymycobactin MA[353,356]
M. avium subsp. *paratuberculosis*[a]	Mycobactin J[357–359]
M. bovis	Carboxymycobactin[360]
M. farcinogenes	Mycobactin F[351,361]
M. fortuitum	Mycobactin F,[351,361] mycobactin H[362]
M. intracellulare	Mycobactin[358]
M. kansasii	Mycobactin K[351]
M. marinum	Mycobactin M, N[351]
M. phlei	Mycobactin P[362]
M. scrofulaceum	Mycobactin[358]
M. senegalese	Mycobactin F[351,361]
M. smegmatis	Mycobactin S,[362] carboxymycobactin MS[355]
M. terrae	Mycobactin R[351]
M. thermoresistible	Mycobactin H[362]
M. tuberculosis	Mycobactin T,[354,362] carboxymycobactin MT[354]
Nocardia asteroides	Nocobactin NA,[363] amamistatin B,[364,365] asterobactin[366]
N. brasiliensis	Nocobactin NB,[367] brasilibactin A[368,369]
N. caviae, *N. phenotolerans*	Nocobactin NC[367]
Nocardia sp. A32030	BE-32030[370]
Nocardia sp. ND20	Formobactin[371]
Nocardia sp. TP-A0674	Nocardimicin[372]

[a] Mycobactin production by *Mycobacterium avium* subsp. *paratuberculosis* is contingent upon selection for survival in iron-free media[373] (#1447); thus, several variations of mycobactin J have been observed.

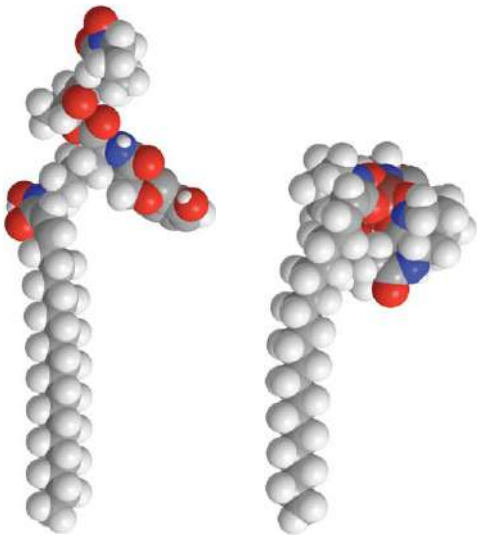


Figure 31 Space-filling models showing the conformation of mycobactin T before and after binding iron.

of the lipophilic tail plays a major role in determining the lipophilicity of the marinobactins, thereby affecting the critical micelle concentration at which marinobactin micelles spontaneously form and bud off from the membrane.[375]

1.04.6.2.2 Biosynthesis

Mycobactin is synthesized through a complex hybrid nonribosomal polypeptide/polyketide biosynthesis process. The biosynthetic pathway of mycobactin was first delineated during analysis of the complete *Mtb* H37Rv genome based on homology with the yersiniabactin biosynthesis genes of *Yersinia pestis*. Cole *et al.*[247] identified the putative mbt mycobactin biosynthesis operon simultaneously with Quadri *et al.*[376] who also confirmed it experimentally. The mycobactin core biosynthesis has been nicely illustrated by De Voss *et al.*[350] The generation of an mbtB knockout mutant of *Mtb* demonstrated that this gene is necessary for mycobactin production, and definitively showed that mycobactin production is required for *Mtb* survival within macrophages.[340] A second operon, mbt-2, was discovered several years later. This operon modifies and attaches the fatty acid that forms the alkyl chain at R^1.[377] The mbt operon contains biosynthesis genes mbtA–H, whereas mbt-2 contains genes mbtK–N. mbtK–N are also referred to as Rv1347c, Rv1344, fadD33, and fadE14, respectively.[377] **Figure 32** illustrates the role of each enzyme in mycobactin biosynthesis.

Note that MbtB, D, E, F, and L must be activated with 4′-phosphopantetheinyl cofactors before they can participate in biosynthesis[376,377] and the specific enzymes that perform this have yet to be conclusively identified. The functions of *mbtH* and *mbtJ* are still unknown, but *mbtH*-like genes in *S. coelicolor* have been found to be necessary for the production of aminocoumarin.[386] The branch point at which the biosynthetic pathways of mycobactin and CMB bifurcate is also unknown.

Both *mbt* operons are regulated by IdeR, an essential iron regulator that suppresses mycobactin synthesis both in the presence of iron and upon exposure of cells to oxidative stress.[387] During oxidative stress, mycobactin synthesis is proposed to be downregulated to limit the iron-catalyzed production of reactive oxygen species through the Fenton reaction. IdeR also regulates other genes involved in iron acquisition and storage.[388,389]

The oxazoline ring has long been thought to be synthesized from serine or threonine. However, the antigen DDM from *Mtb* resembles mycobactin T, but lacks the N^6-hydoxyl groups and incorporates a methyl group in the R^4 position.[385] This suggests the involvement of the uncommon amino acid α-methyl serine in mycobactin biosynthesis. NMR and MS/MS fragmentation analyses of DDM and synthetic analogues confirmed the presence of the methyl group at R^4. The only other report of a substituent at this position can be found in

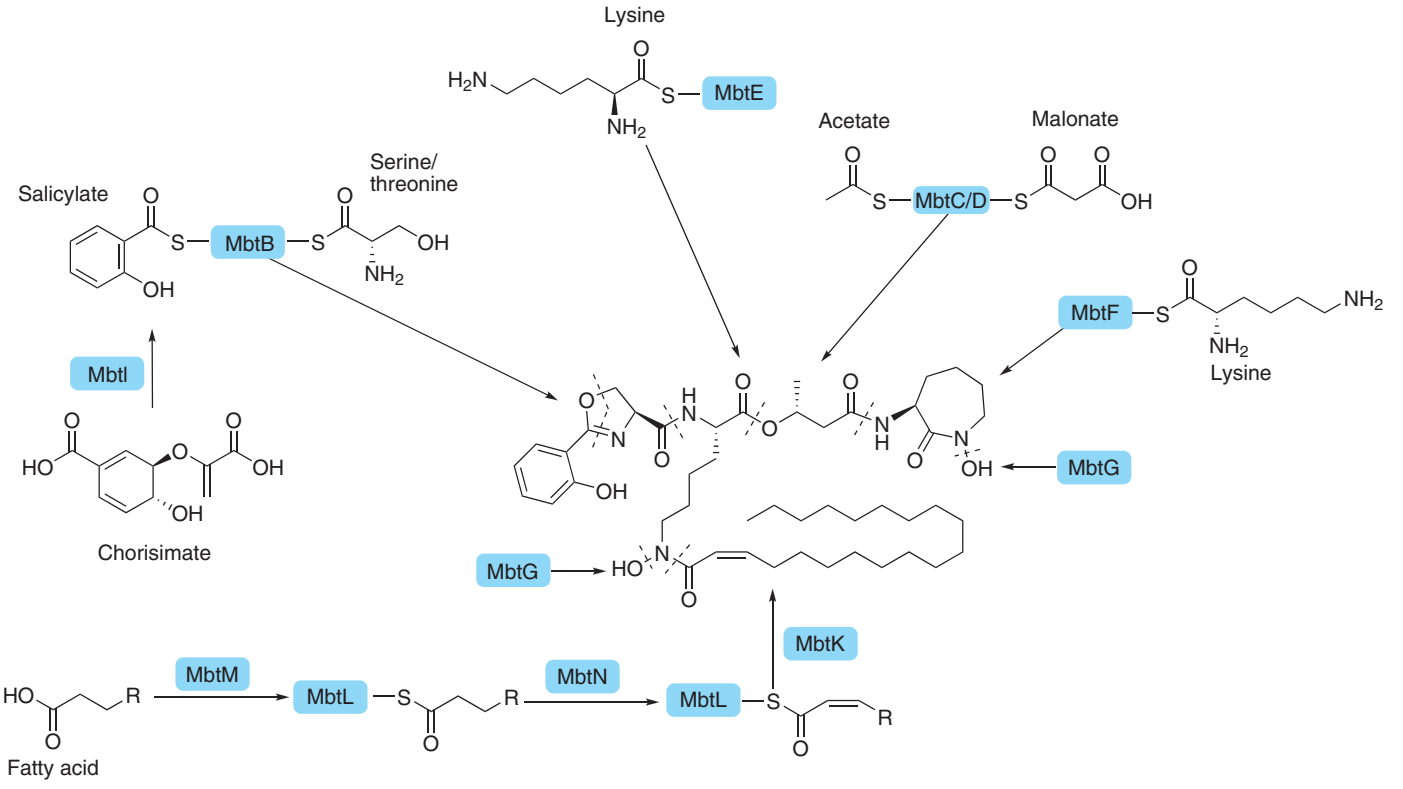


Figure 32 Schematic of mycobactin biosynthesis; enzymes are outlined in blue:

1. MbtI alone synthesizes salicylate from chorismate, with an isochorismate intermediate. This process is dependent upon a Mg^{2+} cofactor.[378,379] 6-Methylsalicylic acid is synthesized through an entirely different polyketide-based biosynthesis from acetate and malonate,[380,381] making the presence of both the methylated and unmethylated salicylate components in *Mycobacterium fortuitum* mycobactins remarkable.[351]
2. MbtA adenylates salicylic acid to create salicyl-AMP, which then forms a thioester with MbtB.[376]
3. MbtB catalyzes the formation of the peptide bond with either serine or threonine, followed by cyclization of the peptide and final dehydration to form the oxazoline ring. The serine/threonine-activating domain of MbtB has shown preference for both serine and threonine.[382]
4. The MbtC/D complex synthesizes the butyrate subunit from acetate and malonate[350] or from propionate depending on the species.[383] The butyrate is then transferred to MbtE.
5. MbtE forms the peptide bond between the salicyl-oxazonline and linear lysine subunit, and the ester bond linking the butyrate subunit (the exact order of these steps is unknown).[376]
6. MbtF attaches and cyclizes the final lysine.[376]
7. MbtM activates a fatty acid by adenylation and transfers it to MbtL.[377]
8. MbtN dehydrogenates the alkyl chain while it is bound to MbtL.[377]
9. MbtK acylates the linear lysine residue, attaching the lipophilic alkyl chain.[384]
10. MbtG is thought to hydroxylate the lysine residues as the final step in mycobactin synthesis. This hypothesis is supported by evidence that N^6-hydroxylysine is neither incorporated into mycobactin[383] nor is it a substrate for acylation by MbtK.[384] It is further supported by the discovery of DDM,[385] which is presumed to be an intermediate in mycobactin synthesis.

the siderophores isolated from *Nocardia* sp. A32030 that also exhibit a methyl group at R^4. It is possible that this methyl group is cleaved to form the finished mycobactin T structure, but this seems unlikely. Both of these structures were elucidated using modern NMR technology not available during the initial elucidation of the mycobactin T structure.[390]

An MbtM orthologue knockout (*fadD33* in *M. smegmatis*) produced 90% less mycobactin than the wild type, and only produced mycobactins containing a saturated alkyl chain. This suggests that MbtK is capable of attaching alternative substrates to mycobactin that do not result directly from the *mbt-2* pathway. However, it should be noted that complementation with the *mbtM* gene from *Mtb* failed to restore wild-type mycobactin production in *M. smegmatis*, so the two orthologues may have different functions.[391]

p-Aminosalicylic acid (PAS) was one of the standard drugs used to treat tuberculosis before the introduction of rifampicin and pyrazinamide (PZA).[392] PAS is thought to act in part by inhibiting mycobactin synthesis. PAS has been shown to decrease mycobactin production,[393] but confoundingly, increased the production of CMB in *M. smegmatis*.[394] The mechanism through which this occurs is unknown and invites further investigation. PAS may interact with an enzyme involved with the differentiation between mycobactin and CMB.

Wild-type *M. avium* subsp. *paratuberculosis* is dependent on exogenous mycobactin for growth in the laboratory. Interestingly, when an IS*900* gene repeat was transformed into *M. smegmatis*, the property of mycobactin dependence was conferred unto the recombinant strain. IS*900* encodes p43, a protein with unknown function. The mechanism through which mycobactin dependence is conferred is unknown, but merits further investigation as a route toward mycobactin inhibition. p43 may be responsible for the mycobactin dependence and slow growth rate of *M. avium* subsp. *paratuberculosis*.[395]

1.04.6.3 Iron Acquisition in Mtb

Bacteria use siderophores to acquire iron when it is otherwise too scarce in the environment. In the case of pathogenic mycobacteria, the scarcity is compounded by two factors: first, that Fe^{3+} has very low solubility at physiological pH,[396] and second, that the natural host response to infection includes sequestration of available iron. Sequestration is accomplished by upregulation of the iron-binding proteins transferrin, lactoferrin, and ferritin, and internalization of iron by host cells.[397]

Mycobactins can compete directly for iron with these three proteins to overcome host sequestration.[398] There is also evidence to suggest that pathogenic mycobacteria can manipulate the host to traffic iron into infected cells.[399,400] Aside from being an essential cofactor in many processes, iron is especially important for production of many mycobacterial lipids. Testament to this is the requirement of iron supplementation for the formation of *M. smegmatis* biofilms.[401] The aforementioned *mbtB* knockout mutant showed attenuated growth in low-iron conditions and in macrophages.[340] Another mutant that constitutively expressed a homologue of the iron repressor IdeR also showed attenuated growth in BALB/C mice.[402] Gene expression data has shown that mycobacteria are exposed to iron restriction in human macrophages[403] and mice,[404] and similar data supporting this hypothesis have been reviewed by Waddell and Butcher.[405] Furthermore, iron supplementation has been associated with the exacerbation of mycobacterial and other infections in experimental models[406,407] and clinical practice.[408–410] A modern analysis of data from a pre-HIV 1929 study indicated that iron overload significantly increased the risk of death from tuberculosis.[411] All these data support that iron availability is a limiting factor for the progression of mycobacterial infections.

Despite this compelling evidence for the importance of mycobactin, the *mbtB* mutant is still able to grow in culture medium supplemented with exogenous iron,[340] suggesting that some alternative form of iron acquisition exists that has yet to be characterized. There is evidence that the xenosiderophore ferricrocin and fungal siderophore rhizoferrin are internalized by *M. smegmatis*.[412,413] It is also possible that mycobacteria can take up iron from other common xenosiderophores and coordinating compounds that might be found in the immediate environment; however, this does not seem to be as effective as mycobactin-mediated iron acquisition in iron-restricted environments.

1.04.6.3.1 Current models for mycobacterial iron trafficking

Conventionally, mycobactins were thought to be associated only with the cell wall and membrane of the mycobacteria. They have low solubility in water (5–10 $\mu g\,ml^{-1}$) and are typically extracted from the cell pellet,

not from culture supernatants.[339] They have been reported to be entirely cell-associated when grown in the absence of the detergent Tween, but can be found in culture supernatants when Tween is added to the growth medium.[414] They have also been observed by electron microscopy to exist within the cell wall near the cytoplasmic membrane.[415] Accordingly, mycobactin was presumed to function simply as a means of ferrying iron across the thick mycobacterial cell wall and membrane.[416]

'Exochelins',[341] discovered in 1975, are water-soluble siderophores consisting of two types: the true peptidic exochelins of the saprophytic mycobacteria that are structurally unrelated to mycobactins, and the chloroform-extractable CMBs. The structures of the CMBs were elucidated in both *M. avium*[353] and *Mtb*[354] simultaneously and determined to be a variation of the mycobactins. To date, only *M. smegmatis* has been shown to produce both exochelins and CMBs.[417]

The discovery of extracellular siderophores brought into question the importance of lipophilic mycobactin. Models were proposed in which CMBs/exochelins scavenged iron from the extracellular milieu, then transferred it into the cell by means of a putative membrane-spanning reductase. In the case of excess iron and high concentrations of ferric CMBs/exochelins, the surplus may be donated to mycobactin residing in the cell membrane.[418,419] However, the growth-promoting effect of mycobactin could not be ignored, and more recent studies have shown that mycobactins are more vital to iron acquisition than are exochelins. This was demonstrated by a mycobactin-deficient *M. smegmatis* mutant that was fivefold more susceptible to growth inhibition by the iron chelator EDDA (ethylenediamine-*N*,*N*′-di(2-hydroxyphenylacetic acid)) than either the wild-type or exochelin-deficient mutant.[391]

Recently, an ABC transporter system consisting of two proteins, IrtA and IrtB, was found to be required for effective utilization of ferric CMB, but not for its secretion. The same study also found that mycobactin is not required for iron uptake from ferric CMB,[420] but this was based on the uptake from culture filtrates, not purified CMB. Another study concluded that IrtA specifically exports siderophores whereas IrtB in conjunction with Rv2895c imports ferric siderophores,[421] but this has yet to be corroborated by follow-up studies. It is unknown whether the IrtAB system transports lipophilic mycobactins in addition to CMBs.

Luo *et al.*[422] showed that the lipophilic mycobactins were able to diffuse freely in and around macrophages when added exogenously, and that gallium-loaded mycobactins were found preferentially in lipid bodies in direct contact with phagosomes. Interestingly, recent studies have shown that oxygenated mycolic acids from *Mtb* and *M. avium* induce the formation of 'foamy' macrophages that are filled with many lipid bodies. *Mtb* living within the phagosomes of foamy macrophages were shown to colocalize with, and subsequently engulf these lipid bodies.[423] Not only does this provide the bacteria with a lipid-rich environment in which to persist,[423] but it may represent a new route for trafficking iron that involves subverting host lipid bodies to serve as reservoirs of nutrients and ferrimycobactins.[424]

These findings complement research discussed earlier on the formation of marinobactin micelles,[347,375] and the findings of Xu and comrades that mycobacterial lipids in *Mtb*- and *M. avium*-infected macrophages are capable of trafficking out of the mycobacterial vacuole.[425] A different scenario for mycobacterial iron acquisition emerges, in which the lipophilic mycobactins act as extracellular siderophores, and iron is trafficked through intracellular lipid bodies back to phagosomes infected by *Mtb*.[426]

Figure 33 shows representations of mycobactin-mediated and CMB-mediated acquisition of extracellular iron in pathogenic mycobacteria. **Figure 33(a)** depicts a model of mycobactin-mediated iron acquisition within macrophages that takes advantage of lipid body trafficking.[422] Mycobactins leave the bacterium either by trafficking out of the phagosome with other mycobacterial lipids[425] or by forming micelles.[375] In order to cross the thick cell wall, it might be necessary for mycobactins to use a specific mechanism that has yet to be discovered. The micelles diffuse throughout the macrophage and bind iron from the macrophage iron pool, including iron plundered from ferritin and imported transferrin.[398] Ferric mycobactins then traffic preferentially to macrophage lipid bodies[422] that are engulfed by infected phagosomes.[423] The ferric mycobactins translocate through the cell wall once again. At this point the ferric mycobactins may exchange iron with CMBs on an equilibration basis.[398] This iron can then be imported through the IrtAB system. Alternatively, the mycobactins may diffuse across the membrane or handoff their iron payload to other mycobactins within the cell membrane. Because membrane affinity is modulated by the length and degree of unsaturation of in the alkyl chain,[375] these recipient mycobactins in the membrane could form a gradient of mycobactins with differential membrane affinities. In this way they could assemble a 'bucket brigade' to transport the iron into the bacterium.[427]

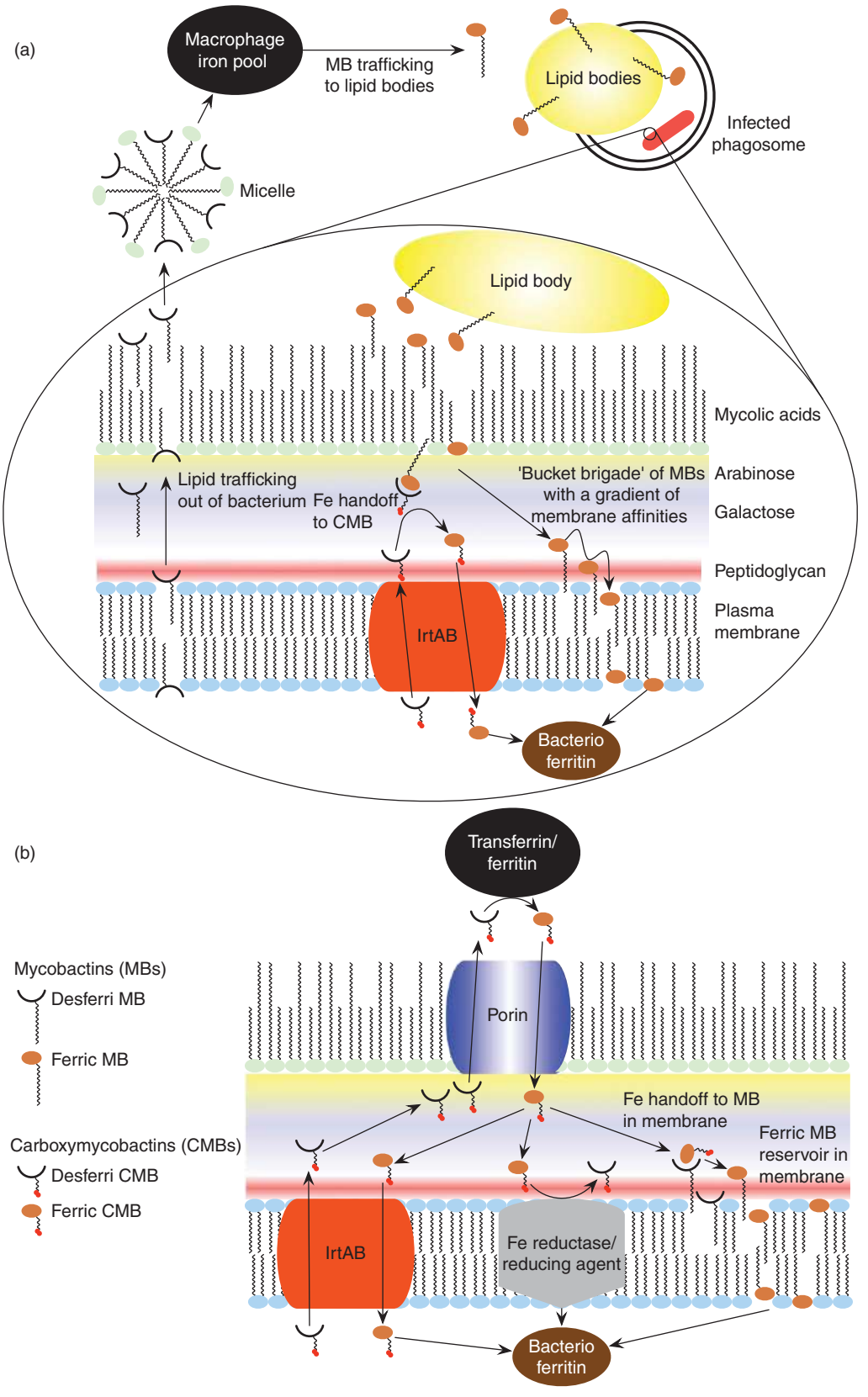


Figure 33 Models of mycobacterial iron trafficking. (a) Lipophilic mycobactin-mediated iron acquisition utilizing lipid body trafficking within macrophages. (b) An alternative carboxymycobactin-mediated iron acquisition.

Figure 33(b) depicts a model of CMB-mediated iron acquisition. CMBs are transported out of the bacterium by means of IrtAB and a putative porin-based mechanism[419] or other as-yet undefined method. The CMBs then scavenge iron from host iron-binding proteins, either within a macrophage or in the relatively aqueous extracellular milieu. Ferric CMBs are then either transported back into the cell through IrtAB, are relieved of iron by an undetermined reductive process in the periplasm, or hand excess iron off to mycobactin. Mycobactins in the cytoplasmic membrane create a reservoir of readily accessible mycobactin-bound ferric iron.[415]

The lipid-rich environment within a foamy macrophage is more conducive to diffusion of lipophilic mycobactins. Mycobacteria residing outside of such a lipid-rich environment may require a more water-friendly method of iron acquisition. It can be hypothesized that lipophilic mycobactin is the principal iron scavenger within the lipid-rich environment of the macrophage, whereas CMB serves as the principal scavenger for bacteria living in the relatively aqueous extracellular milieu. The reality is likely to be a hybrid of these mechanisms, adapted by the bacterium to its particular condition and environment. Many components of this model are speculative; it does not account for normal secretion of CMBs in the IrtAB knockout[420] or for the translocation of mycobactin across the formidable mycobacterial cell wall. The trafficking of mycobactins, CMBs, and iron within mycobacteria will likely prove to be much more complex, pending the discovery of other siderophore transport mechanisms.

1.04.6.3.2 Mycobactins and their analogues as antimicrobials

Mycobactin analogues have been contemplated for use as either mycobactin synthesis inhibitors or vehicles for infiltrating the mycobacterial cell wall to deliver drugs since 1945.[339] Hu and collaborators synthesized mycobactin S and unexpectedly found that it inhibited *Mtb* growth with 99% growth inhibition (MIC_{99}) at 12.5 $\mu g\,ml^{-1}$.[428] The remarkable part of this story is that mycobactin S is identical to mycobactin T except for the chiral center at d (see **Figure 30**). It should be noted that this inhibitory activity has not been confirmed with naturally produced mycobactin S. Miller's laboratory synthesized more mycobactin analogues, the most active of which included a Boc-protected amino group at R^5 and demonstrated an MIC_{99} of less than 0.2 $\mu g\,ml^{-1}$.[429]

Transvalencin Z is a compound that was isolated from *Nocardia transvalensis*. It resembles the salicyl-oxazoline half of mycobactin up to the ester bond and lacks the long acyl chain. This compound was shown to have antimicrobial activity[430] and may act by interfering with the siderophore pathways on organisms with siderophores containing salicyl-oxazoline or salicyl-thiazoline moieties. Cross-reactivity of this structural feature is tentatively supported by the finding that mycobactin S and CMB S induce the expression of *P. aeruginosa* extracytoplasmic function sigma factor PA1910.[431] Although the function of PA1910 is not fully understood, the *Pseudomonas* siderophore pyochelin also contains a salicyl-thiazoline.[342] This suggests the possibility of siderophore cross-reactivity or inhibition. Alternatively, a similar fragment of amamistatin B was synthesized and shown to inhibit histone deacetylase as an antitumor agent.[432]

The idea of a siderophore 'Trojan horse' has also gained traction as a way to transport drugs that would otherwise fail to penetrate the mycobacterial cell wall. A number of these compounds coupled to β-lactams have successfully been synthesized.[433] Efforts to create therapeutic mycobactin analogues have also focused on creating competitive iron chelators to starve the bacteria of iron.[434]

The elucidation of the mycobactin biosynthetic pathway initiated the search for inhibitors against enzymes within the pathway. In the quest for an inhibitor of salicylate-activating enzymes (examples include MbtA in *Mtb* and YbtE in *Y. pestis*), two independent research groups generated the same bisubstrate inhibitor: salicyl-5′-*O*-(*N*-salicylsulfamoyl)adenosine (salicyl-adenosine monosulfate (AMS)), as depicted in **Figure 34(b)**.[435,436] This inhibitor is a modification of the activated salicyl-AMP (**Figure 34(a)**) and mimics both substrates of MbtA, occupying both binding pockets of the enzyme. Salicyl-AMS showed inhibitory activity against both *Mtb* and *Y. pestis* under low-iron conditions, but only against *Mtb* under iron-supplemented conditions.

Following the design of salicyl-AMS, Courtney Aldrich and coworkers embarked on a systematic exploration of the salicyl-AMS structure–activity relationship (SAR). They synthesized a multitude of analogues with different modifications to the salicyl–glycosyl linker,[436,437] glycosyl subunit,[438] salicyl subunit,[439] and adenosine subunit[440,441] of salicyl-AMS. Luis Quadri and associates have also examined analogues of this compound.[435,442] The results suggested that the ribose subunit is critical for recognition by a putative transporter.[438] This could explain the ability of this compound to overcome the membrane permeability problem that has plagued other acyl adenylate-based inhibitors.[443] Salicyl-AMS analogues were designed rationally with the help of a complete

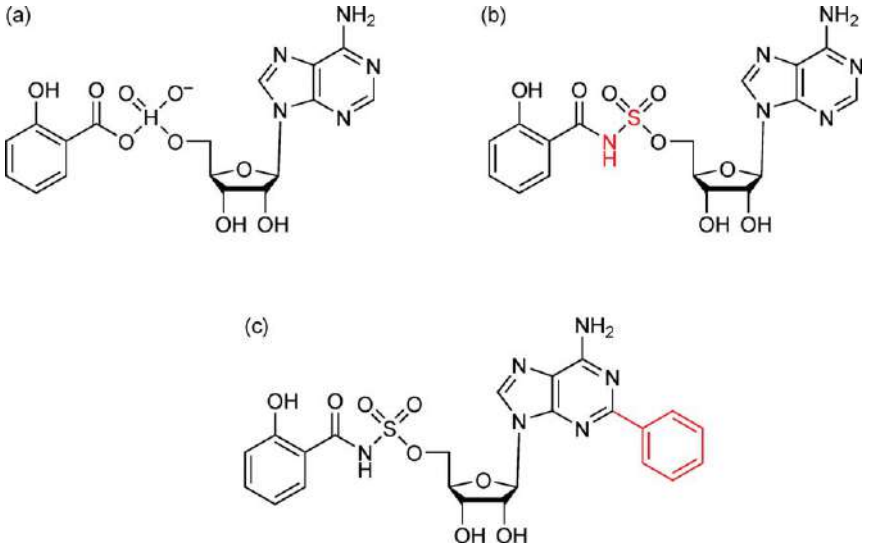


Figure 34 (a) Salicyl-AMP, the natural product of MbtA; (b) salicyl-AMS, the MbtA bisubstrate inhibitor; and (c) 2-phenyl-salicyl-AMS, optimized inhibitor.[440]

homology model of MbtA based on the crystal structure of DhbE from *Bacillus subtilis* complexed with an adenylated 2,3-dihydroxybenzoic acid. This epic quest led to a compound exhibiting an MIC_{99} of 0.39 μmol l^{-1} in iron-replete conditions and 0.049 μmol l^{-1} in iron-deficient conditions.[440] This compound is pictured in **Figure 34(c)**. The potency of this compound and activity under both iron-replete and iron-deficient condition suggests that it may be active in additional pathways beyond MbtA. Salicyl-AMS has also been tested for stability for up to 24 h in human plasma and 1 h in human liver microsomes.[444]

1.04.6.3.3 Other therapeutic uses of mycobacterial siderophores

CMBs[445,446] and the nocardial siderophores[370,447] have been repeatedly examined for use as antitumor agents, chiefly because their strong iron-chelating activity is able to deprive cancer cells of iron. This has been shown to inhibit DNA replication by limiting iron availability to ribonucleotide reductase. CMB T (referred to as exochelin) was able to induce apoptosis of human breast cancer cells without harming normal breast tissue. CMB T outperformed desferoxamine as a ribonucleotide reductase inhibitor; the greater effect was attributed to the lipophilicity of CMB enabling it to penetrate the cell membrane and scavenge intracellular iron, whereas desferoxamine was only able to chelate extracellular iron. Practically speaking, this would require local delivery of CMB, since any CMB delivered systemically would almost certainly have chelated iron before reaching the target tumor cells.

The amamistatins from *Nocardia asteroides* were also investigated for their antitumor activity.[364,365] Several analogues of these were subsequently synthesized and tested.[432,448,449] Interestingly, the amamistatin fragment mentioned earlier showed antitumor activity and histone deacetylase inhibition, which did not seem to be related to iron chelation activity. Some mycobactin analogues were also found to have monoamine oxidase (MAO) inhibitory properties as well as some antitubercular activity.[450] This is reminiscent of the weak MAO inhibitory activity of isoniazid[451] and was cited as a possible complicating factor for the development of antimicrobials based on these scaffolds. Siderophores from *Nocardia* sp. TP-A0674 showed inhibitory activity of muscarinic M3 receptor, an important neurotransmitter.[372] The muscarinic M3 receptor is a potential target for treatment of a variety of disorders including respiratory, gastrointestinal, and urinary tract disorders.

1.04.6.3.4 Immunology of mycobactin

It was not known whether mycobactins participated in the adaptive immune response until the discovery of the antigen DDM. Moody *et al.* showed that DDM is recognized by the CD1a antigen display protein and serves as a

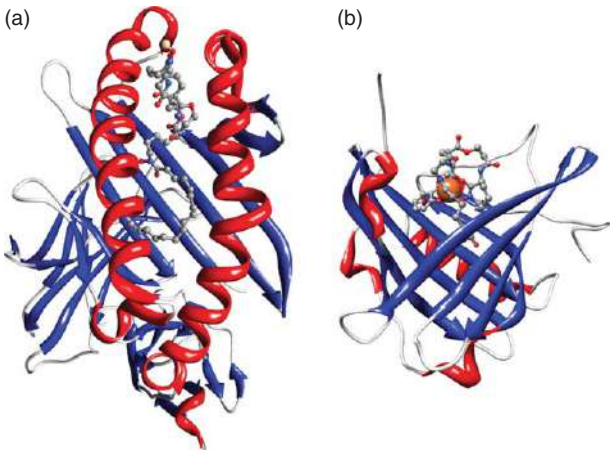


Figure 35 (a) DDM analogue in the binding pocket of CD1a (PDB 1XZ0).[452] (b) Lipocalin 2 bound to ferric carboxymycobactin T (PDB 1X8U).[453] Both images were rendered using Chimera.[199]

potent T cell activator.[385] Natural mycobactin did not activate T cells, and T cell activation seemed to be sensitive to the length of the alkyl chain at R^1. T cell activation by DDM may play an important role in the host immune response and the formation of granulomas. CD1a has subsequently been cocrystallized with a synthetic DDM analogue.[452] The crystal structure of this analogue in the binding pocket of CD1a is shown in **Figure 35**(**a**).

Lipocalin 2, also called siderocalin,[453] neutrophil gelatinase-associated lipocalin,[454] and several other names,[425,455] is a secreted human neutrophil granule protein.[456] It is secreted during bacterial infection as part of innate immunity[457] and binds several bacterial siderophores including the CMBs from *Mtb* and *M. smegmatis*,[453] enterobactin from *E. coli*,[458] and parabactin from *Paracoccus*.[453] In this way, lipocalin 2 sequesters siderophore-bound iron to prevent harvest by the bacterium. **Figure 35(b)** shows the crystal structure of lipocalin 2 bound to ferric CMB T.

Lipocalin 2 is released from neutrophil granules as a monomer or disulfide-linked homodimer or heterodimer with neutrophil gelatinase-B.[454,459] Neutrophil gelatinase-B is also known as matrix metalloproteinase-9 (MMP-9), which interestingly has been shown to play a role in granuloma development.[460] Serum concentrations of lipocalin 2 have been recorded as high as 404 ng ml^{-1} during bacterial infection and can be used to differentiate between bacterial and viral infections.[461]

Lipocalin 2 has been shown to bind CMB with a carboxylated alkyl chain containing eight methylene groups at R^1. It may also be able to bind CMBs of slightly differing chain lengths with reduced affinity.[453] Previously, only chains containing up to seven methylene groups had been reported.[354] Lipocalin 2 knockout mice have greater susceptibility to infection by *E. coli*[457] and *Mtb*.[462] Intratracheal infection with *Mtb* yielded significantly increased *Mtb* growth in alveolar epithelium, but not inside macrophages. The lipocalin 2 knockout mice died of infection more quickly than wild type.

It has been suggested that bacteria may further modify their siderophores specifically to elude capture by lipocalin 2.[463,464] Although this has yet to be experimentally demonstrated, it could account for the diverse structural variability of mycobactins, even within a single species. Enterobactin undergoes a shift in coordination chemistry upon acidification that results in a conformational change in the molecules. This shift is sterically incompatible with the lipocalin 2 binding pocket, leading to release of the iron and degraded siderophore.[465] It has yet to be shown whether or not a similar release of iron might occur with CMB. There is no evidence that CMB undergoes a similar structural change, but CMB does dissociate from iron at low pH.[351] A recent study examined the role of electrostatics in siderophore capture by lipocalin 2. The authors concluded that electrostatic attraction between the siderophore and lipocalin 2 binding pocket is important for siderophore binding. Because of this, catecholate siderophores are the natural prey for lipocalin 2, while hydroxamate siderophores do not bind as well because the complex forms a neutral molecule.[466] Conversely, mycobactin utilizes a primarily hydroxamate chelation modality and forms a neutral complex with Fe^{3+}.

Nevertheless, lipocalin 2 binds CMBs S and T with a high affinity, and the lipocalin 2 binding pocket is more completely filled by CMB than by the catecholate enterobactin.[453]

Mycobactins have been studied extensively over the course of the past 100 years, yet many mysteries still surround their exact role and function in mycobacterial biology and disease. The importance of mycobactin in pathogenic mycobacterial diseases is unassailable and the evidence of an age-old arms race between the pathogen and the host is fascinating. In spite of this, mycobactin remains an elusive creature that has never successfully been observed *in vivo*.[467] The diligence of future research has much more to reveal about mycobactin and its nefarious iron-smuggling operations.

1.04.6.4 Mycolactones

Mycolactones are a group of polyketide synthase-derived macrolide toxins produced by *M. ulcerans*. Mycolactone secretion is the primary virulence determinant for *M. ulcerans* infection, known as Buruli ulcer. Mycolactone appears to be solely responsible for the pathology of this disease, causing necrosis, apoptosis, and immunosuppression of host cells. Buruli ulcer is an emerging disease that has been reported in a growing number of countries across the globe, most notably in Australia, Southeast Asia, and West Africa, where it has become a major public health problem. In endemic areas of West Africa, the prevalence of Buruli ulcer can exceed that of tuberculosis and leprosy.[468] Buruli ulcer is rarely lethal, but often leads to severe disfigurement and disability.

Mycobacterium ulcerans infection can result from invasion of a low number of *M. ulcerans* bacilli[469] and incubation of the infection has been estimated to last anywhere from 1 to 7 months.[470–472] The role of mycolactone in the progression of Buruli ulcer is depicted in **Figure 36**. The infection starts as a subcutaneous nodule, or as a shallower raised papule in Australia,[468] where a different form of mycolactone renders *M. ulcerans* less virulent. Infection occurs most often on the arms and legs because of the lower optimal growing temperature of *M. ulcerans*.[473] Evidence suggests that *M. ulcerans* has an intracellular stage, 3–24 h after introduction into the host.[474,475] Many mycobacteria are phagocytosed during the initial immune response, but once mycolactone production is established the toxin destroys the engulfing phagocytes. Mycolactone knockout strains are more easily internalized by phagocytes.[476] Subsequent mycolactone production inhibits phagocytosis, the inflammatory response, and adaptive immunity.

If untreated, the infection spreads, forming necrotic lesions within the subcutaneous adipose tissue. Eventually, the overlying tissue will ulcerate to form a painless necrotic ulcer with undermined edges. Mycolactone is required for the infection to progress to ulceration.[477] Ulcers can spread extensively to cover large patches, leading to the destruction of limbs and joints. The bacteria grow as extracellular microcolonies with multiple foci spread throughout the base of the lesion. Although inflammatory infiltrates can surround the expanding lesion, the necrotic zone remains free from inflammatory cells and granulomas.[478] The WHO now recommends antibiotic therapy (rifampicin, streptomycin) in the treatment of Buruli ulcer[479]; however, some infections do not respond to antibiotic therapy and wide excision of the ulcer followed by skin graft remains the only effective treatment in these cases.[480] Antibiotic treatment reverses the immunosuppressive effects of mycolactone and leads to a local cellular immune response around necrotic lesions, the formation of

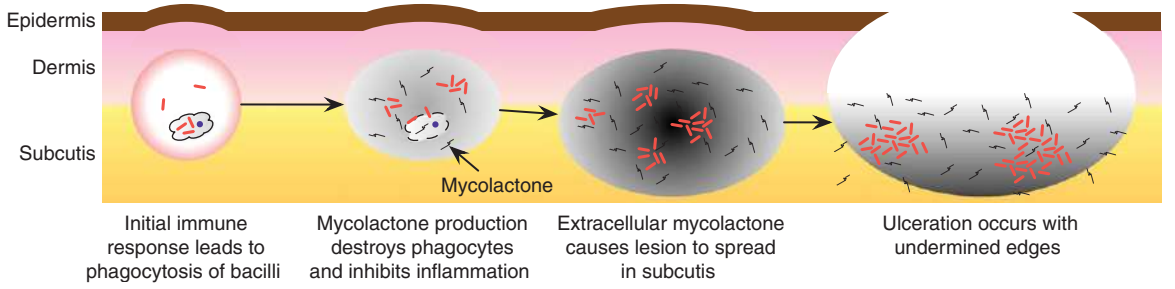


Figure 36 The role of mycolactone in the progression of Buruli ulcer.

granulomas, and phagocytosis of mycobacteria.[481,482] In small lesions, this inflammatory response can clear the infection and allow healing. Granulomas also form in the late ulcerative stage of the disease, near the phase of spontaneous clearance of the infection.[483] This corresponds to an increase in IFN-γ and a decrease in bacillary load.

Buruli ulcer is not spread from human to human. It is thought to spread to humans from an unknown vector in the environment. The disease is associated with proximity to slow-moving water and Buruli ulcer outbreaks often occur following local disturbance of the water table. Classically, Buruli ulcer has been known to be coincident with local trauma at the infection site, including the site of immunizations and scratches.[484,485] *Mycobacterium ulcerans* has been detected in the salivary glands of Naucoridae aquatic insects, and experimentally infected insects transmit the infection to mice through biting. Mycolactone is required for colonization of the salivary glands of these insects.[486,487] Biofilm formation is important in the *M. ulcerans* life cycle and has been associated with aquatic plants[488] and the raptorial legs of Naucoridae where they secrete saliva into their prey.[486] Biofilms formed by *M. ulcerans* create a large reservoir of mycolactone within the extracellular matrix and vesicles from the matrix are highly cytotoxic. Extracellular matrix also enhances colonization of both insect vectors and mammalian hosts and serves as an inhibitory barrier to the passage of antibiotics.[489] In contrast to this evidence, aquatic insect vectors for *M. ulcerans* infection have been called into question by study of their environmental distribution.[490] *Mycobacterium ulcerans* has also been detected in mosquitoes in the vicinity of a Buruli ulcer outbreak in Australia,[471] and prevention from insect bites decreases the risk of outbreaks.[491,492] The first isolation and full characterization of *M. ulcerans* from the environment was recently published.[493] The bacterium was isolated from a water strider (sp. *Gerris remigis*) in Benin, produces mycolactone, and is capable of causing Buruli ulcer in mice.

Definitive studies on the relationship between *M. ulcerans* exposure and contraction of Buruli ulcer are lacking. A preliminary serological analysis of patients revealed that 75% of patients and 38% of their healthy household contacts had circulating antibodies to *M. ulcerans* antigens, whereas controls from nonendemic areas did not.[494] This suggests that many people exposed to the bacterium do not develop Buruli ulcer, possibly by resolving the infection in its early stages.

Analysis of the *M. ulcerans* genome indicated that it recently evolved from *M. marinum* and is in the process of adapting to occupy a dark, aerobic niche requiring mycolactone.[495,496] Sequence analysis across an array of mycolactone-producing mycobacteria allowed the creation of a phylogenetic tree. The result indicated that acquisition of the mycolactone synthesis plasmid drove the speciation from a common *M. marinum* progenitor into a human pathogen.[497,498]

1.04.6.4.1 Structure

Figure 37 shows the mycolactone core structure and **Table 4** lists the side chains that have been described. The mycolactone core consists of a 12-membered ring and linear carbon chain containing two hydroxyl groups. The core is highly conserved, whereas the ester-linked side chain at R is variable between strains.

A survey of mycolactone-producing strains of *M. ulcerans* showed that mycolactone A/B was the most abundant mycolactone produced by West African and Malaysian isolates and that trace amounts of mycolactone A/B can be found in nearly all strains of *M. ulcerans*. Mycolactone C was the most abundant in Australian isolates and D was the most abundant from a Chinese isolate.[501] Japanese *M. ulcerans* strain 8756 produced

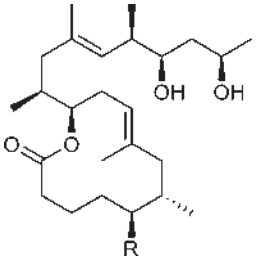
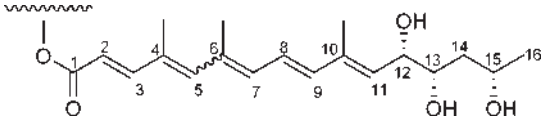
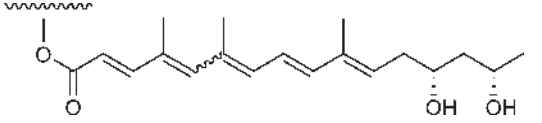
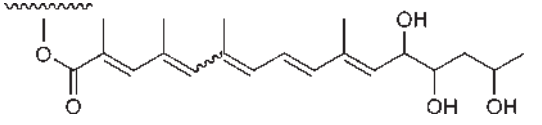
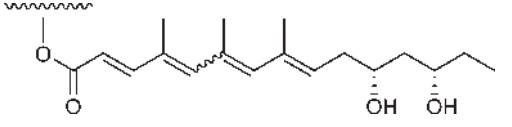
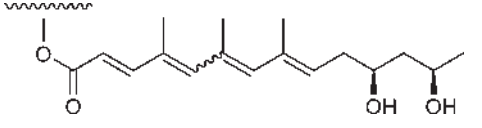
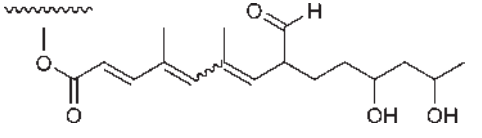


Figure 37 Mycolactone core structure.

Table 4 Mycolactone side chains from different strain and species of mycobacteria; all side chains reported include both *Z*- and *E*-$\Delta^{4,5}$ isomers in differing ratios

Mycolactone	*Species*	*R*	*Reference(s)*
A/B A = *Z*-$\Delta^{4,5}$ B = *E*-$\Delta^{4,5}$	*Mycobacterium ulcerans* (West Africa, Malaysia, Japan)		477, 499, 500
C	*M. ulcerans* (Australia)		501, 502
D	*M. ulcerans* (China, Japan)		501, 503
E	*Mycobacterium liflandii*		504–506
F	*Mycobacterium marinum, M. pseudoshottsii*		507, 508
G	*M. marinum* (engineered)		509

mycolactone D as its major metabolite, but more recent genetic analysis has suggested that Japanese strain 753 (*M. ulcerans* subsp. *shinshuense*) contains a plasmid encoding for the production of mycolactone A/B.[510] Thus, different strains from Japan may produce either mycolactone A/B or mycolactone D as major metabolites.

In 2004, 52 isolates of mycobacteria were screened for the presence of mycolactone. Several species produced cytopathic lipids less potent than mycolactone, but none produced a similar cytopathology in L929 cells.[511] In 2005, the newly described species *Mycobacterium liflandii*, a frog pathogen, was found to produce mycolactone E.[504] In 2006, globally distributed isolates of the fish pathogens *M. marinum* and *Mycobacterium pseudoshottsii* were found to produce mycolactone F. Mycolactone-producing strains of *M. marinum* and *M. pseudoshottsii* will not grow above 30 °C,[507] making them less likely to infected a human host than *M. ulcerans*, which has an optimal growing temperature of 30–33 °C. These strains also included insertion sequence (IS) *2404*, which had previously been used as an *M. ulcerans*-specific PCR marker. Synthetic studies have proven invaluable in deducing the complete structure and stereochemistry of the mycolactones, including resolving the conflicting proposed structures for mycolactone E.[506]

1.04.6.4.2 Function

Mycolactone is solely responsible for the pathology of *M. ulcerans* infection. Mycolactone A/B reversibly arrests the growth of murine fibroblasts at 25 pg ml^{-1}.[477] At higher concentrations, cell death occurs within 72 h and 20–30% of cell death is mediated by apoptosis.[512] This is supported by high rates of apoptosis found in clinical Buruli ulcer lesions.[513] Subcutaneous injection of purified mycolactone in guinea pigs mimics the pathology of Buruli ulcer, forming painless open lesions with minimal inflammation.[477] This is characteristic of *M. ulcerans* infection and contrasts with dermal *Mtb* or *M. marinum* infections, which induce an acute inflammatory response and form granulomatous, pus-filled lesions. Mycolactone-deficient *M. ulcerans* mutants also formed pus-filled lesions, but wild-type pathology was restored when the mutant cells were coated with exogenous mycolactone.[476]

Synthetic fluorescent mycolactone analogue localizes in the cytoplasm of infected cells.[514] Mycolactone causes cytoskeletal rearrangement[512] and an increase in intracellular calcium. This calcium increase may arise through damage to calcium-containing compartments of the cytoskeleton rather than the activation of cell signaling pathways.[514] Mycolactone is believed to cause the lack of pain associated with Buruli ulcer lesions by inflicting nerve damage. *Mycobacterium ulcerans* infiltrates and damages the nerve bundles surrounding lesions and causes degeneration of myelin-forming Schwann cells.[515] Mycolactone alone causes nerve damage including intraneural hemorrhage, loss of Schwann cell nuclei, and a decrease in myelin fiber density in the mouse-footpad model, leading to painless erosion of the footpad.[516]

The first reports of immunosuppression by mycolactone preceded its isolation when these properties were observed from the culture supernatant.[517] The acetone-soluble lipids from *M. ulcerans* culture filtrate inhibited lipopolysaccharide (LPS)-induced release of TNF-α and IL-10 from human monocytes and production of IL-2 from activated T lymphocytes. Mycolactone also interfered with TNF-α-induced NF-κB activation.[518] Production of macrophage inflammatory proteins 1α and 1β were strongly inhibited in dendritic cells, thus preventing recruitment of inflammatory cell to the site of infection.[519] Production of mycolactone by intracellular *M. ulcerans* also inhibited the production of macrophage inflammatory protein 2 in a dose-dependent manner.[520] Mycolactone has been detected by liquid chromatography (LC)–MS/MS in the spleen, kidneys, and liver of mice infected with a subcutaneous injection of 10^4 *M. ulcerans* to the tail.[521] Whole blood cells of infected mice also showed inhibited ability to secrete IL-2, whereas mice infected with mycolactone-deficient *M. ulcerans* showed less inhibition of IL-2 secretion. This is a testament to the diffusibility of mycolactone in infection implies the ability to cause systemic immunosuppressive effects. Immunosuppression by mycolactone is thought to act through an unknown posttranscriptional mechanism.[522]

The structure–function relationship of mycolactone is complex. The mycolactone core is highly conserved, suggesting that its structure is critical for survival in the natural environment. The ester-linked side chain has been observed in several varieties, but the effect this has on function has only been investigated in relation to human pathology since the definitive environmental niche of *M. ulcerans* is unknown. Hydrogenation of double bond or acetylation of the hydroxyl groups of mycolactone leads to loss of cytopathicity. Several other bacterial toxins are known to interfere with host cell-cycle progression including several pathogenic strains for *E. coli*, *Pasteurella multocida*, and *Helicobacter pylori*. Notably, the *H. pylori* toxin VacA also triggers G_1 cell-cycle arrest and

has been implicated in the development of peptic ulcers.[523] Other macrocyclic compounds produced by actinomycetes that have immunosuppressive properties include rapamycin, FK506, and cyclosporine A. Rapamycin and FK506 both target the intracellular receptor FKBP1A, whereas cyclosporine A targets cyclophilin.[496] The inhibition profile of cytokines indicates that mycolactone possesses a different mechanism of action from these immunosuppressive drugs. A similar 12-membered macrolide, called pladienolide, was recently discovered in *Streptomyces platensis*.[524] Pladienolide has shown potent antitumor activity and has recently been shown to inhibit the activity of splicing factor SF3b.[525]

1.04.6.4.3 Biosynthesis

The polyketide synthase genes responsible for mycolactone biosynthesis are encoded on a giant plasmid.[526] The mycolactone core is synthesized by two modular polyketide synthases: MLSA1 and MLSA2. The mycolactone side chain is synthesized by a third synthase, MLSB. The plasmid also contains three polyketide-modifying enzymes. A full functional characterization of this plasmid followed.[527] **Table 5** lists the genes involved in mycobactin biosynthesis.

The modular subunits of MLSA and MLSB share an unprecedented level of homology, indicating that they are the product of duplication.[526] The giant plasmid encoding the mycolactone synthases is unstable[498] and prone to mutation leading to loss of mycolactone production during passage *in vitro*.[477,498] Despite its instability, the production of mycolactone is well conserved in strains from diverse global origins, indicating that the native environment of *M. ulcerans* applies a strong selection bias for the retention of mycolactone production.

It has now been well established that *cyp140A7* encodes a P-450 hydroxylase that is responsible for the hydroxylation of C-12 on the mycolactone side chain. This gene is absent in strains that primarily produce mycolactones that are unhydroxylated at C-12. Only one *M. ulcerans* isolate from Papua New Guinea predominantly produces mycolactone A/B despite its lack of *cyp140A7*.[498] Hydroxylation at this site has a significant impact on the potency of mycolactone: mycolactone A/B is cytopathic at 0.01 ng ml^{-1}, whereas mycolactone C requires a concentration of 800 ng ml^{-1} for the manifestation of cytopathic effects. Mycolactone A/B is also the most potent IL-2 inhibitor in T cells, followed by mycolactones E and F, whereas C and G were the least potent.[509] This helps to explain why Australian *M. ulcerans* strains that produce mycolactone C as their major metabolite are less virulent compared to West African strains. It should be noted that most strains of *M. ulcerans* still produce a small amount of mycolactone A/B,[501] perhaps indicating the action of a promiscuous hydroxylase in the absence of Cyp140A7. In contrast to mycolactone C, mycolactone F, which also lacks a third hydroxyl group, is nearly as cytotoxic as mycolactone A/B. In an interesting biosynthetic engineering experiment, *M. marinum* strain DL045, which had previously been shown to produce mycolactone F,[507] was complemented with *cyp140A7*. The result was a novel mycolactone containing an aldehyde group instead of a hydroxyl group at the expected location (C-10 since the length of mycolactone F is shortened). This structure was dubbed mycolactone G and exhibited decreased immunosuppressive activity.[509]

Given the wide variability in potency between mycolactone A/B and mycolactone C, the human pathology of mycolactone C-producing *M. ulcerans* may still be due to the small amount of mycolactone A/B that is

Table 5 Genes involved in mycolactone biosynthesis

Gene	*Required for mycolactone production*	*Function*	*Reference(s)*
mlsA1	Yes	Synthesis of the mycolactone core	526
mlsA2	Yes	Synthesis of the final core module and possible ring-closing	526
mslB	Yes	Synthesis of the mycolactone side chain	526
mup038	?	Type II thioesterase	527
mup045	Yes	FabH-like type III ketosynthases	474, 526
cyp140A7 (*mup053*)	No	P-450 hydroxylase that performs C-12 hydroxylation of the R group	526

produced. The differential effectiveness of the different mycolactones for survival in the natural environmental niche of *M. ulcerans* has yet to be elucidated.

A differential analysis of the giant mycolactone synthesis plasmids was carried out on mycobacteria-producing mycolactones A/B, D, E, and F.[510] Mycolactone D-producing *M. ulcerans* from China contains a methylmalonate acyltransferase in module 7 of *mlsB*, corresponding to the additional methyl group observed in the structure. Rearrangement in *mlsB* perfectly matches the structural differences between mycolactone A/B and mycolactone E. Mycolactone E-producing *M. liflandii* has a module 4 deletion that corresponds to its shortened chain length, while a transferase domain change in the load module accounts for the additional terminal methyl group. Mycolactone F-producing *M. marinum* shares the module 4 deletion with *M. liflandii*, and a change in the type of ketoreductase contained in modules 1 and 2 results in the reversed stereochemistry of the two hydroxyl groups. This is consistent with another study profiling the stereospecificity of ketoreductase domains in mycolactone biosynthesis.[528] Plasmids from both mycolactone E- and F-producing mycobacteria lack *cyp140A7* due to an IS*2606*-mediated deletion.[510] The macrocycle formation activity of the ACP and thioesterase domains of MLSA2 have been studied as a novel mode for macrocycle formation to occur.[529]

Mycolactone biosynthesis is a fascinating study of how modular polyketide synthases develop in a rapidly evolving organism. The striking cytotoxic potency of mycolactone and immunosuppressive properties evoke many intriguing therapeutic possibilities for this compound. Intensive research on mycolactone and *M. ulcerans* promises to shed light on the dreadful and long-neglected blight of Buruli ulcer.

1.04.7 Conclusions and Outlook

TB and its mycobacterial relatives are currently one of the most fascinating and active areas of research in chemical biology and probably will be for the foreseeable future. The complexities of the natural products biosynthesized by this genus of microbial life is truly unparalleled in nature and the nearly complete dominance of these products in the pathogenesis of these organisms stands in stark contrast to the dominance of proteins in pathogenesis of other diseases. The long coevolution of mycobacterial disease with *Homo sapiens* certainly underlies the intimate connection of the complex products of these bacilli and the human immune system. Not only is the disease TB a persistent and continuing global health problem but also the basic scientific understanding of the host–pathogen dynamic has spillover effects into other areas of human diseases that are unrelated to the highly visible examples of TB and leprosy. Learning to understand and manipulate such structures has led to the discovery of entire arms of the human immune system that were previously unknown. Although the field has advanced our understanding of these interactions and identified some of the important mechanisms, it seems certain that we have barely even scratched the surface. As examples, consider that the TB vaccine strain BCG is one of the primary therapeutics used today to treat bladder cancer – yet the molecular mechanisms by which this therapy acts are virtually completely unstudied. Also consider that components of the mycobacterial envelope (and their derivatives) are widely being used to develop new vaccine adjuvants for human use but again the molecular mechanisms of these interactions are dramatically understudied. Understanding the structures and the SARs for the varied biological activities of these molecules will continue to reveal critical insights into both new treatments for the diseases caused by these bacteria as well as providing knowledge and tools for understanding and curing other major diseases of humankind.

Abbreviations

ABC transporter	ATP-binding cassette transporter
ACP	acyl carrier protein
AMP	adenosine monophosphate
AMS	adenosine monosulfate
APC	antigen-presenting cell
AT	acyl transferase
ATP	adenosine triphosphate
BCG	bacillus Calmette-Guérin

CD	cluster of differentiation
cDNA	complementary deoxyribonucleic acid
CMB	carboxymycobactin
CoA	coenzyme A
DAT	diacyl trehalose
DC	dendritic cell
DDM	dideoxymycobactin
DH	dehydratase
DTH	delayed-type hypersensitivity
EDDA	ethylenediamine-*N*,*N*′-di(2-hydroxyphenylacetic acid)
EI/MS	electron ionization/mass spectrometry
ELISA	enzyme-linked immunosorbent assay
ER	enoyl reductase
FAB/MS	fast atom bombardment/mass spectrometry
FAS	fatty acid synthase
FBP	fibronectin-binding protein
GC/MS	gas chromatography/mass spectrometry
GMM	glucose monomycolates
GPL	glycopeptidolipid
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
IFN-γ	interferon-γ
IL	interleukin
INH	isoniazid
IS	insertion sequence
KAS	β-ketoacyl synthase
KR	β-ketoacyl reductase
LAM	lipoarabinomannan
LC	liquid chromatography
LPS	lipopolysaccharide
MAGP	mycolic acid–arabinogalactan–peptidoglycan
MALDI-TOF	matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
MAO	monoamine oxidase
MAS	mycocerosic acid synthases
MB	mycobactin
MHC	major histocompatibility complex
MIC	minimum inhibitory concentration
MMG	glycerol monomycolate
MMP	matrix metalloproteinase
mRNA	messenger ribonucleic acid
MS	mass spectrometry
Mtb	*Mycobacterium tuberculosis*
Myc-PL	6-*O*-mycolyl-β-D-mannosyl-1-phosphoheptaprenol
NF-κB	nuclear factor kappa-light-chain-enhancer of activated B cells
NK cells	natural killer cells
NMR	nuclear magnetic resonance
NO	nitric oxide
nsGPL	nonserovar-specific glycopeptidolipid
NTM	nontuberculous mycobacteria
PAPS	3′-phosphoadenosine-5′-phosphosulfate
PAS	*para*-aminosalicylic acid

PAT	polyacyl trehalose
PBMC	peripheral blood monocyte
PCR	polymerase chain reaction
PDIM	phthiocerol dimycocerosate
PGL	phenolic glycolipid
PIM	phosphatidyl-*myo*-inositol mannoside
PKS	polyketide synthase
PPD	purified protein-derivative
PZA	pyrazinamide
RNA	ribonucleic acid
SAM	*S*-adenosyl-L-methionine
SAR	structure–activity relationship
SL-1	sulfolipid-1
ssGPL	serovar-specific glycopeptidolipid
STM	scanning tunneling microscopy
TAT	triacyl trehalose
TB	tuberculosis
TDCM	trehalose dicorynomycolate
TDM	trehalose-6,6′-dimycolate
TE	thioesterase
Th	T helper
TLC	thin-layer chromatography
TMM	trehalose-6-monomycolate
TNF-α	tumor necrosis factor-α
VEGF	vascular endothelial growth factor
WHO	World Health Organization

References

1. J. Kazda, *The Ecology of Mycobacteria*; Kluwer Academic: Dordrecht, Boston, MA, 2000.
2. WHO. *Global Tuberculosis Control Report* 2008; 2008. htttp://www.who.int/tb/publications/global_report/en/index.html
3. D. E. Minnikin, *Lipids: Complex Lipids, Their Chemistry Biosynthesis and Roles*; Academic Press: London, 1982; Vol.1.
4. M. Daffé; P. Draper, *Adv. Microb. Physiol.* **1998**, *39*, 131–203.
5. P. J. Brennan; H. Nikaido, *Annu. Rev. Biochem.* **1995**, *64*, 29–63.
6. V. Puech; M. Chami; A. Lemassu; M. A. Laneelle; B. Schiffler; P. Gounon; N. Bayan; R. Benz; M. Daffé, *Microbiology* **2001**, *147*, 1365–1382.
7. T. Mineda; N. Ohara; H. Yukitake; T. Yamada, *New Microbiol.* **1998**, *21*, 1–7.
8. T. R. Paul; T. J. Beveridge, *J. Bacteriol.* **1992**, *174*, 6508–6517.
9. J. Liu; E. Y. Rosenberg; H. Nikaido, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11254–11258.
10. N. Rastogi, *Res. Microbiol.* **1991**, *142*, 464–476.
11. H. Engelhardt; C. Heinz; M. Niederweis, *J. Biol. Chem.* **2002**, *277*, 37567–37572.
12. M. Faller; M. Niederweis; G. E. Schulz, *Science* **2004**, *303*, 1189–1192.
13. M. Mahfoud; S. Sukumaran; P. Hulsmann; K. Grieger; M. Niederweis, *J. Biol. Chem.* **2006**, *281*, 5908–5915.
14. C. Hoffmann; A. Leis; M. Niederweis; J. M. Plitzko; H. Engelhardt, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3963–3967.
15. B. Zuber; M. Chami; C. Houssin; J. Dubochet; G. Griffiths; M. Daffé, *J. Bacteriol.* **2008**, *190*, 5672–5680.
16. M. Daffé, In *Handbook of Corynebacterium glutamicum*; L. Eggeling, M. Bott, Eds.; CRC Press, Inc.: Boca Raton, FL, 2005; pp 121–148.
17. R. A. Slayden; C. E. Barry, In *Mycobacterium tuberculosis Protocols*; Humana Press: Totowa, NJ, 2001; Vol. 54, pp 229–245.
18. C. E. Barry; R. E. Lee; K. Mdluli; A. E. Sampson; B. G. Schroeder; R. A. Slayden; Y. Yuan, *Prog. Lipid Res.* **1998**, *37*, 143–179.
19. M. Ohta; Y. T. Pan; R. A. Laine; A. D. Elbein, *Eur. J. Biochem.* **2002**, *269*, 3142–3149.
20. Y. Fujita; T. Naka; T. Doi; I. Yano, *Microbiology* **2005**, *151*, 1443–1452.
21. Y. Fujita; T. Naka; M. R. McNeil; I. Yano, *Microbiology* **2005**, *151*, 3403–3416.
22. L. G. Wayne; R. C. Good; E. C. Bottger; R. Butler; M. Dorsch; T. Ezaki; W. Gross; V. Jonas; J. Kilburn; P. Kirschner; M. I. Krichevsky; M. Ridell; T. M. Shinnick; B. Springer; E. Stackebrandt; I. Tarnok; Z. Tarnok; H. Tasaka; V. Vincent; N. G. Warren; C. A. Knott; R. Johnson, *Int. J. Syst. Bacteriol.* **1996**, *46*, 280–297.

23. M. Kai; Y. Fujita; Y. Maeda; N. Nakata; S. Izumi; I. Yano; M. Makino, *FEBS Lett.* **2007**, *581*, 3345–3350.
24. R. Goude; T. Parish, *Future Microbiol.* **2008**, *3*, 299–313.
25. A. Bhatt; V. Molle; G. S. Besra; W. R. Jacobs, Jr.; L. Kremer, *Mol. Microbiol.* **2007**, *64*, 1442–1454.
26. K. Takayama; C. Wang; G. S. Besra, *Clin. Microbiol. Rev.* **2005**, *18*, 81–101.
27. C. E. Barry; D. C. Crick; M. R. McNeil, *Infect. Disord. Drug Targets* **2007**, *7*, 182–202.
28. R. A. Slayden; C. E. Barry, *Microb. Infect.* **2000**, *2*, 659–669.
29. H. T. Wright; K. A. Reynolds, *Curr. Opin. Microbiol.* **2007**, *10*, 447–453.
30. Y. M. Zhang; S. W. White; C. O. Rock, *J. Biol. Chem.* **2006**, *281*, 17541–17544.
31. A. Banerjee; E. Dubnau; A. Quemard; V. Balasubramanian; K. S. Um; T. Wilson; D. Collins; G. de Lisle; W. R. Jacobs, Jr. *Science* **1994**, *263*, 227–230.
32. D. A. Rozwarski; G. A. Grant; D. H. Barton; W. R. Jacobs, Jr.; J. C. Sacchettini, *Science* **1998**, *279*, 98–102.
33. E. K. Schroeder; N. de Souza; D. S. Santos; J. S. Blanchard; L. A. Basso, *Curr. Pharm. Biotechnol.* **2002**, *3*, 197–225.
34. G. S. Timmins; V. Deretic, *Mol. Microbiol.* **2006**, *62*, 1220–1227.
35. A. R. Baulard; J. C. Betts; J. Engohang-Ndong; S. Quan; R. A. McAdam; P. J. Brennan; C. Locht; G. S. Besra, *J. Biol. Chem.* **2000**, *275*, 28326–28331.
36. F. Frenois; J. Engohang-Ndong; C. Locht; A. R. Baulard; V. Villeret, *Mol. Cell* **2004**, *16*, 301–307.
37. L. G. Dover; A. Alahari; P. Gratraud; J. M. Gomes; V. Bhowruth; R. C. Reynolds; G. S. Besra; L. Kremer, *Antimicrob. Agents Chemother.* **2007**, *51*, 1055–1063.
38. A. Alahari; X. Trivelli; Y. Guerardel; L. G. Dover; G. S. Besra; J. C. Sacchettini; R. C. Reynolds; G. D. Coxon; L. Kremer, *PLoS ONE* **2007**, *2*, e1343.
39. A. E. DeBarber; K. Mdluli; M. Bosman; L. G. Bekker; C. E. Barry, III, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9677–9682.
40. R. A. Slayden; R. E. Lee; J. W. Armour; A. M. Cooper; I. M. Orme; P. J. Brennan; G. S. Besra, *Antimicrob. Agents Chemother.* **1996**, *40*, 2813–2819.
41. L. M. McMurry; M. Oethinger; S. B. Levy, *Nature* **1998**, *394*, 531–532.
42. M. R. Kuo; H. R. Morbidoni; D. Alland; S. F. Sneddon; B. B. Gourlie; M. M. Staveski; M. Leonard; J. S. Gregory; A. D. Janjigian; C. Yee; J. M. Musser; B. Kreiswirth; H. Iwamoto; R. Perozzo; W. R. Jacobs, Jr.; J. C. Sacchettini; D. A. Fidock, *J. Biol. Chem.* **2003**, *278*, 20851–20859.
43. R. J. Heath; J. R. Rubin; D. R. Holland; E. Zhang; M. E. Snow; C. O. Rock, *J. Biol. Chem.* **1999**, *274*, 11110–11114.
44. T. J. Sullivan; J. J. Truglio; M. E. Boyne; P. Novichenok; X. Zhang; C. F. Stratton; H. J. Li; T. Kaur; A. Amin; F. Johnson; R. A. Slayden; C. Kisker; P. J. Tonge, *ACS Chem. Biol.* **2006**, *1*, 43–53.
45. P. B. Jones; N. M. Parrish; T. A. Houston; A. Stapon; N. P. Bansal; J. D. Dick; C. A. Townsend, *J. Med. Chem.* **2000**, *43*, 3304–3314.
46. M. H. Cynamon; R. J. Speirs; J. T. Welch, *Antimicrob. Agents Chemother.* **1998**, *42*, 462–463.
47. O. Zimhony; J. S. Cox; J. T. Welch; C. Vilcheze; W. R. Jacobs, Jr., *Nat. Med.* **2000**, *6*, 1043–1047.
48. B. Phetsuksiri; M. Jackson; H. Scherman; M. McNeil; G. S. Besra; A. R. Baulard; R. A. Slayden; A. E. DeBarber; C. E. Barry, III; M. S. Baird; D. C. Crick; P. J. Brennan, *J. Biol. Chem.* **2003**, *278*, 53123–53130.
49. Y. Yuan; D. Mead; B. G. Schroeder; Y. Zhu; C. E. Barry, III, *J. Biol. Chem.* **1998**, *273*, 21282–21290.
50. D. Portevin; C. de Sousa-D'Auria; H. Montrozier; C. Houssin; A. Stella; M. A. Laneelle; F. Bardou; C. Guilhot; M. Daffé, *J. Biol. Chem.* **2005**, *280*, 8862–8874.
51. D. Portevin; C. De Sousa-D'Auria; C. Houssin; C. Grimaldi; M. Chami; M. Daffé; C. Guilhot, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 314–319.
52. D. J. Lea-Smith; J. S. Pyke; D. Tull; M. J. McConville; R. L. Coppel; P. K. Crellin, *J. Biol. Chem.* **2007**, *282*, 11000–11008.
53. B. G. Schroeder; C. E. Barry, III, *Bioorg. Chem.* **2001**, *29*, 164–177.
54. Y. Chang; B. G. Fox, *Biochemistry (Mosc.)* **2006**, *45*, 13476–13486.
55. D. H. Dyer; K. S. Lyle; I. Rayment; B. G. Fox, *Protein Sci.* **2005**, *14*, 1508–1517.
56. K. M. George; Y. Yuan; D. R. Sherman; C. E. Barry, III, *J. Biol. Chem.* **1995**, *270*, 27292–27298.
57. Y. Yuan; R. E. Lee; G. S. Besra; J. T. Belisle; C. E. Barry, IIII, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6630–6634.
58. F. Boissier; F. Bardou; V. Guillet; S. Uttenweiler-Joseph; M. Daffé; A. Quemard; L. Mourey, *J. Biol. Chem.* **2006**, *281*, 4434–4445.
59. C. C. Huang; C. V. Smith; M. S. Glickman; W. R. Jacobs, Jr.; J. C. Sacchettini, *J. Biol. Chem.* **2002**, *277*, 11559–11569.
60. D. F. Iwig; A. Uchida; J. A. Stromberg; S. J. Booker, *J. Am. Chem. Soc.* **2005**, *127*, 11612–11613.
61. Y. Yuan; C. E. Barry, III, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 12828–12833.
62. Y. Yuan; D. C. Crane; J. M. Musser; S. Sreevatsan; C. E. Barry, III, *J. Biol. Chem.* **1997**, *272*, 10041–10049.
63. Y. Yuan; Y. Zhu; D. D. Crane; C. E. Barry, III, *Mol. Microbiol.* **1998**, *29*, 1449–1458.
64. M. S. Glickman; S. M. Cahill; W. R. Jacobs, Jr., *J. Biol. Chem.* **2001**, *276*, 2228–2233.
65. E. Dubnau; M. A. Laneelle; S. Soares; A. Benichou; T. Vaz; D. Prome; J. C. Prome; M. Daffé; A. Quemard, *Mol. Microbiol.* **1997**, *23*, 313–322.
66. A. Quemard; M. A. Laneelle; H. Marrakchi; D. Prome; E. Dubnau; M. Daffé, *Eur. J. Biochem.* **1997**, *250*, 758–763.
67. E. Dubnau; J. Chan; C. Raynaud; V. P. Mohan; M. A. Laneelle; K. Yu; A. Quemard; I. Smith; M. Daffé, *Mol. Microbiol.* **2000**, *36*, 630–637.
68. M. S. Glickman; J. S. Cox; W. R. Jacobs, Jr., *Mol. Cell* **2000**, *5*, 717–727.
69. V. Rao; N. Fujiwara; S. A. Porcelli; M. S. Glickman, *J. Exp. Med.* **2005**, *201*, 535–543.
70. M. S. Glickman, *J. Biol. Chem.* **2003**, *278*, 7844–7849.
71. J. Liu; C. E. Barry, III; G. S. Besra; H. Nikaido, *J. Biol. Chem.* **1996**, *271*, 29545–29551.
72. C. E. Barry, III, *Trends Microbiol.* **2001**, *9*, 237–241.
73. M. Watanabe; Y. Aoyagi; M. Ridell; D. E. Minnikin, *Microbiology* **2001**, *147*, 1825–1837.
74. A. Alahari; L. Alibaud; X. Trivelli; R. Gupta; G. Lamichhane; R. C. Reynolds; W. R. Bishai; Y. Guerardel; L. Kremer, *Mol. Microbiol.* **2009**, *71*, 1263–1277.
75. R. Brosch; S. V. Gordon; T. Garnier; K. Eiglmeier; W. Frigui; P. Valenti; S. Dos Santos; S. Duthoy; C. Lacroix; C. Garcia-Pelayo; J. K. Inwald; P. Golby; J. N. Garcia; R. G. Hewinson; M. A. Behr; M. A. Quail; C. Churcher; B. G. Barrell; J. Parkhill; S. T. Cole, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 5596–5601.

76. M. A. Behr; B. G. Schroeder; J. N. Brinkman; R. A. Slayden; C. E. Barry, III, *J. Bacteriol.* **2000**, *182*, 3394–3399.
77. D. N. Dao; K. Sweeney; T. Hsu; S. S. Gurcha; I. P. Nascimento; D. Roshevsky; G. S. Besra; J. Chan; S. A. Porcelli; W. R. Jacobs, *PLoS Pathog.* **2008**, *4*, e1000081.
78. V. Rao; F. Gao; B. Chen; W. R. Jacobs, Jr.; M. S. Glickman, *J. Clin. Invest.* **2006**, *116*, 1660–1667.
79. H. Noll, *J. Biol. Chem.* **1957**, *224*, 149–164.
80. H. Noll; H. Bloch, *J. Biol. Chem.* **1955**, *214*, 251–265.
81. H. Noll; H. Bloch; J. Asselineau; E. Lederer, *Biochim. Biophys. Acta* **1956**, *20*, 299–309.
82. S. Toriyama; Y. Ikuya; M. Masui; K. Masamichi; E. Kusunose, *FEBS Lett.* **1978**, *95*, 111–115.
83. R. Toubiana; E. Ribi; C. McLaughlin; S. M. Strain, *Cancer Immunol. Immunother.* **1977**, *2*, 189–193.
84. A. Liav; M. B. Goren, *Carbohydr. Res.* **1986**, *155*, 229–235.
85. F. Numata; H. Ishida; K. Nishimura; I. Sekikawa; I. Azuma, *J. Carbohydr. Chem.* **1986**, *5*, 127–138.
86. J. J. O. Kilburn; K. K. Takayama; E. E. L. Armstrong, *Biochem. Biophys. Res. Commun.* **1982**, *108*, 132–139.
87. A. Laszlo; H. H. Baer; M. B. Goren; V. Handzel; F. Papa, *Res. Microbiol.* **1994**, *145*, 563–572.
88. M. Nishizawa; D. M. Garcia; R. Minagawa; Y. Noguchi; H. Imagawa; H. Yamada; R. Watanabe; Y. C. Yoo; I. Azuma, *Synlett* **1996**, 452–454.
89. R. Watanabe; Y. C. Yoo; K. Hata; M. Mitobe; Y. Koike; M. Nishizawa; D. M. Garcia; Y. Nobuchi; H. Imagawa; H. Yamada; I. Azuma, *Vaccine* **1999**, *17*, 1484–1492.
90. I. D. Jenkins; M. B. Goren, *Chem. Phys. Lipids* **1986**, *41*, 225–235.
91. Y. Chapleur; B. Castro; R. Toubiana, *J. Chem. Soc. Perkin Trans. 1* **1980**, 1940–1943.
92. A. Liav; H. M. Flowers; M. B. Goren, *Carbohydr. Res.* **1984**, *133*, 53–58.
93. G. Retzinger; S. Meredith; K. Takayama; R. Hunter; F. Kezdy, *J. Biol. Chem.* **1981**, *256*, 8208–8216.
94. K. A. De Smet; A. Weston; I. N. Brown; D. B. Young; B. D. Robertson, *Microbiology* **2000**, *146* (Pt. 1), 199–208.
95. J. G. Streeter; M. L. Gomez, *Appl. Environ. Microbiol.* **2006**, *72*, 4250–4255.
96. Y. T. Pan; V. Koroth Edavana; W. J. Jourdian; R. Edmondson; J. D. Carroll; I. Pastuszak; A. D. Elbein, *Eur. J. Biochem.* **2004**, *271*, 4259–4269.
97. M. Tzvetkov; C. Klopprogge; O. Zelder; W. Liebl, *Microbiology* **2003**, *149*, 1659–1673.
98. H. N. Murphy; G. R. Stewart; V. V. Mischenko; A. S. Apt; R. Harris; M. S. McAlister; P. C. Driscoll; D. B. Young; B. D. Robertson, *J. Biol. Chem.* **2005**, *280*, 14524–14529.
99. P. J. Woodruff; B. L. Carlson; B. Siridechadilok; M. R. Pratt; R. H. Senaratne; J. D. Mougous; L. W. Riley; S. J. Williams; C. R. Bertozzi, *J. Biol. Chem.* **2004**, *279*, 28835–28843.
100. G. S. Besra; T. Sievert; R. E. Lee; R. A. Slayden; P. J. Brennan; K. Takayama, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12735–12739.
101. C. Wang; B. Hayes; M. M. Vestling; K. Takayama, *Biochem. Biophys. Res. Commun.* **2006**, *340*, 953–960.
102. P. Domenech; H. Kobayashi; K. LeVier; G. C. Walker; C. E. Barry, III, *J. Bacteriol.* **2009**, *191*, 477–485.
103. Y. Fukui; T. Hirai; T. Uchida; M. Yoneda, *Biken J.* **1965**, *8*, 189–199.
104. G. Harth; B. Y. Lee; J. Wang; D. L. Clemens; M. A. Horwitz, *Infect. Immun.* **1996**, *64*, 3038–3047.
105. M. Jackson; C. Raynaud; M. A. Laneelle; C. Guilhot; C. Laurent-Winter; D. Ensergueix; B. Gicquel; M. Daffé, *Mol. Microbiol.* **1999**, *31*, 1573–1587.
106. M. Tropis; X. Meniche; A. Wolf; H. Gebhardt; S. Strelkov; M. Chami; D. Schomburg; R. Kramer; S. Morbach; M. Daffé, *J. Biol. Chem.* **2005**, *280*, 26573–26585.
107. V. Puech; C. Guilhot; E. Perez; M. Tropis; L. Y. Armitige; B. Gicquel; M. Daffé, *Mol. Microbiol.* **2002**, *44*, 1109–1122.
108. D. R. Ronning; T. Klabunde; G. S. Besra; V. D. Vissa; J. T. Belisle; J. C. Sacchettini, *Nat. Struct. Biol.* **2000**, *7*, 141–146.
109. D. H. Anderson; G. Harth; M. A. Horwitz; D. Eisenberg, *J. Mol. Biol.* **2001**, *307*, 671–681.
110. D. R. Ronning; V. Vissa; G. S. Besra; J. T. Belisle; J. C. Sacchettini, *J. Biol. Chem.* **2004**, *279*, 36771–36777.
111. N. Sathyamoorthy; K. Takayama, *J. Biol. Chem.* **1987**, *262*, 13417–13423.
112. J. T. Belisle; V. D. Vissa; T. Sievert; K. Takayama; P. J. Brennan; G. S. Besra, *Science* **1997**, *276*, 1420–1422.
113. J. Boucau; A. K. Sanki; B. J. Voss; S. J. Sucheck; D. R. Ronning, *Anal. Biochem.* **2009**, *385*, 120–127.
114. A. K. Sanki; J. Boucau; D. R. Ronning; S. J. Sucheck, *Glycoconj. J.* **2009**, *26*, 589–596.
115. L. Nguyen; S. Chinnapapagari; C. J. Thompson, *J. Bacteriol.* **2005**, *187*, 6603–6611.
116. Y. Hui; C. W. Chang, *Org. Lett.* **2002**, *4*, 2245–2248.
117. J. D. Rose; J. A. Maddry; R. N. Comber; W. J. Suling; L. N. Wilson; R. C. Reynolds, *Carbohydr. Res.* **2002**, *337*, 105–120.
118. J. Wang; B. Elchert; Y. Hui; J. Y. Takemoto; M. Bensaci; J. Wennergren; H. Chang; R. Rai; C. W. Chang, *Bioorg. Med. Chem.* **2004**, *12*, 6397–6413.
119. S. Gobec; I. Plantan; J. Mravljak; U. Svajger; R. A. Wilson; G. S. Besra; S. L. Soares; R. Appelberg; D. Kikelj, *Eur. J. Med. Chem.* **2007**, *42*, 54–63.
120. S. Gobec; I. Plantan; J. Mravljak; R. A. Wilson; G. S. Besra; D. Kikelj, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3559–3562.
121. A. K. Sanki; J. Boucau; P. Srivastava; S. S. Adams; D. R. Ronning; S. J. Sucheck, *Bioorg. Med. Chem.* **2008**, *16*, 5672–5682.
122. H. Yamagami; T. Matsumoto; N. Fujiwara; T. Arakawa; K. Kaneda; I. Yano; K. Kobayashi, *Infect. Immun.* **2001**, *69*, 810–815.
123. S. S. Syed; R. L. Hunter, Jr., *Ann. Clin. Lab. Sci.* **1997**, *27*, 375–383.
124. K. Kobayashi; K. Kaneda; T. Kasama, *Microsc. Res. Tech.* **2001**, *53*, 241–245.
125. P. Parham, *The Immune System*, 3rd ed.; Garland Science: New York, 2009.
126. R. L. Hunter; M. Olsen; C. Jagannath; J. K. Actor, *Am. J. Pathol.* **2006**, *168*, 1249–1261.
127. V. M. Lima; V. L. Bonato; K. M. Lima; S. A. Dos Santos; R. R. Dos Santos; E. D. Goncalves; L. H. Faccioli; I. T. Brandao; J. M. Rodrigues-Junior; C. L. Silva, *Infect. Immun.* **2001**, *69*, 5305–5312.
128. R. E. Geisel; K. Sakamoto; D. G. Russell; E. R. Rhoades, *J. Immunol.* **2005**, *174*, 5007–5015.
129. E. R. Rhoades; R. E. Geisel; B. A. Butcher; S. McDonough; D. G. Russell, *Tuberculosis (Edinb.)* **2005**, *85*, 159–176.
130. M.-P. Puissegur; C. Botanch; J.-L. Duteyrat; G. Delsol; C. Caratero; F. Altare, *Cell. Microbiol.* **2004**, *6*, 423–433.
131. T. Baba; Y. Natsuhara; K. Kaneda; I. Yano, *Cell. Mol. Life Sci.* **1997**, *53*, 227–232.
132. L. E. Via; P. L. Lin; S. M. Ray; J. Carrillo; S. S. Allen; S. Y. Eum; K. Taylor; E. Klein; U. Manjunatha; J. Gonzales; E. G. Lee; S. K. Park; J. A. Raleigh; S. N. Cho; D. N. McMurray; J. L. Flynn; C. E. Barry, III, *Infect. Immun.* **2008**, *76*, 2333–2340.

133. S. Burthe; M. Bennett; A. Kipar; X. Lambin; A. Smith; S. Telfer; M. Begon, *Parasitology* **2008**, *135*, 309–317.
134. R. D. Cavanagh; X. Lambin; T. Ergon; M. Bennett; I. M. Graham; D. van Soolingen; M. Begon, *Proc. Biol. Sci.* **2004**, *271*, 859–867.
135. R. Cavanagh; M. Begon; M. Bennett; T. Ergon; I. M. Graham; P. E. De Haas; C. A. Hart; M. Koedam; K. Kremer; X. Lambin; P. Roholl; D. Soolingen Dv, *J. Clin. Microbiol.* **2002**, *40*, 3281–3285.
136. I. Sugawara; T. Udagawa; S. C. Hua; M. Reza-Gholizadeh; K. Otomo; Y. Saito; H. Yamada, *J. Med. Microbiol.* **2002**, *51*, 131–137.
137. K. Yasuda, *Osaka City Med. J.* **1999**, *45*, 159–174.
138. J. S. Seggev; M. B. Goren; R. I. Carr; E. Rubenstein; C. H. Kirkpatrick, *Exp. Lung Res.* **1988**, *14*, 431–444.
139. C. W. Borders; A. Courtney; K. Ronen; M. Pilar Laborde-Lahoz; T. V. Guidry; S. A. Hwang; M. Olsen; R. L. Hunter, Jr.; T. J. Hollmann; R. A. Wetsel; J. K. Actor, *Scand. J. Immunol.* **2005**, *62*, 123–130.
140. I. P. Oswald; C. M. Dozois; J. F. Petit; G. Lemaire, *Infect. Immun.* **1997**, *65*, 1364–1369.
141. I. P. Oswald; C. M. Dozois; S. Fournout; J. F. Petit; G. Lemaire, *Eur. Cytokine Netw.* **1999**, *10*, 533–540.
142. R. L. Perez; J. Roman; S. Roser; C. Little; M. Olsen; J. Indrigo; R. L. Hunter; J. K. Actor, *J. Interferon Cytokine Res.* **2000**, *20*, 795–804.
143. J. Indrigo; R. L. Hunter, Jr.; J. K. Actor, *Microbiology* **2002**, *148*, 1991–1998.
144. J. Indrigo; R. L. Hunter, Jr.; J. K. Actor, *Microbiology* **2003**, *149*, 2049–2059.
145. K. J. Welsh; A. N. Abbott; S. A. Hwang; J. Indrigo; L. Y. Armitige; M. R. Blackburn; R. L. Hunter, Jr.; J. K. Actor, *Microbiology* **2008**, *154*, 1813–1824.
146. H. M. Algood; P. L. Lin; J. L. Flynn, *Clin. Infect. Dis.* **2005**, *41* (Suppl. 3), S189–S193.
147. A. Cooper; D. Dalton; T. Stewart; J. Griffin; D. Russell; I. Orme, *J. Exp. Med.* **1993**, *178*, 2243–2247.
148. H. Takimoto; H. Maruyama; K. I. Shimada; R. Yakabe; I. Yano; Y. Kumazawa, *Clin. Exp. Immunol.* **2006**, *144*, 134–141.
149. S. Axelrod; H. Oschkinat; J. Enders; B. Schlegel; V. Brinkmann; S. H. Kaufmann; A. Haas; U. E. Schaible, *Cell. Microbiol.* **2008**, *10*, 1530–1545.
150. B. J. Spargo; L. M. Crowe; T. Ioneda; B. L. Beaman; J. H. Crowe, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 737–740.
151. L. M. Crowe; B. J. Spargo; T. Ioneda; B. L. Beaman; J. H. Crowe, *Biochim. Biophys. Acta* **1994**, *1194*, 53–60.
152. R. G. White; P. Jolles; D. Samour; E. Lederer, *Immunology* **1964**, *7*, 158–171.
153. A. Bekierkunst, *J. Bacteriol.* **1968**, *96*, 958–961.
154. A. Bekierkunst; E. Yarkoni; I. Flechner; S. Morecki; E. Vilkas; E. Lederer, *Infect. Immun.* **1971**, *4*, 256–263.
155. V. L. Moore; Q. N. Myrvik; M. Kato, *Infect. Immun.* **1972**, *6*, 5–8.
156. A. Bekierkunst; E. Yarkoni, *Infect. Immun.* **1973**, *7*, 631–638.
157. T. J. Meyer; E. Ribi; I. Azuma, *Cell. Immunol.* **1975**, *16*, 11–24.
158. C. A. McLaughlin; R. Parker; W. J. Hadlow; R. Toubiana; E. Ribi, *Cell. Immunol.* **1978**, *38*, 14–24.
159. T. V. Guidry; R. L. Hunter, Jr.; J. K. Actor, *Microbiology* **2006**, *152*, 3765–3775.
160. T. Ulrichs; S. H. Kaufmann, *J. Pathol.* **2006**, *208*, 261–269.
161. D. G. Russell, *Nat. Rev. Microbiol.* **2007**, *5*, 39–47.
162. G. Neufeld; T. Cohen; S. Gengrinovitch; Z. Poltorak, *FASEB J.* **1999**, *13*, 9–22.
163. J. R. Jackson; M. P. Seed; C. H. Kircher; D. A. Willoughby; J. D. Winkler, *FASEB J.* **1997**, *11*, 457–465.
164. I. Sakaguchi; N. Ikeda; M. Nakayama; Y. Kato; I. Yano; K. Kaneda, *Infect. Immun.* **2000**, *68*, 2043–2052.
165. F. Alatas; O. Alatas; M. Metintas; A. Ozarslan; S. Erginel; H. Yildirim, *Chest* **2004**, *125*, 2156–2159.
166. N. Saita; N. Fujiwara; I. Yano; K. Soejima; K. Kobayashi, *Infect. Immun.* **2000**, *68*, 5991–5997.
167. A. Bekierkunst; L. Wang; R. Toubiana; E. Lederer, *Infect. Immun.* **1974**, *10*, 1044–1050.
168. J. Klostergaard; M. Leroux; M. Hung, *J. Immunol.* **1991**, *147*, 2802–2808.
169. A. Tabata; K. Kaneda; H. Watanabe; T. Abo; I. Yano, *Microbiol. Immunol.* **1996**, *40*, 651–658.
170. M. Nishizawa; H. Yamamoto; H. Imagawa; V. Barbier-Chassefiere; E. Petit; I. Azuma; D. Papy-Garcia, *J. Org. Chem.* **2007**, *72*, 1627–1633.
171. M. Turneer; J. P. Van Vooren; J. De Bruyn; E. Serruys; P. Dierckx; J. C. Yernault, *J. Clin. Microbiol.* **1988**, *26*, 1714–1719.
172. P. Launois; R. DeLeys; M. N. Niang; A. Drowart; M. Andrien; P. Dierckx; J. L. Cartel; J. L. Sarthou; J. P. Van Vooren; K. Huygen, *Infect. Immun.* **1994**, *62*, 3679–3687.
173. M. Naito; T. Fukuda; K. Sekiguchi; T. Yamada, *Biochem. J.* **2000**, *347*, 725–731.
174. C. Abou-Zeid; T. L. Ratliff; H. G. Wiker; M. Harboe; J. Bennedsen; G. A. Rook, *Infect. Immun.* **1988**, *56*, 3046–3051.
175. P. Peake; A. Gooley; W. J. Britton, *Infect. Immun.* **1993**, *61*, 4828–4834.
176. R. Pasula; P. Wisniowski; W. J. Martin, II, *Infect. Immun.* **2002**, *70*, 1287–1292.
177. C. S. Andersen; E. M. Agger; I. Rosenkrands; J. M. Gomes; V. Bhowruth; K. J. Gibson; R. V. Petersen; D. E. Minnikin; G. S. Besra; P. Andersen, *J. Immunol.* **2009**, *182*, 424–432.
178. M. A. Horwitz; B. W. Lee; B. J. Dillon; G. Harth, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1530–1534.
179. S. L. Baldwin; C. D'Souza; A. D. Roberts; B. P. Kelly; A. A. Frank; M. A. Lui; J. B. Ulmer; K. Huygen; D. M. McMurray; I. M. Orme, *Infect. Immun.* **1998**, *66*, 2951–2959.
180. A. W. Olsen; A. Williams; L. M. Okkels; G. Hatch; P. Andersen, *Infect. Immun.* **2004**, *72*, 6148–6150.
181. K. M. Lima; S. A. Santos; V. M. Lima; A. A. Coelho-Castelo; J. M. Rodrigues, Jr.; C. L. Silva, *Gene Ther.* **2003**, *10*, 678–685.
182. S. A. Khader; G. K. Bell; J. E. Pearl; J. J. Fountain; J. Rangel-Moreno; G. E. Cilley; F. Shen; S. M. Eaton; S. L. Gaffen; S. L. Swain; R. M. Locksley; L. Haynes; T. D. Randall; A. M. Cooper, *Nat. Immunol.* **2007**, *8*, 369–377.
183. A. A. Pathan; C. R. Sander; H. A. Fletcher; I. Poulton; N. C. Alder; N. E. Beveridge; K. T. Whelan; A. V. Hill; H. McShane, *PLoS ONE* **2007**, *2*, e1052.
184. J. W. Pan; N. Fujiwara; S. Oka; R. Maekura; T. Ogura; I. Yano, *Microbiol. Immunol.* **1999**, *43*, 863–869.
185. N. Fujiwara; J. W. Pan; K. Enomoto; Y. Terano; T. Honda; I. Yano, *FEMS Immunol. Med. Microbiol.* **1999**, *24*, 141–149.
186. T. Ulrichs; S. A. Porcelli, *Rev. Immunogenet.* **2000**, *2*, 416–432.
187. B. E. Willcox; C. R. Willcox; L. G. Dover; G. Besra, *Curr. Top. Microbiol. Immunol.* **2007**, *314*, 73–110.
188. M. Sugita; E. P. Grant; E. van Donselaar; V. W. Hsu; R. A. Rogers; P. J. Peters; M. B. Brenner, *Immunity* **1999**, *11*, 743–752.

189. G. Bricard; S. A. Porcelli, *Cell. Mol. Life Sci.* **2007**, *64*, 1824–1840.
190. D. M. Zajonc; M. D. Crispin; T. A. Bowden; D. C. Young; T. Y. Cheng; J. Hu; C. E. Costello; P. M. Rudd; R. A. Dwek; M. J. Miller; M. B. Brenner; D. B. Moody; I. A. Wilson, *Immunity* **2005**, *22*, 209–219.
191. R. E. Hunger; P. A. Sieling; M. T. Ochoa; M. Sugaya; A. E. Burdick; T. H. Rea; P. J. Brennan; J. T. Belisle; A. Blauvelt; S. A. Porcelli; R. L. Modlin, *J. Clin. Invest.* **2004**, *113*, 701–708.
192. P. A. Sieling; D. Chatterjee; S. A. Porcelli; T. I. Prigozy; R. J. Mazzaccaro; T. Soriano; B. R. Bloom; M. B. Brenner; M. Kronenberg; P. J. Brennan; R. L. Modlin, *Science* **1995**, *69* (5221), 227–230.
193. D. B. Moody; B. B. Reinhold; V. N. Reinhold; G. S. Besra; S. A. Porcelli, *Immunol. Lett.* **1999**, *65*, 85–91.
194. D. B. Moody; B. B. Reinhold; M. R. Guy; E. M. Beckman; D. E. Frederique; S. T. Furlong; S. Ye; V. N. Reinhold; P. A. Sieling; R. L. Modlin; G. S. Besra; S. A. Porcelli, *Science* **1997**, *278*, 283–286.
195. L. Tentori; G. Graziani; S. A. Porcelli; M. Sugita; M. B. Brenner; R. Madaio; E. Bonmassar; A. Giuliani; A. Aquino, *Antimicrob. Agents Chemother.* **1998**, *42*, 550–554.
196. Y. Enomoto; M. Sugita; I. Matsunaga; T. Naka; A. Sato; T. Kawashima; K. Shimizu; H. Takahashi; Y. Norose; I. Yano, *Biochem. Biophys. Res. Commun.* **2005**, *337*, 452–456.
197. M. Gilleron; S. Stenger; Z. Mazorra; F. Wittke; S. Mariotti; G. Bohmer; J. Prandi; L. Mori; G. Puzo; G. De Libero, *J. Exp. Med.* **2004**, *199*, 649–659.
198. T. Batuwangala; D. Shepherd; S. D. Gadola; K. J. Gibson; N. R. Zaccai; A. R. Fersht; G. S. Besra; V. Cerundolo; E. Y. Jones, *J. Immunol.* **2004**, *172*, 2382–2388.
199. E. F. Pettersen; T. D. Goddard; C. C. Huang; G. S. Couch; D. M. Greenblatt; E. C. Meng; T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.
200. D. B. Moody; T. Ulrichs; W. Muhlecker; D. C. Young; S. S. Gurcha; E. Grant; J. P. Rosat; M. B. Brenner; C. E. Costello; G. S. Besra; S. A. Porcelli, *Nature* **2000**, *404*, 884–888.
201. I. Matsunaga; A. Bhatt; D. C. Young; T. Y. Cheng; S. J. Eyles; G. S. Besra; V. Briken; S. A. Porcelli; C. E. Costello; W. R. Jacobs, Jr.; D. B. Moody, *J. Exp. Med.* **2004**, *200*, 1559–1569.
202. A. de Jong; E. C. Arce; T. Y. Cheng; R. P. van Summeren; B. L. Feringa; V. Dudkin; D. Crich; I. Matsunaga; A. J. Minnaard; D. B. Moody, *Chem. Biol.* **2007**, *14*, 1232–1242.
203. M. Sugita; N. van Der Wel; R. A. Rogers; P. J. Peters; M. B. Brenner, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8445–8450.
204. R. Ryll; K. Watanabe; N. Fujiwara; H. Takimoto; R. Hasunuma; Y. Kumazawa; M. Okada; I. Yano, *Microbes Infect.* **2001**, *3*, 611–619.
205. A. Chackerian; J. Alt; V. Perera; S. M. Behar, *Infect. Immun.* **2002**, *70*, 6302–6309.
206. S. M. Behar; C. C. Dascher; M. J. Grusby; C. R. Wang; M. B. Brenner, *J. Exp. Med.* **1999**, *189*, 1973–1980.
207. J. K. Actor; M. Olsen; R. L. Hunter, Jr.; Y. J. Geng, *J. Interferon Cytokine Res.* **2001**, *21*, 1089–1096.
208. M. Munoz; M. A. Laneelle; M. Luquin; J. Torrelles; E. Julian; V. Ausina; M. Daffé, *FEMS Microbiol. Lett.* **1997**, *157*, 251–259.
209. D. E. Minnikin; G. Dobson; M. Goodfellow; M. Magnusson; M. Ridell, *J. Gen. Microbiol.* **1985**, *131*, 1375–1381.
210. D. E. Minnikin; G. Dobson; D. Sesardic; M. Ridell, *J. Gen. Microbiol.* **1985**, *131*, 1369–1374.
211. S. Alugupalli; F. Portaels; L. Larsson, *J. Bacteriol.* **1994**, *176*, 2962–2969.
212. M. A. Ariza; P. L. Valero-Guillen, *FEMS Microbiol. Lett.* **1994**, *119*, 279–282.
213. M. Daffé; C. Lacave; M. A. Laneelle; M. Gillois; G. Laneelle, *Eur. J. Biochem.* **1988**, *172*, 579–584.
214. R. Saavedra; E. Segura; R. Leyva; L. A. Esparza; L. M. Lopez-Marin, *Clin. Diagn. Lab. Immunol.* **2001**, *8*, 1081–1088.
215. R. Saavedra; E. Segura; E. P. Tenorio; L. M. Lopez-Marin, *Microbes Infect.* **2006**, *8*, 533–540.
216. C. Rousseau; O. Neyrolles; Y. Bordat; S. Giroux; T. D. Sirakova; M. C. Prevost; P. E. Kolattukudy; B. Gicquel; M. Jackson, *Cell. Microbiol.* **2003**, *5*, 405–415.
217. V. S. Dubey; T. D. Sirakova; P. E. Kolattukudy, *Mol. Microbiol.* **2002**, *45*, 1451–1459.
218. S. B. Walters; E. Dubnau; I. Kolesnikova; F. Laval; M. Daffé; I. Smith, *Mol. Microbiol.* **2006**, *60*, 312–330.
219. G. Middlebrook; C. M. Coleman; W. B. Schaefer, *Proc. Natl. Acad. Sci. U.S.A.* **1959**, *45*, 1801–1804.
220. M. Tsukamura; S. Mizuno, *Microbiol. Immunol.* **1986**, *30*, 589–591.
221. M. B. Goren; P. D'Arcy Hart; M. R. Young; J. A. Armstrong, *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2510–2514.
222. C. Y. Soto; M. Cama; I. Gibert; M. Luquin, *FEMS Microbiol. Lett.* **2000**, *187*, 103–107.
223. L. Zhang; M. B. Goren; T. J. Holzer; B. R. Andersen, *Infect. Immun.* **1988**, *56*, 2876–2883.
224. L. Zhang; D. English; B. R. Andersen, *J. Immunol.* **1991**, *146*, 2730–2736.
225. Y. Okamoto; Y. Fujita; T. Naka; M. Hirai; I. Tomiyasu; I. Yano, *Microb. Pathog.* **2006**, *40*, 245–253.
226. J. D. Mougous; C. J. Petzold; R. H. Senaratne; D. H. Lee; D. L. Akey; F. L. Lin; S. E. Munchel; M. R. Pratt; L. W. Riley; J. A. Leary; J. M. Berger; C. R. Bertozzi, *Nat. Struct. Mol. Biol.* **2004**, *11*, 721–729.
227. T. D. Sirakova; A. K. Thirumala; V. S. Dubey; H. Sprecher; P. E. Kolattukudy, *J. Biol. Chem.* **2001**, *276*, 16833–16839.
228. C. A. Rivera-Marrero; J. D. Ritzenthaler; S. A. Newburn; J. Roman; R. D. Cummings, *Microbiology-Sgm* **2002**, *148*, 783–792.
229. P. Kumar; M. W. Schelle; M. Jain; F. L. Lin; C. J. Petzold; M. D. Leavell; J. A. Leary; J. S. Cox; C. R. Bertozzi, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11221–11226.
230. K. Bhatt; S. S. Gurcha; A. Bhatt; G. S. Besra; W. R. Jacobs, Jr., *Microbiology* **2007**, *153*, 513–520.
231. J. Lynett; R. W. Stokes, *Microbiology* **2007**, *153*, 3133–3140.
232. P. Domenech; M. B. Reed; C. S. Dowd; C. Manca; G. Kaplan; C. E. Barry, *J. Biol. Chem.* **2004**, *279*, 21257–21265.
233. S. E. Converse; J. D. Mougous; M. D. Leavell; J. A. Leary; C. R. Bertozzi; J. S. Cox, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6121–6126.
234. C. Rousseau; O. C. Turner; E. Rush; Y. Bordat; T. D. Sirakova; P. E. Kolattukudy; S. Ritter; I. M. Orme; B. Gicquel; M. Jackson, *Infect. Immun.* **2003**, *71*, 4684–4690.
235. P. Saxena; G. Yadav; D. Mohanty; R. S. Gokhale, *J. Biol. Chem.* **2003**, *278*, 44780–44790.
236. R. S. Gokhale; P. Saxena; T. Chopra; D. Mohanty, *Nat. Prod. Rep.* **2007**, *24*, 267–277.
237. K. C. Onwueme; J. A. Ferreras; J. Buglino; C. D. Lima; L. E. Quadri, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 4608–4613.
238. H. Demarteau Ginsburg; E. Lederer; R. Ryhage; S. Stallberg-Stenhagen; E. Stenhagen, *Nature* **1959**, *183*, 117–119.
239. M. B. Goren; O. Brokl; W. B. Schaefer, *Infect. Immun.* **1974**, *9*, 150–158.

240. P. J. Brennan, *Int. J. Lepr. Other Mycobact. Dis.* **1983**, *51*, 387–396.
241. M. Daffé; M. A. Laneélle, *Biochim. Biophys. Acta* **1989**, *1002*, 333–337.
242. M. Daffé; F. Papa; A. Laszlo; H. L. David, *J. Gen. Microbiol.* **1989**, *135*, 2759–2766.
243. M. Daffé, *Res. Microbiol.* **1991**, *142*, 405–410.
244. M. Mathur; P. E. Kolattukudy, *J. Biol. Chem.* **1992**, *267*, 19388–19395.
245. A. K. Azad; T. D. Sirakova; N. D. Fernandes; P. E. Kolattukudy, *J. Biol. Chem.* **1997**, *272*, 16741–16745.
246. T. D. Sirakova; A. M. Fitzmaurice; P. Kolattukudy, *J. Bacteriol.* **2002**, *184*, 6796–6802.
247. S. T. Cole; R. Brosch; J. Parkhill; T. Garnier; C. Churcher; D. Harris; S. V. Gordon; K. Eiglmeier; S. Gas; C. E. Barry, III; F. Tekaia; K. Badcock; D. Basham; D. Brown; T. Chillingworth; R. Connor; R. Davies; K. Devlin; T. Feltwell; S. Gentles; N. Hamlin; S. Holroyd; T. Hornsby; K. Jagels; A. Krogh; J. McLean; S. Moule; L. Murphy; K. Oliver; J. Osborne; M. A. Quail; M. A. Rajandream; J. Rogers; S. Rutter; K. Seeger; J. Skelton; R. Squares; S. Squares; J. E. Sulston; K. Taylor; S. Whitehead; B. G. Barrell, *Nature* **1998**, *393*, 537–544.
248. C. M. Sassetti; D. H. Boyd; E. J. Rubin, *Mol. Microbiol.* **2003**, *48*, 77–84.
249. V. Pelicic; M. Jackson; J. M. Reyrat; W. R. Jacobs, Jr.; B. Gicquel; C. Guilhot, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10955–10960.
250. M. Hensel; J. E. Shea; C. Gleeson; M. D. Jones; E. Dalton; D. W. Holden, *Science* **1995**, *269*, 400–403.
251. A. M. Fitzmaurice; P. E. Kolattukudy, *J. Biol. Chem.* **1998**, *273*, 8033–8039.
252. K. Tahlan; S. K. Ahn; A. Sing; T. D. Bodnaruk; A. R. Willems; A. R. Davidson; J. R. Nodwell, *Mol. Microbiol.* **2007**, *63*, 951–961.
253. J. S. Cox; B. Chen; M. McNeil; W. R. Jacobs, Jr., *Nature* **1999**, *402*, 79–83.
254. P. Domenech; M. B. Reed; C. E. Barry, III, *Infect. Immun.* **2005**, *73*, 3492–3501.
255. M. Jain; J. S. Cox, *PLoS Pathog.* **2005**, *1*, e2.
256. L. R. Camacho; D. Ensergueix; E. Perez; B. Gicquel; C. Guilhot, *Mol. Microbiol.* **1999**, *34*, 257–267.
257. L. R. Camacho; P. Constant; C. Raynaud; M. A. Laneelle; J. A. Triccas; B. Gicquel; M. Daffé; C. Guilhot, *J. Biol. Chem.* **2001**, *276*, 19845–19854.
258. C. Rousseau; T. D. Sirakova; V. S. Dubey; Y. Bordat; P. E. Kolattukudy; B. Gicquel; M. Jackson, *Microbiology* **2003**, *149*, 1837–1847.
259. T. D. Sirakova; V. S. Dubey; M. H. Cynamon; P. E. Kolattukudy, *J. Bacteriol.* **2003**, *185*, 2999–3008.
260. T. D. Sirakova; V. S. Dubey; H. J. Kim; M. H. Cynamon; P. E. Kolattukudy, *Infect. Immun.* **2003**, *71*, 3794–3801.
261. O. A. Trivedi; P. Arora; A. Vats; M. Z. Ansari; R. Tickoo; V. Sridharan; D. Mohanty; R. S. Gokhale, *Mol. Cell* **2005**, *17*, 631–643.
262. J. Buglino; K. C. Onwueme; J. A. Ferreras; L. E. Quadri; C. D. Lima, *J. Biol. Chem.* **2004**, *279*, 30634–30642.
263. M. S. Li; I. M. Monahan; S. J. Waddell; J. A. Mangan; S. L. Martin; M. J. Everett; P. D. Butcher, *Microbiology* **2001**, *147*, 2293–2305.
264. L. S. Schlesinger; M. A. Horwitz, *J. Immunol.* **1991**, *147*, 1983–1994.
265. M. A. Neill; S. J. Klebanoff, *J. Exp. Med.* **1988**, *167*, 30–42.
266. V. Ng; G. Zanazzi; R. Timpl; J. F. Talts; J. L. Salzer; P. J. Brennan; A. Rambukkana, *Cell* **2000**, *103*, 511–524.
267. M. B. Reed; P. Domenech; C. Manca; H. Su; A. K. Barczak; B. N. Kreiswirth; G. Kaplan; C. E. Barry, III, *Nature* **2004**, *431*, 84–87.
268. P. Constant; E. Perez; W. Malaga; M. A. Laneelle; O. Saurel; M. Daffé; C. Guilhot, *J. Biol. Chem.* **2002**, *277*, 38148–38158.
269. E. Perez; P. Constant; A. Lemassu; F. Laval; M. Daffé; C. Guilhot, *J. Biol. Chem.* **2004**, *279*, 42574–42583.
270. K. C. Onwueme; C. J. Vos; J. Zurita; C. E. Soll; L. E. Quadri, *J. Bacteriol.* **2005**, *187*, 4760–4766.
271. E. Perez; P. Constant; F. Laval; A. Lemassu; M. A. Laneelle; M. Daffé; C. Guilhot, *J. Biol. Chem.* **2004**, *279*, 42584–42592.
272. R. Simeone; P. Constant; W. Malaga; C. Guilhot; M. Daffé; C. Chalut, *FEBS J.* **2007**, *274*, 1957–1969.
273. R. Pinto; B. M. Saunders; L. R. Camacho; W. J. Britton; B. Gicquel; J. A. Triccas, *J. Infect. Dis.* **2004**, *189*, 105–112.
274. G. S. Hotter; B. J. Wards; P. Mouat; G. S. Besra; J. Gomes; M. Singh; S. Bassett; P. Kawakami; P. R. Wheeler; G. W. de Lisle; D. M. Collins, *J. Bacteriol.* **2005**, *187*, 2267–2277.
275. E. Infante; L. D. Aguilar; B. Gicquel; R. H. Pando, *Clin. Exp. Immunol.* **2005**, *141*, 21–28.
276. G. M. Scandurra; R. B. Williams; J. A. Triccas; R. Pinto; B. Gicquel; B. Slobedman; A. Cunningham; W. J. Britton, *Microbes Infect.* **2007**, *9*, 87–95.
277. C. Rousseau; N. Winter; E. Pivert; Y. Bordat; O. Neyrolles; P. Ave; M. Huerre; B. Gicquel; M. Jackson, *Cell. Microbiol.* **2004**, *6*, 277–287.
278. S. W. Hunter; H. Gaylord; P. J. Brennan, *J. Biol. Chem.* **1986**, *261*, 12345–12351.
279. S. W. Hunter; P. J. Brennan, *J. Biol. Chem.* **1990**, *265*, 9272–9279.
280. S. W. Hunter; R. C. Murphy; K. Clay; M. B. Goren; P. J. Brennan, *J. Biol. Chem.* **1983**, *258*, 10481–10487.
281. W. B. Severn; R. H. Furneaux; R. Falshaw; P. H. Atkinson, *Carbohydr. Res.* **1998**, *308*, 397–408.
282. J. Nigou; M. Gilleron; G. Puzo, *Biochimie* **2003**, *85*, 153–166.
283. D. Chatterjee; K. Lowell; B. Rivoire; M. R. McNeil; P. J. Brennan, *J. Biol. Chem.* **1992**, *267*, 6234–6239.
284. J. Nigou; L. G. Dover; G. S. Besra, *Biochemistry (Mosc.)* **2002**, *41*, 4392–4398.
285. J. Kordulakova; M. Gilleron; G. Puzo; P. J. Brennan; B. Gicquel; K. Mikusova; M. Jackson, *J. Biol. Chem.* **2003**, *278*, 36285–36295.
286. M. L. Schaeffer; K. H. Khoo; G. S. Besra; D. Chatterjee; P. J. Brennan; J. T. Belisle; J. M. Inamine, *J. Biol. Chem.* **1999**, *274*, 31625–31631.
287. L. Kremer; S. S. Gurcha; P. Bifani; P. G. Hitchen; A. Baulard; H. R. Morris; A. Dell; P. J. Brennan; G. S. Besra, *Biochem. J.* **2002**, *363*, 437–447.
288. G. S. Besra; C. B. Morehouse; C. M. Rittner; C. J. Waechter; P. J. Brennan, *J. Biol. Chem.* **1997**, *272*, 18460–18466.
289. T. R. McCarthy; J. B. Torrelles; A. S. MacFarlane; M. Katawczik; B. Kutzbach; L. E. Desjardin; S. Clegg; J. B. Goldberg; L. S. Schlesinger, *Mol. Microbiol.* **2005**, *58*, 774–790.
290. A. K. Mishra; L. J. Alderwick; D. Rittmann; R. V. Tatituri; J. Nigou; M. Gilleron; L. Eggeling; G. S. Besra, *Mol. Microbiol.* **2007**, *65*, 1503–1517.
291. D. C. Alexander; J. R. Jones; T. Tan; J. M. Chen; J. Liu, *J. Biol. Chem.* **2004**, *279*, 18824–18833.

292. Y. S. Morita; C. B. Sena; R. F. Waller; K. Kurokawa; M. F. Sernee; F. Nakatani; R. E. Haites; H. Billman-Jacobe; M. J. McConville; Y. Maeda; T. Kinoshita, *J. Biol. Chem.* **2006**, *281*, 25143–25155.
293. D. Kaur; M. R. McNeil; K. H. Khoo; D. Chatterjee; D. C. Crick; M. Jackson; P. J. Brennan, *J. Biol. Chem.* **2007**, *282*, 27133–27140.
294. D. Kaur; S. Berg; P. Dinadayala; B. Gicquel; D. Chatterjee; M. R. McNeil; V. D. Vissa; D. C. Crick; M. Jackson; P. J. Brennan, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13664–13669.
295. S. Berg; J. Starbuck; J. B. Torrelles; V. D. Vissa; D. C. Crick; D. Chatterjee; P. J. Brennan, *J. Biol. Chem.* **2005**, *280*, 5651–5663.
296. R. Goude; A. G. Amin; D. Chatterjee; T. Parish, *J. Bacteriol.* **2008**, *190*, 4335–4341.
297. J. B. Torrelles; K. H. Khoo; P. A. Sieling; R. L. Modlin; N. Zhang; A. M. Marques; A. Treumann; C. D. Rithner; P. J. Brennan; D. Chatterjee, *J. Biol. Chem.* **2004**, *279*, 41227–41239.
298. E. Hayakawa; F. Tokumasu; G. A. Nardone; A. J. Jin; V. A. Hackley; J. A. Dvorak, *Biophys. J.* **2007**, *93*, 4018–4030.
299. K. H. Khoo; A. Dell; H. R. Morris; P. J. Brennan; D. Chatterjee, *J. Biol. Chem.* **1995**, *270*, 12380–12389.
300. P. F. Barnes; D. Chatterjee; J. S. Abrams; S. Lu; E. Wang; M. Yamamura; P. J. Brennan; R. L. Modlin, *J. Immunol.* **1992**, *149*, 541–547.
301. D. Chatterjee; S. W. Hunter; M. McNeil; P. J. Brennan, *J. Biol. Chem.* **1992**, *267*, 6228–6233.
302. C. Vignal; Y. Guerardel; L. Kremer; M. Masson; D. Legrand; J. Mazurier; E. Elass, *J. Immunol.* **2003**, *171*, 2014–2023.
303. D. Chatterjee; K. H. Khoo, *Glycobiology* **1998**, *8*, 113–120.
304. V. Briken; S. A. Porcelli; G. S. Besra; L. Kremer, *Mol. Microbiol.* **2004**, *53*, 391–403.
305. D. N. Dao; L. Kremer; Y. Guerardel; A. Molano; W. R. Jacobs, Jr.; S. A. Porcelli; V. Briken, *Infect. Immun.* **2004**, *72*, 2067–2074.
306. K. H. Khoo; J. B. Tang; D. Chatterjee, *J. Biol. Chem.* **2001**, *276*, 3863–3871.
307. J. B. Torrelles; R. Knaup; A. Kolareth; T. Slepushkina; T. M. Kaufman; P. Kang; P. J. Hill; P. J. Brennan; D. Chatterjee; J. T. Belisle; J. M. Musser; L. S. Schlesinger, *J. Biol. Chem.* **2008**, *283*, 31417–31428.
308. J. S. Schorey; L. Sweet, *Glycobiology* **2008**, *18*, 832–841.
309. W. W. Barrow; P. J. Brennan, *Infect. Immun.* **1982**, *36*, 678–684.
310. P. J. Brennan; M. Heifets; B. P. Ullom, *J. Clin. Microbiol.* **1982**, *15*, 447–455.
311. G. O. Aspinall; D. W. Gammon; R. K. Sood; D. Chatterjee; B. Rivoire; P. J. Brennan, *Carbohydr. Res.* **1992**, *237*, 57–77.
312. P. J. Brennan; G. O. Aspinall; J. E. Shin, *J. Biol. Chem.* **1981**, *256*, 6817–6822.
313. J. T. Belisle; K. Klaczkiewicz; P. J. Brennan; W. R. Jacobs, Jr.; J. M. Inamine, *J. Biol. Chem.* **1993**, *268*, 10517–10523.
314. J. T. Belisle; M. R. McNeil; D. Chatterjee; J. M. Inamine; P. J. Brennan, *J. Biol. Chem.* **1993**, *268*, 10510–10516.
315. S. T. Howard; E. Rhoades; J. Recht; X. Pang; A. Alsup; R. Kolter; C. R. Lyons; T. F. Byrd, *Microbiology* **2006**, *152*, 1581–1590.
316. E. L. Wright; M. Pourshafie; W. W. Barrow, *FEMS Microbiol. Lett.* **1992**, *77*, 209–216.
317. E. L. Wright; S. Zywno-van Ginkel; N. Rastogi; W. W. Barrow, *J. Clin. Microbiol.* **1996**, *34*, 2475–2478.
318. W. W. Barrow; E. L. Wright; K. S. Goh; N. Rastogi, *Antimicrob. Agents Chemother.* **1993**, *37*, 652–661.
319. K. Shimada; H. Takimoto; I. Yano; Y. Kumazawa, *Microbiol. Immunol.* **2006**, *50*, 243–251.
320. H. Kano; T. Doi; Y. Fujita; H. Takimoto; I. Yano; Y. Kumazawa, *Biol. Pharm. Bull.* **2005**, *28*, 335–339.
321. J. Recht; R. Kolter, *J. Bacteriol.* **2001**, *183*, 5718–5724.
322. J. H. Patterson; M. J. McConville; R. E. Haites; R. L. Coppel; H. Billman-Jacobe, *J. Biol. Chem.* **2000**, *275*, 24900–24906.
323. J. Recht; A. Martinez; S. Torello; R. Kolter, *J. Bacteriol.* **2000**, *182*, 4348–4351.
324. G. Carter; M. Wu; D. C. Drummond; L. E. Bermudez, *J. Med. Microbiol.* **2003**, *52*, 747–752.
325. R. Freeman; H. Geier; K. M. Weigel; J. Do; T. E. Ford; G. A. Cangelosi, *Appl. Environ. Microbiol.* **2006**, *72*, 7554–7558.
326. A. K. Ojha; S. Varma; D. Chatterji, *Microbiology* **2002**, *148*, 3039–3048.
327. R. Mukherjee; M. Gomez; N. Jayaraman; I. Smith; D. Chatterji, *Microbiology* **2005**, *151*, 2385–2392.
328. H. Billman-Jacobe; M. J. McConville; R. E. Haites; S. Kovacevic; R. L. Coppel, *Mol. Microbiol.* **1999**, *33*, 1244–1253.
329. Y. Miyamoto; T. Mukai; N. Nakata; Y. Maeda; M. Kai; T. Naka; I. Yano; M. Makino, *J. Bacteriol.* **2006**, *188*, 86–95.
330. D. Jeevarajah; J. H. Patterson; E. Taig; T. Sargeant; M. J. McConville; H. Billman-Jacobe, *J. Bacteriol.* **2004**, *186*, 6792–6799.
331. D. Jeevarajah; J. H. Patterson; M. J. McConville; H. Billman-Jacobe, *Microbiology* **2002**, *148*, 3079–3087.
332. H. Billman-Jacobe, *Curr. Sci.* **2004**, *86*, 111–114.
333. T. M. Eckstein; F. S. Silbaq; D. Chatterjee; N. J. Kelly; P. J. Brennan; J. T. Belisle, *J. Bacteriol.* **1998**, *180*, 5567–5573.
334. J. N. Maslow; V. R. Irani; S. H. Lee; T. M. Eckstein; J. M. Inamine; J. T. Belisle, *Microbiology* **2003**, *149*, 3193–3202.
335. R. R. Crichton; J. L. Pierre, *BioMetals* **2001**, *14*, 99–112.
336. J. B. Neilands, *J. Biol. Chem.* **1995**, *270*, 26723–26726.
337. F. W. Twort; G. L. Y. Ingram, *Proc. R. Soc. Lond. B, Biol. Sci.* **1912**, *84*, 517–542.
338. J. Francis; H. M. Macturk; J. Madinaveitia; G. A. Snow, *Biochem. J.* **1953**, *55*, 596–607.
339. G. A. Snow, *Bacteriol. Rev.* **1970**, *34*, 99–125.
340. De J. J. Voss; K. Rutter; B. G. Schroeder; H. Su; Y. Zhu; C. E. Barry, III, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1252–1257.
341. L. P. Macham; C. Ratledge, *J. Gen. Microbiol.* **1975**, *89*, 379–382.
342. C. D. Cox; K. L. Rinehart, Jr.; M. L. Moore; J. C. Cook, Jr., *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4256–4260.
343. S. A. Ong; T. Peterson; J. B. Neilands, *J. Biol. Chem.* **1979**, *254*, 1860–1865.
344. G. L. Griffiths; S. P. Sigel; S. M. Payne; J. B. Neilands, *J. Biol. Chem.* **1984**, *259*, 383–385.
345. M. A. F. Jalal; M. B. Hossain; D. Van der Helm; J. Sanders-Loehr; L. A. Actis; J. H. Crosa, *J. Am. Chem. Soc.* **1989**, *111*, 292–296.
346. H. Drechsel; H. Stephan; R. Lotz; H. Haag; H. Zähner; K. Hantke; G. Jung, *Liebigs Ann. Chem.* **1995**, 1727–1733.
347. J. S. Martinez; G. P. Zhang; P. D. Holt; H. T. Jung; C. J. Carrano; M. G. Haygood; A. Butler, *Science* **2000**, *287*, 1245–1247.
348. C. Ratledge; P. V. Patel, *J. Gen. Microbiol.* **1976**, *93*, 141–152.
349. K. Schneider; I. Rose; S. Vikineswary; A. L. Jones; M. Goodfellow; G. Nicholson; W. Beil; R. D. Sussmuth; H. P. Fiedler, *J. Nat. Prod.* **2007**, *70*, 932–935.
350. J. J. De Voss; K. Rutter; B. G. Schroeder; C. E. Barry, III, *J. Bacteriol.* **1999**, *181*, 4443–4451.
351. G. A. Snow; A. J. White, *Biochem. J.* **1969**, *115*, 1031–1050.
352. Y. M. Lin; M. J. Miller; U. Mollmann, *BioMetals* **2001**, *14*, 153–157.

353. S. J. Lane; P. S. Marshall; R. J. Upton; C. Ratledge; M. Ewing, *Tetrahedron Lett.* **1995**, *36*, 4129–4132.
354. J. Gobin; C. H. Moore; J. R. Reeve, Jr.; D. K. Wong; B. W. Gibson; M. A. Horwitz, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5189–5193.
355. S. J. Lane; P. S. Marshall; R. J. Upton; C. Ratledge, *BioMetals* **1998**, *11*, 13–20.
356. D. K. Wong; J. Gobin; M. A. Horwitz; B. W. Gibson, *J. Bacteriol.* **1996**, *178*, 6394–6398.
357. W. G. McCullough; R. S. Merkal, *Curr. Microbiol.* **1982**, *7*, 337–341.
358. R. Barclay; D. F. Ewing; C. Ratledge, *J. Bacteriol.* **1985**, *164*, 896–903.
359. B. D. Schwartz; J. J. De Voss, *Tetrahedron Lett.* **2001**, *42*, 3653–3655.
360. J. Gobin; D. K. Wong; B. W. Gibson; M. A. Horwitz, *Infect. Immun.* **1999**, *67*, 2035–2039.
361. R. M. Hall; C. Ratledge, *J. Gen. Microbiol.* **1985**, *131*, 1691–1696.
362. A. J. White; G. A. Snow, *Biochem. J.* **1969**, *111*, 785–792.
363. C. Ratledge; G. A. Snow, *Biochem. J.* **1974**, *139*, 407–413.
364. K. Suenaga; S. Kokubo; C. Shinohara; T. Tsuji; D. Uemura, *Tetrahedron Lett.* **1999**, *40*, 1945–1948.
365. S. Kokubo; K. Suenaga; C. Shinohara; T. Tsuji; D. Uemura, *Tetrahedron* **2000**, *56*, 6435–6440.
366. A. Nemoto; Y. Hoshino; K. Yazawa; A. Ando; Y. Mikami; H. Komaki; Y. Tanaka; U. Grafe, *J. Antibiot.* **2002**, *55*, 593–597.
367. C. Ratledge, Mycobactins and Nocobactins. In *CRC Handbook of Microbiology*; 2nd ed.; A. I. Laskin, H. A. Lechevalier, Eds.; CRC Press, 1982; Vol. IV, pp 575–580.
368. M. Tsuda; M. Yamakawa; S. Oka; Y. Tanaka; Y. Hoshino; Y. Mikami; A. Sato; H. Fujiwara; Y. Ohizumi; J. Kobayashi, *J. Nat. Prod.* **2005**, *68*, 462–464.
369. J. M. Mitchell; J. T. Shaw, *Org. Lett.* **2007**, *9*, 1679–1681.
370. M. Tsukamoto; K. Murooka; S. Nakajima; S. Abe; H. Suzuki; K. Hirano; H. Kondo; K. Kojiri; H. Suda, *J. Antibiot. (Tokyo)* **1997**, *50*, 815–821.
371. Y. Murakami; S. Kato; M. Nakajima; M. Matsuoka; H. Kawai; K. ShinYa; H. Seto, *J. Antibiot.* **1996**, *49*, 839–845.
372. Y. Ikeda; H. Nonaka; T. Furumai; H. Onaka; Y. Igarashi, *J. Nat. Prod.* **2005**, *68*, 1061–1065.
373. R. Barclay; V. Furst; I. Smith, *J. Med. Microbiol.* **1992**, *37*, 286–290.
374. E. Hough; D. Rogers, *Biochem. Biophys. Res. Commun.* **1974**, *57*, 73–77.
375. G. Xu; J. S. Martinez; J. T. Groves; A. Butler, *J. Am. Chem. Soc.* **2002**, *124*, 13408–13415.
376. L. E. N. Quadri; J. Sello; T. A. Keating; P. H. Weinreb; C. T. Walsh, *Chem. Biol.* **1998**, *5*, 631–645.
377. R. Krithika; U. Marathe; P. Saxena; M. Z. Ansari; D. Mohanty; R. S. Gokhale, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 2069–2074.
378. A. J. Harrison; M. Yu; T. Gardenborg; M. Middleditch; R. J. Ramsay; E. N. Baker; J. S. Lott, *J. Bacteriol.* **2006**, *188*, 6081–6091.
379. J. Zwahlen; S. Kolappan; R. Zhou; C. Kisker; P. J. Tonge, *Biochemistry (Mosc.)* **2007**, *46*, 954–964.
380. A. T. Hudson; I. M. Campbell; R. Bentley, *Biochemistry (Mosc.)* **1970**, *9*, 3988–3992.
381. J. G. Dain; R. Bentley, *Bioorg. Chem.* **1971**, *1*, 374–379.
382. T. Duerfahrt; K. Eppelmann; R. Muller; M. A. Marahiel, *Chem. Biol.* **2004**, *11*, 261–271.
383. J. E. Tateson, *Biochem. J.* **1970**, *118*, 747–753.
384. B. A. Frankel; J. S. Blanchard, *Arch. Biochem. Biophys.* **2008**, *477*, 259–266.
385. D. B. Moody; D. C. Young; T. Y. Cheng; J. P. Rosat; C. Roura-Mir; P. B. O'Connor; D. M. Zajonc; A. Walz; M. J. Miller; S. B. Levery; I. A. Wilson; C. E. Costello; M. B. Brenner, *Science* **2004**, *303*, 527–531.
386. M. Wolpert; B. Gust; B. Kammerer; L. Heide, *Microbiology* **2007**, *153*, 1413–1423.
387. G. M. Rodriguez; M. I. Voskuil; B. Gold; G. K. Schoolnik; I. Smith, *Infect. Immun.* **2002**, *70*, 3371–3381.
388. B. Gold; G. M. Rodriguez; S. A. Marras; M. Pentecost; I. Smith, *Mol. Microbiol.* **2001**, *42*, 851–865.
389. S. Yellaboina; S. Ranjan; V. Vindal; A. Ranjan, *FEBS Lett.* **2006**, *580*, 2567–2576.
390. G. A. Snow, *Biochem. J.* **1965**, *97*, 166–175.
391. B. B. LaMarca; W. Zhu; J. E. Arceneaux; B. R. Byers; M. D. Lundrigan, *J. Bacteriol.* **2004**, *186*, 374–382.
392. D. A. Mitchison, *Int. J. Tuberc. Lung Dis.* **2000**, *4*, 796–806.
393. C. Ratledge; K. A. Brown, *Am. Rev. Respir. Dis.* **1972**, *106*, 774–776.
394. T. Adilakshmi; P. D. Ayling; C. Ratledge, *J. Bacteriol.* **2000**, *182*, 264–271.
395. S. A. Naser; R. F. Gillespie; N. A. Naser; F. A. K. El-Zaatari, *Curr. Microbiol.* **1998**, *37*, 373–379.
396. J. R. Chipperfield; C. Ratledge, *BioMetals* **2000**, *13*, 165–168.
397. S. T. Ong; J. Z. S. Ho; B. Ho; J. L. Ding, *Immunobiology* **2006**, *211*, 295–314.
398. J. Gobin; M. A. Horwitz, *J. Exp. Med.* **1996**, *183*, 1527–1532.
399. D. Wagner; J. Maser; B. Lai; Z. Cai; C. E. Barry, III; K. Honer Zu Bentrup; D. G. Russell; L. E. Bermudez, *J. Immunol.* **2005**, *174*, 1491–1500.
400. A. Kahnert; P. Seiler; M. Stein; S. Bandermann; K. Hahnke; H. Mollenkopf; S. H. Kaufmann, *Eur. J. Immunol.* **2006**, *36*, 631–647.
401. A. Ojha; G. F. Hatfull, *Mol. Microbiol.* **2007**, *66*, 468–483.
402. Y. C. Manabe; B. J. Saviola; L. Sun; J. R. Murphy; W. R. Bishai, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 12844–12848.
403. J. Y. Hou; J. E. Graham; J. E. Clark-Curtiss, *Infect. Immun.* **2002**, *70*, 3714–3726.
404. J. Timm; F. A. Post; L. G. Bekker; G. B. Walther; H. C. Wainwright; R. Manganelli; W. T. Chan; L. Tsenova; B. Gold; I. Smith; G. Kaplan; J. D. McKinney, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14321–14326.
405. S. J. Waddell; P. D. Butcher, *Curr. Mol. Med.* **2007**, *7*, 287–296.
406. N. Lounis; C. Truffot-Pernot; J. Grosset; V. R. Gordeuk; J. R. Boelaert, *J. Clin. Virol.* **2001**, *20*, 123–126.
407. M. S. Gomes; J. R. Boelaert; R. Appelberg, *J. Clin. Virol.* **2001**, *20*, 117–122.
408. A. Trousseau, *Lectures on Clinical Medicine*, 3rd ed.; The New Sydenham Society: London, 1872; Vol. 5.
409. I. T. Gangaidzo; V. M. Moyo; E. Mvundura; G. Aggrey; N. L. Murphree; H. Khumalo; T. Saungweme; I. Kasvosve; Z. A. Gomo; T. Rouault; J. R. Boelaert; V. R. Gordeuk, *J. Infect. Dis.* **2001**, *184*, 936–939.
410. J. R. Boelaert; S. J. Vandecasteele; R. Appelberg; V. R. Gordeuk, *J. Infect. Dis.* **2007**, *195*, 1745–1753.
411. V. R. Gordeuk; C. E. McLaren; A. P. MacPhail; G. Deichsel; T. H. Bothwell, *Blood* **1996**, *87*, 3470–3476.
412. B. F. Matzanke; R. Bohnke; U. Mollmann; R. Reissbrodt; V. Schunemann; A. X. Trautwein, *BioMetals* **1997**, *10*, 193–203.
413. B. F. Matzanke; U. Mollmann; R. Reissbrodt; V. Schunemann; A. X. Trautwein, *Hyperfine Interact.* **1998**, *112*, 123–128.

414. L. P. Macham; C. Ratledge; J. C. Nocton, *Infect. Immun.* **1975**, *12*, 1242–1251.
415. C. Ratledge; P. V. Patel; J. Mundy, *J. Gen. Microbiol.* **1982**, *128*, 1559–1565.
416. C. Ratledge; B. J. Marshall, *Biochim. Biophys. Acta* **1972**, *279*, 58–74.
417. C. Ratledge; M. Ewing, *Microbiology (UK)* **1996**, *142*, 2207–2212.
418. M. C. Stephenson; C. Ratledge, *J. Gen. Microbiol.* **1979**, *110*, 193–202.
419. C. Ratledge, *Tuberculosis* **2004**, *84*, 110–130.
420. G. M. Rodriguez; I. Smith, *J. Bacteriol.* **2006**, *188*, 424–430.
421. A. Farhana; S. Kumar; S. S. Rathore; P. C. Ghosh; N. Z. Ehtesham; A. K. Tyagi; S. E. Hasnain, *PLoS ONE* **2008**, *3*, e2087.
422. M. Luo; E. A. Fadeev; J. T. Groves, *Nat. Chem. Biol.* **2005**, *1*, 149–153.
423. P. Peyron; J. Vaubourgeix; Y. Poquet; F. Levillain; C. Botanch; F. Bardou; M. Daffé; J. F. Emile; B. Marchou; P. J. Cardona; C. de Chastellier; F. Altare, *PLoS Pathog.* **2008**, *4*, e1000204.
424. N. J. Garton; S. J. Waddell; A. L. Sherratt; S. M. Lee; R. J. Smith; C. Senner; J. Hinds; K. Rajakumar; R. A. Adegbola; G. S. Besra; P. D. Butcher; M. R. Barer, *PLoS Med.* **2008**, *5*, e75.
425. S. Xu; A. Cooper; S. Sturgill-Koszycki; T. van Heyningen; D. Chatterjee; I. Orme; P. Allen; D. G. Russell, *J. Immunol.* **1994**, *153*, 2568–2578.
426. C. E. Barry, III; H. Boshoff, *Nat. Chem. Biol.* **2005**, *1*, 127–128.
427. J. S. Martinez; J. N. Carter-Franklin; E. L. Mann; J. D. Martin; M. G. Haygood; A. Butler, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3754–3759.
428. J. Hu; M. J. Miller, *J. Am. Chem. Soc.* **1997**, *119*, 3462–3468.
429. Y. P. Xu; M. J. Miller, *J. Org. Chem.* **1998**, *63*, 4314–4322.
430. A. Mukai; T. Fukai; Y. Matsumoto; J. Ishikawa; Y. Hoshino; K. Yazawa; K. I. Harada; Y. Mikami, *J. Antibiot.* **2006**, *59*, 366–369.
431. M. A. Llamas; M. J. Mooij; M. Sparrius; C. M. Vandenbroucke-Grauls; C. Ratledge; W. Bitter, *Mol. Microbiol.* **2008**, *67*, 458–472.
432. K. A. Fennell; M. J. Miller, *Org. Lett.* **2007**, *9*, 1683–1685.
433. A. J. Walz; M. J. Miller, *Tetrahedron Lett.* **2007**, *48*, 5103–5105.
434. A. Nayyar; R. Jain, *Curr. Med. Chem.* **2005**, *12*, 1873–1886.
435. J. A. Ferreras; J. S. Ryu; F. Di Lello; D. S. Tan; L. E. Quadri, *Nat. Chem. Biol.* **2005**, *1*, 29–32.
436. R. V. Somu; H. Boshoff; C. Qiao; E. M. Bennett; C. E. Barry, III; C. C. Aldrich, *J. Med. Chem.* **2006**, *49*, 31–34.
437. J. Vannada; E. M. Bennett; D. J. Wilson; H. I. Boshoff; C. E. Barry, III; C. C. Aldrich, *Org. Lett.* **2006**, *8*, 4707–4710.
438. R. V. Somu; D. J. Wilson; E. M. Bennett; H. I. Boshoff; L. Celia; B. J. Beck; C. E. Barry, III; C. C. Aldrich, *J. Med. Chem.* **2006**, *49*, 7623–7635.
439. C. H. Qiao; A. Gupte; H. I. Boshoff; D. J. Wilson; E. M. Bennett; R. V. Somu; C. E. Barry; C. C. Aldrich, *J. Med. Chem.* **2007**, *50*, 6080–6094.
440. J. Neres; N. P. Labello; R. V. Somu; H. I. Boshoff; D. J. Wilson; J. Vannada; L. Chen; C. E. Barry; E. M. Bennett; C. C. Aldrich, *J. Med. Chem.* **2008**, *51*, 5349–5370.
441. A. Gupte; H. I. Boshoff; D. J. Wilson; J. Neres; N. P. Labello; R. V. Somu; C. Xing; C. E. Barry; C. C. Aldrich, *J. Med. Chem.* **2008**, *51*, 7495–7507.
442. K. L. Stirrett; J. A. Ferreras; V. Jayaprakash; B. N. Sinha; T. Ren; L. E. Quadri, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 2662–2668.
443. P. Arora; A. Goyal; V. T. Natarajan; E. Rajakumara; P. Verma; R. Gupta; M. Yousuf; O. A. Trivedi; D. Mohanty; A. Tyagi; R. Sankaranarayanan; R. S. Gokhale, *Nat. Chem. Biol.* **2009**, *5*, 166–173.
444. A. Gupte; M. Subramanian; R. P. Remmel; C. C. Aldrich, *J. Labelled Comp. Radiopharm.* **2008**, *51*, 118–122.
445. P. M. B. Pahl; M. A. Horwitz; K. B. Horwitz; L. D. Horwitz, *Breast Cancer Res. Treat.* **2001**, *69*, 69–79.
446. Y. K. Hodges; W. E. Antholine; L. D. Horwitz, *Biochem. Biophys. Res. Commun.* **2004**, *315*, 595–598.
447. H. Sakagami; M. Ishihara; Y. Hoshino; J. Ishikawa; Y. Mikami; T. Fukai, *In Vivo* **2005**, *19*, 277–282.
448. F. Yokokawa; K. Izumi; J. Omata; T. Shioiri, *Tetrahedron* **2000**, *56*, 3027–3034.
449. K. A. Fennell; U. Mollmann; M. J. Miller, *J. Org. Chem.* **2008**, *73*, 1018–1024.
450. V. Jayaprakash; B. N. Sinha; G. Ucar; A. Ercan, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6362–6368.
451. M. J. Hauser; H. Baier, *Drug Intell. Clin. Pharm.* **1982**, *16*, 617–618.
452. C. Roura-Mir; L. S. Wang; T. Y. Cheng; I. Matsunaga; C. C. Dascher; S. L. Peng; M. J. Fenton; C. Kirschning; D. B. Moody, *J. Immunol.* **2005**, *175*, 1758–1766.
453. M. A. Holmes; W. Paulsene; X. Jide; C. Ratledge; R. K. Strong, *Structure* **2005**, *13*, 29–41.
454. L. Kjeldsen; A. H. Johnsen; H. Sengelov; N. Borregaard, *J. Biol. Chem.* **1993**, *268*, 10425–10432.
455. L. Kjeldsen; J. B. Cowland; N. Borregaard, *Biochim. Biophys. Acta* **2000**, *1482*, 272–283.
456. L. Kjeldsen; D. F. Bainton; H. Sengelov; N. Borregaard, *Blood* **1994**, *83*, 799–807.
457. T. H. Flo; K. D. Smith; S. Sato; D. J. Rodriguez; M. A. Holmes; R. K. Strong; S. Akira; A. Aderem, *Nature* **2004**, *432*, 917–921.
458. D. H. Goetz; M. A. Holmes; N. Borregaard; M. E. Bluhm; K. N. Raymond; R. K. Strong, *Mol. Cell* **2002**, *10*, 1033–1043.
459. S. Triebel; J. Blaser; H. Reinke; H. Tschesche, *FEBS Lett.* **1992**, *314*, 386–388.
460. J. L. Taylor; J. M. Hattle; S. A. Dreitz; J. M. Troudt; L. S. Izzo; R. J. Basaraba; I. M. Orme; L. M. Matrisian; A. A. Izzo, *Infect. Immun.* **2006**, *74*, 6135–6144.
461. S. Y. Xu; K. Pauksen; P. Venge, *Scand. J. Clin. Lab. Invest.* **1995**, *55*, 125–131.
462. H. Saiga; J. Nishimura; H. Kuwata; M. Okuyama; S. Matsumoto; S. Sato; M. Matsumoto; S. Akira; Y. Yoshikai; K. Honda; M. Yamamoto; K. Takeda, *J. Immunol.* **2008**, *181*, 8521–8527.
463. M. A. Fischbach; H. Lin; D. R. Liu; C. T. Walsh, *Nat. Chem. Biol.* **2006**, *2*, 132–138.
464. R. J. Abergel; E. G. Moore; R. K. Strong; K. N. Raymond, *J. Am. Chem. Soc.* **2006**, *128*, 10998–10999.
465. R. J. Abergel; M. C. Clifton; J. C. Pizarro; J. A. Warner; D. K. Shuh; R. K. Strong; K. N. Raymond, *J. Am. Chem. Soc.* **2008**, *130*, 11524–11534.
466. T. M. Hoette; R. J. Abergel; J. Xu; R. K. Strong; K. N. Raymond, *J. Am. Chem. Soc.* **2008**, *130*, 17584–17592.
467. R. S. Lambrecht; M. T. Collins, *Microb. Pathog.* **1993**, *14*, 229–238.
468. WHO, *Buruli Ulcer, Mycobacterium Ulcerans Infection*; WHO: Geneva, 2000.
469. S. Rondini; E. Mensah-Quainoo; T. Junghanss; G. Pluschke, *J. Clin. Microbiol.* **2006**, *44*, 4273–4275.

470. T. Y. Quek; M. J. Henry; J. A. Pasco; D. P. O'Brien; P. D. Johnson; A. Hughes; A. C. Cheng; J. Redden-Hoare; E. Athan, *Med. J. Aust.* **2007**, *187*, 561–563.
471. P. D. Johnson; J. Azuolas; C. J. Lavender; E. Wishart; T. P. Stinear; J. A. Hayman; L. Brown; G. A. Jenkin; J. A. Fyfe, *Emerg. Infect. Dis.* **2007**, *13*, 1653–1660.
472. C. J. Lavender; S. N. Senanayake; J. A. Fyfe; J. A. Buntine; M. Globan; T. P. Stinear; J. A. Hayman; P. D. Johnson, *Med. J. Aust.* **2007**, *186*, 62–63.
473. P. MacCallum; J. C. Tolhurst; G. Buckle; H. A. Sissons, *J. Pathol. Bacteriol.* **1948**, *60*, 93–122.
474. E. Coutanceau; L. Marsollier; R. Brosch; E. Perret; P. Goossens; M. Tanguy; S. T. Cole; P. L. Small; C. Demangel, *Cell. Microbiol.* **2005**, *7*, 1187–1196.
475. E. Torrado; A. G. Fraga; A. G. Castro; P. Stragier; W. M. Meyers; F. Portaels; M. T. Silva; J. Pedrosa, *Infect. Immun.* **2007**, *75*, 977–987.
476. S. Adusumilli; A. Mve-Obiang; T. Sparer; W. Meyers; J. Hayman; P. L. Small, *Cell. Microbiol.* **2005**, *7*, 1295–1304.
477. K. M. George; D. Chatterjee; G. Gunawardana; D. Welty; J. Hayman; R. Lee; P. L. Small, *Science* **1999**, *283*, 854–857.
478. M. S. Oliveira; A. G. Fraga; E. Torrado; A. G. Castro; J. P. Pereira; A. L. Filho; F. Milanezi; F. C. Schmitt; W. M. Meyers; F. Portaels; M. T. Silva; J. Pedrosa, *Infect. Immun.* **2005**, *73*, 6299–6310.
479. WHO, *Provisional Guidance on the Role of Specific Antibiotics in the Management of Mycobacterium Ulcerans Disease (Buruli Ulcer)*, WHO: Geneva, 2004.
480. A. Chauty; M. F. Ardant; A. Adeye; H. Euverte; A. Guedenon; C. Johnson; J. Aubry; E. Nuermberger; J. Grosset, *Antimicrob. Agents Chemother.* **2007**, *51*, 4029–4035.
481. D. Schutte; A. Um-Boock; E. Mensah-Quainoo; P. Itin; P. Schmid; G. Pluschke, *PLoS Negl. Trop. Dis.* **2007**, *1*, e2.
482. D. Schutte; A. Umboock; G. Pluschke, *Br. J. Dermatol.* **2009**, *160*, 273–283.
483. A. E. Kiszewski; E. Becerril; L. D. Aguilar; I. T. Kader; W. Myers; F. Portaels; R. Hernandez Pando, *Clin. Exp. Immunol.* **2006**, *143*, 445–451.
484. W. M. Meyers; W. M. Shelly; D. H. Connor; E. K. Meyers, *Am. J. Trop. Med. Hyg.* **1974**, *23*, 919–923.
485. D. S. Walsh; F. Portaels; W. M. Meyers, *Trans. R. Soc. Trop. Med. Hyg.* **2008**, *102*, 969–978.
486. L. Marsollier; J. Aubry; E. Coutanceau; J. P. Andre; P. L. Small; G. Milon; P. Legras; S. Guadagnini; B. Carbonnelle; S. T. Cole, *Cell. Microbiol.* **2005**, *7*, 935–943.
487. L. Marsollier; J. P. Andre; W. Frigui; G. Reysset; G. Milon; B. Carbonnelle; J. Aubry; S. T. Cole, *Cell. Microbiol.* **2007**, *9*, 347–355.
488. L. Marsollier; T. Stinear; J. Aubry; J. P. Saint Andre; R. Robert; P. Legras; A. L. Manceau; C. Audrain; S. Bourdon; H. Kouakou; B. Carbonnelle, *Appl. Environ. Microbiol.* **2004**, *70*, 1097–1103.
489. L. Marsollier; P. Brodin; M. Jackson; J. Kordulakova; P. Tafelmeyer; E. Carbonnelle; J. Aubry; G. Milon; P. Legras; J. P. Andre; C. Leroy; J. Cottin; M. L. Guillou; G. Reysset; S. T. Cole, *PLoS Pathog.* **2007**, *3*, e62.
490. M. E. Benbow; H. Williamson; R. Kimbirauskas; M. D. McIntosh; R. Kolar; C. Quaye; F. Akpabey; D. Boakye; P. Small; R. W. Merritt, *Emerg. Infect. Dis.* **2008**, *14*, 1247–1254.
491. T. Y. Quek; E. Athan; M. J. Henry; J. A. Pasco; J. Redden-Hoare; A. Hughes; P. D. Johnson, *Emerg. Infect. Dis.* **2007**, *13*, 1661–1666.
492. R. Pouillot; G. Matias; C. M. Wondje; F. Portaels; N. Valin; F. Ngos; A. Njikap; L. Marsollier; A. Fontanet; S. Eyangoh, *PLoS Negl. Trop. Dis.* **2007**, *1*, e101.
493. F. Portaels; W. M. Meyers; A. Ablordey; A. G. Castro; K. Chemlal; P. de Rijk; P. Elsen; K. Fissette; A. G. Fraga; R. Lee; E. Mahrous; P. L. Small; P. Stragier; E. Torrado; A. Van Aerde; M. T. Silva; J. Pedrosa, *PLoS Negl. Trop. Dis.* **2008**, *2*, e178.
494. D. Diaz; H. Dobeli; D. Yeboah-Manu; E. Mensah-Quainoo; A. Friedlein; N. Soder; S. Rondini; T. Bodmer; G. Pluschke, *Clin. Vaccine Immunol.* **2006**, *13*, 1314–1321.
495. T. P. Stinear; T. Seemann; S. Pidot; W. Frigui; G. Reysset; T. Garnier; G. Meurice; D. Simon; C. Bouchier; L. Ma; M. Tichit; J. L. Porter; J. Ryan; P. D. Johnson; J. K. Davies; G. A. Jenkin; P. L. Small; L. M. Jones; F. Tekaia; F. Laval; M. Daffé; J. Parkhill; S. T. Cole, *Genome Res.* **2007**, *17*, 192–200.
496. C. Demangel; T. P. Stinear; S. T. Cole, *Nat. Rev. Microbiol.* **2009**, *7*, 50–60.
497. M. J. Yip; J. L. Porter; J. A. Fyfe; C. J. Lavender; F. Portaels; M. Rhodes; H. Kator; A. Colorni; G. A. Jenkin; T. Stinear, *J. Bacteriol.* **2007**, *189*, 2021–2029.
498. T. P. Stinear; H. Hong; W. Frigui; M. J. Pryor; R. Brosch; T. Garnier; P. F. Leadlay; S. T. Cole, *J. Bacteriol.* **2005**, *187*, 1668–1676.
499. G. Gunawardana; D. Chatterjee; K. M. George; P. Brennan; D. Whittern; P. L. C. Small, *J. Am. Chem. Soc.* **1999**, *121*, 6092–6093.
500. S. Fidanze; F. Song; M. Szlosek-Pinaud; P. L. Small; Y. Kishi, *J. Am. Chem. Soc.* **2001**, *123*, 10117–10118.
501. A. Mve-Obiang; R. E. Lee; F. Portaels; P. L. Small, *Infect. Immun.* **2003**, *71*, 774–783.
502. T. C. Judd; A. Bischoff; Y. Kishi; S. Adusumilli; P. L. Small, *Org. Lett.* **2004**, *6*, 4901–4904.
503. H. Hong; J. B. Spencer; J. L. Porter; P. F. Leadlay; T. Stinear, *ChemBioChem* **2005**, *6*, 643–648.
504. A. Mve-Obiang; R. E. Lee; E. S. Umstot; K. A. Trott; T. C. Grammer; J. M. Parker; B. S. Ranger; R. Grainger; E. A. Mahrous; P. L. Small, *Infect. Immun.* **2005**, *73*, 3307–3312.
505. H. Hong; T. Stinear; P. Skelton; J. B. Spencer; P. F. Leadlay, *Chem. Commun. (Camb.)* **2005**, 4306–4308.
506. S. Aubry; R. E. Lee; E. A. Mahrous; P. L. Small; D. Beachboard; Y. Kishi, *Org. Lett.* **2008**, *10*, 5385–5388.
507. B. S. Ranger; E. A. Mahrous; L. Mosi; S. Adusumilli; R. E. Lee; A. Colorni; M. Rhodes; P. L. Small, *Infect. Immun.* **2006**, *74*, 6037–6045.
508. H. J. Kim; Y. Kishi, *J. Am. Chem. Soc.* **2008**, *130*, 1842–1844.
509. H. Hong; T. Stinear; J. Porter; C. Demangel; P. F. Leadlay, *ChemBioChem* **2007**, *8*, 2043–2047.
510. S. J. Pidot; H. Hong; T. Seemann; J. L. Porter; M. J. Yip; A. Men; M. Johnson; P. Wilson; J. K. Davies; P. F. Leadlay; T. P. Stinear, *BMC Genomics* **2008**, *9*, 462.
511. A. K. Daniel; R. E. Lee; F. Portaels; P. L. Small, *Infect. Immun.* **2004**, *72*, 123–132.
512. K. M. George; L. Pascopella; D. M. Welty; P. L. Small, *Infect. Immun.* **2000**, *68*, 877–883.
513. D. S. Walsh; W. M. Meyers; F. Portaels; J. E. Lane; D. Mongkolsirichaikul; K. Hussem; P. Gosi; K. S. Myint, *Am. J. Trop. Med. Hyg.* **2005**, *73*, 410–415.

514. D. S. Snyder; P. L. Small, *Microb. Pathog.* **2003**, *34*, 91–101.
515. M. Goto; K. Nakanaga; T. Aung; T. Hamada; N. Yamada; M. Nomoto; S. Kitajima; N. Ishii; S. Yonezawa; H. Saito, *Am. J. Pathol.* **2006**, *168*, 805–811.
516. J. En; M. Goto; K. Nakanaga; M. Higashi; N. Ishii; H. Saito; S. Yonezawa; H. Hamada; P. L. Small, *Infect. Immun.* **2008**, *76*, 2002–2007.
517. M. Pimsler; T. A. Sponsler; W. M. Meyers, *J. Infect. Dis.* **1988**, *157*, 577–580.
518. A. A. Pahlevan; D. J. Wright; C. Andrews; K. M. George; P. L. Small; B. M. Foxwell, *J. Immunol.* **1999**, *163*, 3928–3935.
519. E. Coutanceau; J. Decalf; A. Martino; A. Babon; N. Winter; S. T. Cole; M. L. Albert; C. Demangel, *J. Exp. Med.* **2007**, *204*, 1395–1403.
520. E. Torrado; S. Adusumilli; A. G. Fraga; P. L. Small; A. G. Castro; J. Pedrosa, *Infect. Immun.* **2007**, *75*, 3979–3988.
521. H. Hong; E. Coutanceau; M. Leclerc; L. Caleechurn; P. F. Leadlay; C. Demangel, *PLoS Negl. Trop. Dis.* **2008**, *2*, e325.
522. R. E. Simmonds; F. V. Lali; T. Smallie; P. L. Small; B. M. Foxwell, *J. Immunol.* **2009**, *182*, 2194–2202.
523. J. C. Atherton; P. Cao; R. M. Peek, Jr.; M. K. Tummuru; M. J. Blaser; T. L. Cover, *J. Biol. Chem.* **1995**, *270*, 17771–17777.
524. T. Sakai; N. Asai; A. Okuda; N. Kawamura; Y. Mizui, *J. Antibiot. (Tokyo)* **2004**, *57*, 180–187.
525. Y. Kotake; K. Sagane; T. Owa; Y. Mimori-Kiyosue; H. Shimizu; M. Uesugi; Y. Ishihama; M. Iwata; Y. Mizui, *Nat. Chem. Biol.* **2007**, *3*, 570–575.
526. T. P. Stinear; A. Mve-Obiang; P. L. Small; W. Frigui; M. J. Pryor; R. Brosch; G. A. Jenkin; P. D. Johnson; J. K. Davies; R. E. Lee; S. Adusumilli; T. Garnier; S. F. Haydock; P. F. Leadlay; S. T. Cole, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 1345–1349.
527. T. P. Stinear; M. J. Pryor; J. L. Porter; S. T. Cole, *Microbiology* **2005**, *151*, 683–692.
528. S. Bali; K. J. Weissman, *ChemBioChem* **2006**, *7*, 1935–1942.
529. J. L. Meier; T. Barrows-Yano; T. L. Foley; C. L. Wike; M. D. Burkart, *Mol. Biosyst.* **2008**, *4*, 663–671.

Biographical Sketches

David B. Kastrinsky was born and raised in Fair Lawn, NJ. He attended Cornell University and received a B.A. in chemistry (*magna cum laude*) in 2000, and later received a Ph.D. in chemistry at the Scripps Research Institute in 2005. His graduate research focused on the synthesis and evaluation of a class of DNA-binding antitumor antibiotics called the duocarmycins. For his postdoctoral work at the National Institutes of Allergy and Infectious Diseases, he is developing methods to synthesize modified β-lactams to inhibit a new class of peptidoglycan-synthesizing enzymes. These enzymes might play a role in promoting the stationary-phase survival of *Mycobacterium tuberculosis* and it is hoped that they will provide therapeutic targets for latent disease.

Nicholas S. McBride was born in Falmouth, MA. He attended university at Worcester Polytechnic Institute, receiving a B.S. in biomedical engineering. His research experiences include tracking red tide organism in the Gulf of Maine, research on spinal cord regeneration, design of real-time tissue culture microscopy equipment, and development of a tissue-engineered myocardial scaffold. His interests turned to infectious diseases of the developing world when he experienced the urgent need for improved healthcare during his journeys abroad. Nicholas has also coordinated a project for community-based development in the informal settlements of Windhoek, Namibia. He is currently a predoctoral fellow working on a collaborative project at the National Institute of Allergy and Infectious Disease, Tuberculosis Research Section and the University of Cambridge, Department of Chemical Engineering and Biotechnology. He is researching biosensors for the detection of lipids from pathogenic mycobacteria, specifically using electrochemical techniques and protein engineering to create a sensor for the mycobactin siderophores of *Mycobacterium tuberculosis*.

Keriann M. Backus was born in Seattle, WA and attended Brown University, completing her B.S. in chemistry in 2007 where she conducted chemistry research and was awarded a research fellowship from Pfizer. She is currently a 2007 Rhodes Scholar and pursuing a Ph.D. in organic chemistry at the Oxford University. She is part of the Oxford–National Institutes of Health (NIH) graduate partnership program and is jointly supervised by Professor Ben Davis at Oxford and Dr. Clifton Barry at the NIH. Her thesis project entails the development of novel, carbohydrate-based ^{18}F-probes for PET imaging of tuberculosis. Keriann's broader research interests encompass chemical biology and infectious disease research and treatment.

Jason J. LeBlanc earned his B.S. in biotechnology in 2000 from Worcester Polytechnic Institute. He then pursued his Ph.D. in biology at Johns Hopkins University, studying the posttranscriptional regulation of retroviral RNAs. After graduation, he was a postdoctoral fellow at Northwestern University for 2 years studying the intracellular assembly of HIV-1 particles. During this time, he was awarded postdoctoral fellowships from the National Institutes of Health (NIH) and the American Cancer Society. Currently, Jason is a scientific research analyst and writer for the Tuberculosis Research Section at the National Institute of Allergy and Infectious Diseases at the NIH. His research interests are in molecular biology and studying the molecular basis of infectious diseases.

Clifton E. Barry, III, received his Ph.D. in organic and bioorganic chemistry in 1989 from Cornell University, studying the biosynthesis of complex natural products. Following post-doctoral research at Johns Hopkins University, he joined NIAID's Rocky Mountain Laboratories. In 1998, he was tenured as the chief of the Tuberculosis Research Section (TBRS) and relocated his laboratory to the NIH main campus in Maryland. He is the principal investigator on several overseas clinical trials studying tuberculosis and antituber-culosis chemotherapy in Korea, China, and Nigeria. Dr. Barry is a member of several editorial boards and has authored more than 120 research publications in tuberculosis since entering the field 16 years ago. In 2009, he was named by ScienceWatch as the most highly cited researcher working in the field of tuberculosis.

1.05 Microbial Type III Polyketide Synthases

Yohei Katsuyama and Sueharu Horinouchi, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

© 2010 Elsevier Ltd. All rights reserved.

1.05.1	**Introduction**	147
1.05.2	**Plant Type III PKSs**	149
1.05.3	**Type III PKSs from Actinobacteria**	151
1.05.3.1	RppA (1,3,6,8-Tetrahydroxynaphthalene Synthase)	151
1.05.3.2	Gcs	154
1.05.3.3	SrsA	156
1.05.3.4	DpgA	157
1.05.3.5	Type III PKSs from *Mycobacterium*	158
1.05.4	**Type III PKSs from Proteobacteria**	159
1.05.4.1	PhlD	159
1.05.4.2	ArsABCD	161
1.05.4.3	Type III PKSs from Myxobacteria	161
1.05.5	**Type III PKSs from Eukaryotic Microbes**	163
1.05.5.1	Type III PKSs from *Dictyostelium discoideum* (Steely 1, 2 (StlA, B))	163
1.05.5.2	Fungal Type III PKS	165
1.05.6	**Conclusions and Future Perspectives**	165
References		168

1.05.1 Introduction

Polyketides are compounds that are synthesized by polyketide synthases (PKSs) via condensation of several acetate units.[1] Since polyketides possess various important biological activities, they have been considered as a useful source of pharmaceutical agents. PKSs share a similar reaction mechanism to fatty acid synthases (FASs) and the most important reaction in common between PKSs and FASs is decarboxylative condensation of malonyl-coenzyme A (CoA) derivatives to acyl-thioesters (acyl carrier protein (ACP) or CoA) to synthesize β-ketoacyl thioesters catalyzed by ketosynthase (KS) activity (**Figure 1**).[2–4] In this reaction, acyltransferase (AT) activity and an ACP are also required. ACP carries an acyl or malonyl moiety at the thiol of a phosphopantetheinyl moiety attached to a serine residue. AT transfers the malonyl or acyl moiety of malonyl-CoA and acyl-CoA to the thiol of the phosphopantetheinyl moiety attached to ACP. This phosphopantetheinyl moiety is attached to ACP by the reaction of phosphopantetheinyl transferase (PPT) activity. In the FAS reaction, decarboxylative condensation is followed by stepwise reduction and dehydration of a β-ketoacyl thioester by ketoreductase (KR), dehydratase (DH), and enoylreductase (ER). Thus, FAS produces acyl-thioesters (usually acyl-ACP thioesters) resembling the original substrate, but with two additional methylene moieties (**Figure 1**). The iterative condensation and reduction result in fatty acid production. The difference between FAS and PKS is that all or some steps of these reductions are eliminated in the PKS reaction (**Figure 1**). Therefore, PKSs yield an acyl chain containing double bonds, hydroxyl residues, and ketone groups. In addition, acyl chains yielded by PKSs are usually modified or cyclized at the end of the reaction.

PKSs are divided into three groups on the basis of the organization of their domains (**Figure 2**).[1] Type I PKSs are large multifunctional proteins divided into modules (**Figure 2**).[5] Each module contains three to six domains and catalyzes condensation of one molecule of malonyl-CoA derivatives. Each module contains, at least, a KS domain, AT domain, and ACP, which are required for decarboxylative condensation of malonyl-CoA. In addition to these domains, some modules contain all or part of the three domains (KR, DH, and ER) responsible for reduction and dehydration. At the last module, type I PKSs

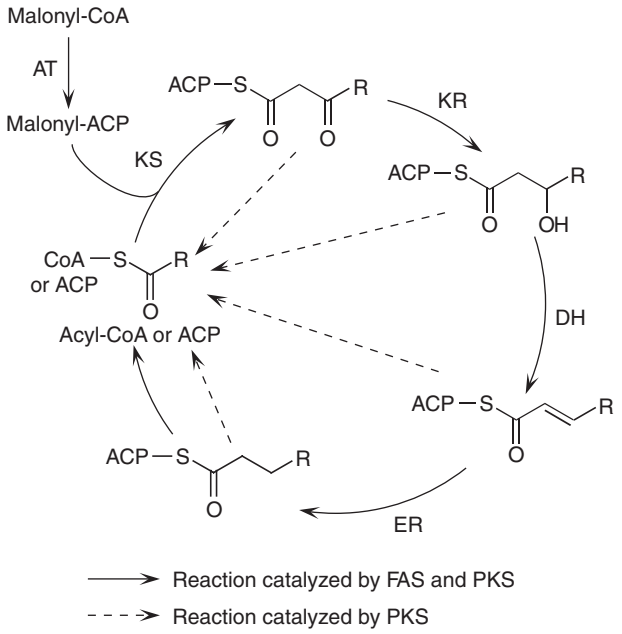


Figure 1 The reactions catalyzed by FAS and PKS. FAS catalyzes reduction and dehydration of β-ketoacyl-ACP synthesized by KS and produces saturated acyl-ACP, which then becomes the substrate of the next extension. In the PKS reactions, all or some of these reductions are eliminated and acyl-ACPs with a hydroxy group, double bond, or ketone group are synthesized.

usually harbor a thioesterase (TE) domain, which is responsible for the release of the resulting acyl chain from ACP (see Chapter 1.10).

In type II PKSs, these domains are on discrete proteins and type II PKSs therefore consist of subunits that possess different functions (**Figure 2**).[6] The subunits required for decarboxylative condensation are KS, chain length factor (CLF), ACP, and AT. CLF, which is also called KS_{β}, is a KS whose catalytic cysteine is not conserved. The complex of these enzymes is called a 'minimal PKS' (see Chapter 1.07).

Type III PKSs are a simple homodimer composed of an approximately 42 kDa KS and are multifunctional proteins catalyzing iterative condensation of malonyl-CoA to form a polyketide chain, cyclization, and aromatization of the resultant polyketide chain (**Figure 2**).[4] They are quite distinct from type I and II PKSs and use free CoA thioesters as a substrate, whereas type I and II PKSs use the substrate attached to the 4′-phosphopantetheine residue of ACP. Type III PKSs are usually responsible for the synthesis of aromatic bioactive compounds, such as flavonoids and stilbenoids in plants.

Until recently, type III PKSs had been found only from higher plants.[7] In 1999, type III PKSs from bacteria, *Streptomyces griseus*[8] and *Pseudomonas fluorescens* Q2-87,[8–10] were first reported. The discovery of bacterial type III PKSs gave rise to the idea that type III PKSs are distributed in a wide variety of microorganisms. In fact, recent explosive accumulation of the bacterial genome sequences has predicted that a variety of microorganisms contain putative type III PKSs. Among these, some were identified as actual type III PKSs through *in vivo* and *in vitro* analyses, such as those from *Streptomyces coelicolor* A3(2), *Mycobacterium tuberculosis*, *Bacillus subtilis*, and *Deinococcus radiodurans*.[7] However, the catalytic properties of most of the bacterial type III PKSs are unknown, which makes these enzymes attractive targets to study possible novel functions and apply them for the production of useful compounds by the so-called metabolic engineering. This chapter summarizes the catalytic functions of some microbial type III PKSs. An important, but thus far overlooked, role of alkylresorcinols, as products of a group of type III PKSs, in biological membranes in bacteria, fungi, and plants has also been implicated.

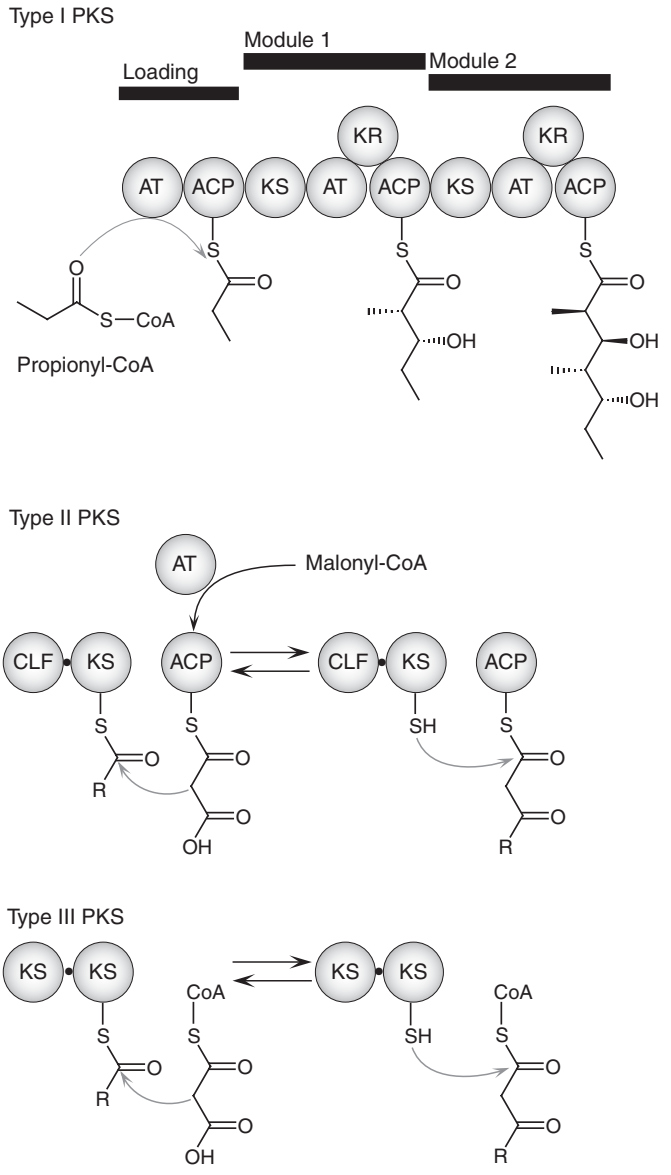


Figure 2 Three types of PKSs. Type I PKS is a multifunctional peptide that is divided into modules and domains. Type II PKS is a complex of discrete enzymes that possess an individual function. Type III PKS is a simple homodimer of ketosynthase.

1.05.2 Plant Type III PKSs

The most typical type III PKS is chalcone synthase (CHS), which catalyzes the key step in the phenylpropanoid pathway (**Figure 3**).[4] First, *p*-coumaroyl-CoA, which is called a starter substrate, attaches to the catalytic cysteine of CHS. Three molecules of malonyl-CoA, which is called an extender substrate, are then decarboxylatively condensed with the *p*-coumaroyl-CoA molecule attached to the enzyme to form a polyketide chain intermediate. The resultant linear polyketide chain is cyclized via intramolecular Claisen condensation and aromatization to yield naringenin chalcone (**Figure 3**). CHS of the type III PKS family catalyzes all of these reactions in a single domain. Similarly, most type III PKSs catalyze iterative condensation of an extender substrate with a starter substrate, cyclization, and aromatization of the resulting polyketide chain. The structural diversity of the products of type III PKSs is therefore dependent upon starter and extender substrate

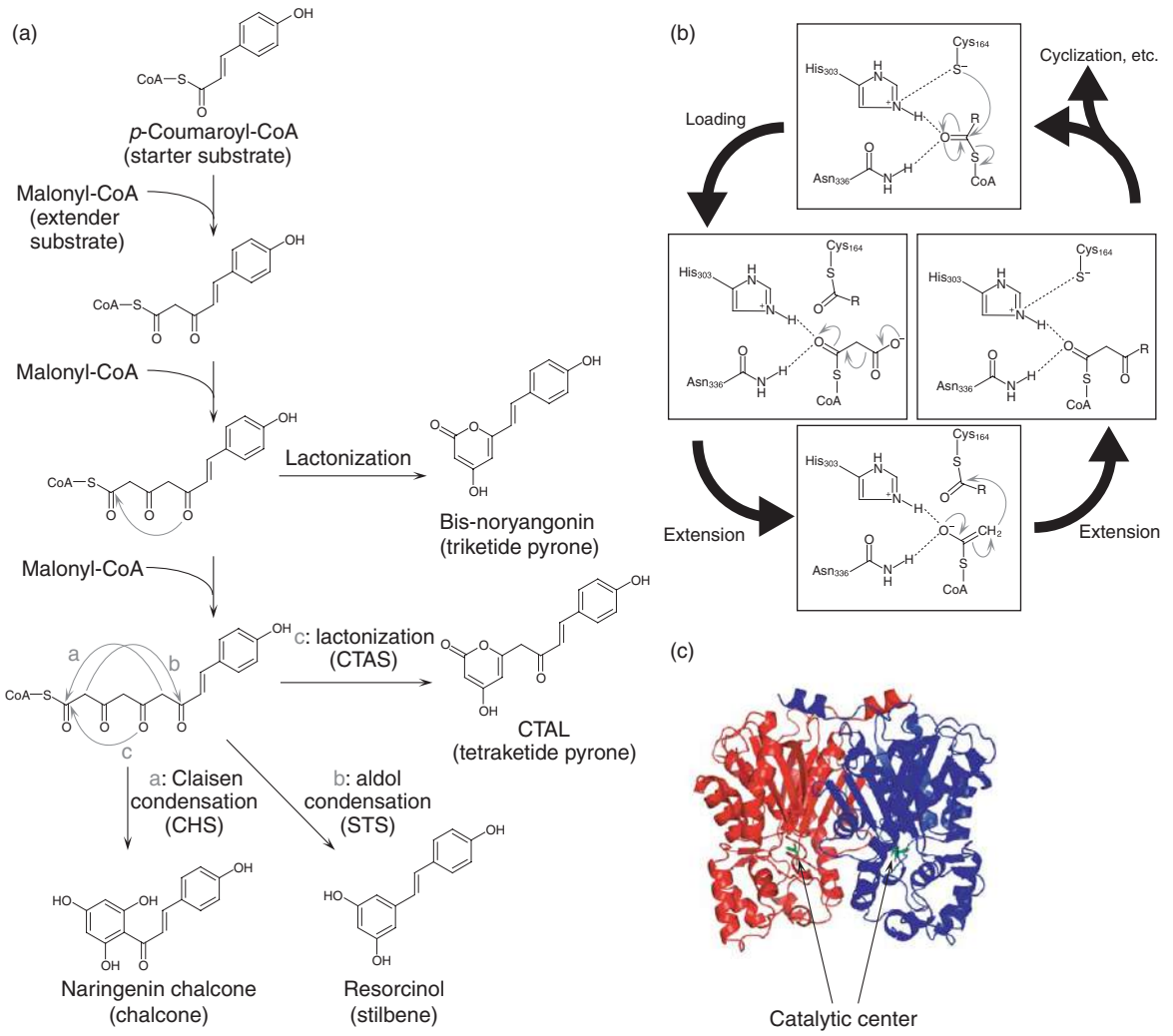


Figure 3 The reactions catalyzed by plant type III PKSs. (a) Chalcone synthase (CHS) catalyzes condensation of three molecules of malonyl-CoA to *p*-coumaroyl-CoA and Claisen condensation to yield naringenin chalcone. Stilbene synthase (STS) catalyzes aldol condensation, instead of Claisen condensation, and synthesizes resorcinol. Coumaroyl triacetic acid synthase (CTAS) catalyzes lactonization of a tetraketide intermediate and synthesizes a coumaroyl triacetic acid lactone (CTAL). When nonenzymatic lactonization occurs, triketide and tetraketide pyrones are synthesized. (b) The catalytic triad consisting of Cys, His, and Asn condenses malonyl-CoA to the starter CoA. (c) The crystal structure of CHS from *Medicago sativa* (alfalfa). Each monomer is colored in red and blue. Naringenin bound to the catalytic center is depicted in green.

specificity, the number of condensation of extender substrates, and the mode of cyclization. For example, stilbene synthase (STS)[4] shows a mode of cyclization different from CHS, synthesizing stilbenes via aldol condensation coupled with decarboxylation (**Figure 3**). Coumaroyl triacetic acid synthase (CTAS)[4] synthesizes a coumaroyl triacetic acid lactone (CTAL) via lactonization, instead of Claisen condensation, of a tetraketide intermediate (**Figure 3**). 2-Pyrone synthase (2-PS)[11] synthesizes a triacetic acid lactone (TAL) via condensation of acetyl-CoA with two malonyl-CoA molecules and subsequent lactonization. Recently, octaketide synthase (OKS), which synthesizes SEK4 and SEK4b from eight molecules of malonyl-CoA, and pentaketide chromone synthase (PCS), which synthesizes pentaketide chromone from five molecules of malonyl-CoA, have been found.[12–14] In addition, curcuminoid synthase (CUS), a type III PKS from the rice *Oryza sativa*, has been found to catalyze decarboxylative condensation of not only malonyl-CoA but also β-keto acids. This unusual condensation of a starter substrate with a β-keto acid, which acts as the second extender substrate, results in the

formation of a curcuminoid scaffold.[15] Most type III PKSs produce pyrones as by-products, as a result of nonenzymatic cyclization (lactonization) of polyketide chain intermediates. The diversity of polyketide products synthesized via type III PKSs is manifested through this functional diversity and post-PKS modification enzymes.

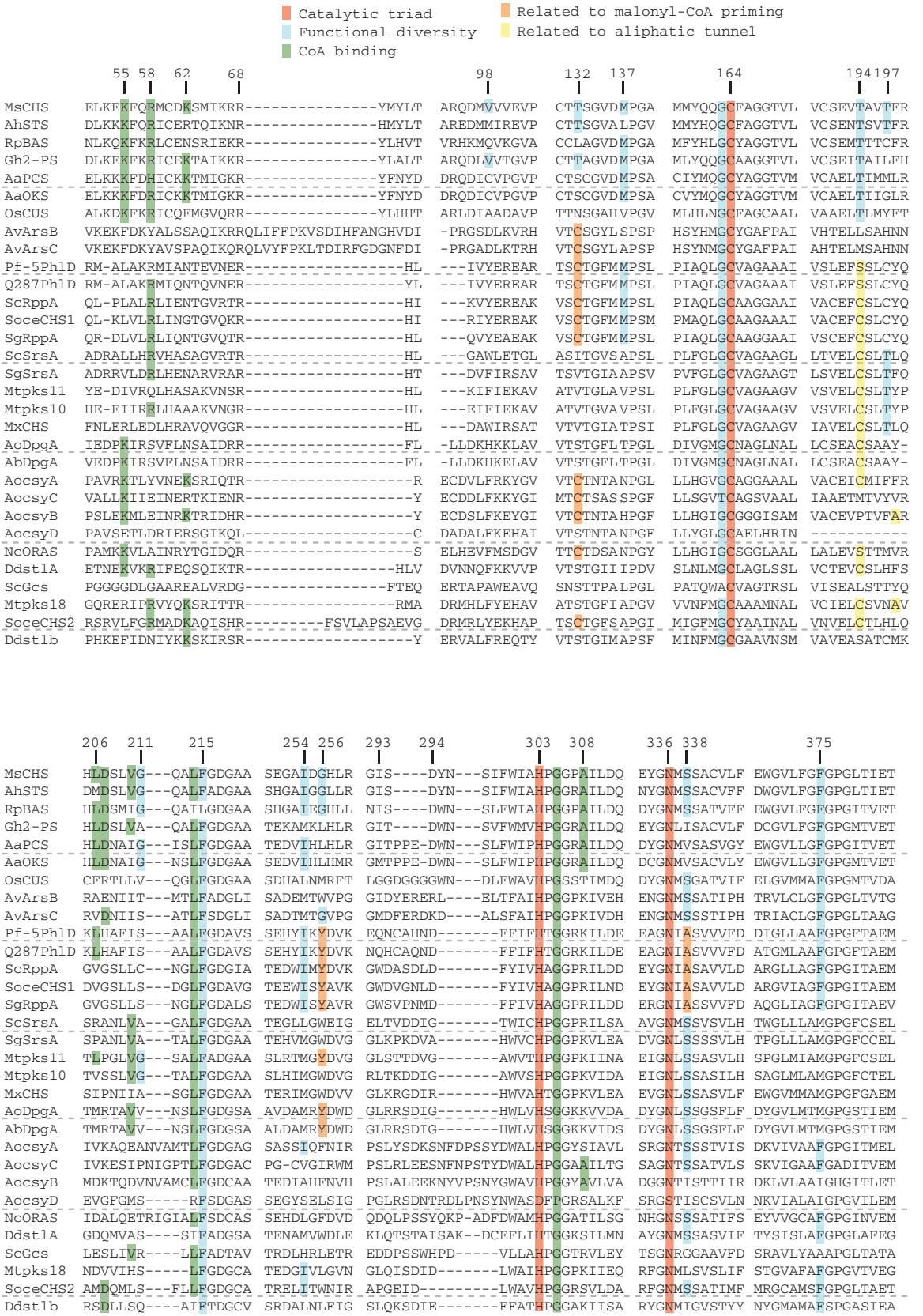
The iterative condensation of malonyl-CoA derivatives by CHS is catalyzed by three amino acid residues, Cys164, His303, and Asn336, forming a catalytic triad (**Figure 2(b)**).[4] These three residues are conserved in all known type III PKSs and are essential for catalysis (**Figure 4**). The crystal structure of CHS, the most typical type III PKS, was solved in 1999 (**Figure 2(c)**).[16] CHS, having an $\alpha\beta\alpha\beta\alpha$ fold, belongs to the thiolase fold family.[4,16] The crystallographic study of type III PKSs as well as their complex with substrates revealed the amino acid residues important for the catalytic diversity of type III PKSs. Alignment of the amino acid residues forming the active site and locating nearby of type III PKSs, including CHS, 2-PS, STS, and PCS,[11,16–18] in **Figure 4**, predicted the residues that are essential and important for polyketide synthesis. For example, Phe215 and Phe265 are the amino acid residues, called gatekeepers, that mediate the CoA binding tunnel.[4] Thr132, Glu192, and Ser338, forming an aldol switch, are involved in forming a hydrogen bond network stabilizing a water molecule, which is responsible for the cryptic TE activity of STS.[17] This TE activity is important for aldol condensation. The reactions catalyzed by type III PKSs are thought to be regulated by a subtle difference in the active site formed by these amino acid residues[16] (see Chapter 1.06).

1.05.3 Type III PKSs from Actinobacteria

1.05.3.1 RppA (1,3,6,8-Tetrahydroxynaphthalene Synthase)

Actinobacteria are high G + C bacteria known to produce a wide variety of secondary metabolites with various useful bioactivities. The polyketides showing interesting bioactivities from actinobacteria had been believed to be synthesized by type I and type II PKSs until RppA (red pigment production), the first bacterial type III PKS, was discovered, cloned, and characterized from *S. griseus* in 1995.[19] *In vitro* experiments revealed that RppA catalyzes the synthesis of 1,3,6,8-tetrahydroxynaphthalene (THN) from five molecules of malonyl-CoA (**Figure 5(a)**).[9] *rppA* from *S. griseus* constitutes an operon with a P-450mel gene, which is a member of the cytochrome P-450 family (**Figure 5(b)**).[20] P-450mel catalyzes oxidative biaryl coupling of two molecules of THN to synthesize 1,4,6,7,9,12-hexahydroxyperylene-3,10-quinone (HPQ). HPQ readily autopolymerizes to generate HPQ melanin (**Figure 5(a)**). Because the THN produced by the action of RppA is used as a substrate by P-450mel, an *rppA* disruptant loses the ability to produce a red-brown melanin pigment.[9] The RppA homologues that catalyze THN synthesis were also present in other actinobacteria, *Saccharopolyspora erythraea*,[21] *S. avermitilis*,[22] *S. antibioticus* IFO13271,[22] and *S. coelicolor* A3(2).[23] These RppA homologues constitute an operon with a P-450mel-like gene or a quinone-forming monooxygenase gene, *momA* (**Figure 5(b)**). MomA catalyzes oxidation of THN to yield flaviolin (**Figure 5(a)**).[24] In addition, hybrid compounds composed of a polyketide and a terpenoid containing a naphtoquinone ring, such as naphterpin,[25] furaquinocins,[26] and napyradiomycin,[27] are predicted to be biosynthesized via the THN formed by RppA homologues (**Figure 5(c)**). These *rppA* homologues sit near various post polyketide modification enzyme genes, including *nphB* whose gene product adds a geranyl moiety to the THN scaffold to synthesize the intermediates of naphterpin.[25,28]

RppA is different in the reaction from type III PKSs previously discovered from plants in two ways.[29] RppA utilizes malonyl-CoA as the starter substrate, which is usually incorporated as an extender substrate by most type III PKSs. The RppA reaction involves two carbon–carbon intramolecular cyclizations: one, aldol condensation, and the other, Claisen condensation (**Figure 5(a)**). Funa *et al.*[22] suggested that Tyr224 (256 in CHS numbering) is important for the malonyl-CoA priming by a bioinformatic approach and site-directed mutagenesis. Tyr224 mutants lost the ability to synthesize THN, although they retained the ability to synthesize pyrones from hexanoyl-CoA. Interestingly, this tyrosine is conserved in other bacterial type III PKSs, such as PhlD and DpgA (these enzymes are described later), that involve malonyl-CoA priming (**Figure 4**). They also showed that substitution of Ala305 (338 in CHS numbering) by bulky amino acid residues restricts the number of malonyl-CoA condensation, suggesting that the volume of the active site is important for the malonyl-CoA condensation. The relationship between the active site cavity and the number of condensation reaction is also examined with plant type III PKSs.[18] Substitution of the amino acids at positions 197, 256, and 338 in CHS


Catalytic triad
Functional diversity
CoA binding
Related to malonyl-CoA priming
Related to aliphatic tunnel
55 58 62 68 98 132 137 164 194 197
MsCHS ELKEKFQRMCDKSMIKRR----------YMYLT ARQDMVVVEVP CTTSGVDMPGA MMYQQGCFAGGTVL VCSEVTAVTFR
AhSTS DLKKKFQRICERTQIKNR----------HMYLT AREDMMIREVP CTTSGVALPGV MMYHQGCFAGGTVL VCSENTSVTFR
RpBAS NLKQKFKRLCENSRIEKR----------YLHVT VRHKMQVKGVA CCLAGVDMPGA MFYHLGCYAGGTVL VCSEMTTTCFR
Gh2-PS DLKEKFKRICEKTAIKKR----------YLALT ARQDLVVTGVP CTTAGVDMPGA MLYQQGCAAGGTVL VCSEITAILFH
AaPCS ELKKKFDHICKKTMIGKR----------YFNYD DRQDICVPGVP CTSCGVDMPSA CIYMQGCYAGGTVM VCAELTIMMLR
AaOKS ELKKKFDRICKKTMIGKR----------YFNYD DRQDICVPGVP CTSCGVDMPSA CVYMQGCYAGGTVM VCAELTIIGLR
OsCUS ALKDKFKRICQEMGVQRR----------YLHHT ARLDIAADAVP TTNSGAHVPGV MLHLNGCFAGCAAL VAAELTLMYFT
AvArsB VKEKFDKYALSSAQIKRRQLIFFPKVSDIHFANGHVDI -PRGSDLKVRH VTCSGYLSPSP HSYHMGCYGAFPAI VHTELLSAHNN
AvArsC VKEKFDKYAVSPAQIKQRQLVYFPKLTDIRFGDGNFDI -PRGADLKTRH VTCSGYLAPSP HSYNMGCYGAFPAI AHTELMSAHNN
Pf-5PhlD RM-ALAKRMIANTEVNER----------HL ---IVYEREAR TSCTGFMMPSL PIAQLGCVAGAAAI VSLEFSSLCYQ
Q287PhlD RM-ALAKRMIQNTQVNER----------YL ---IVYEREAR TSCTGFMMPSL PIAQLGCVAGAAAI VSLEFSSLCYQ
ScRppA QL-PLALRLIENTGVRTR----------HI ---KVYEREAK VSCTGFMMPSL PIAQLGCAAGGAAI VACEFCSLCYQ
SoceCHS1 QL-KLVLRLINGTGVQKR----------HI ---RIYEREAK VSCTGFMMPSM PMAQLGCAAGAAAI VACEFCSLCYQ
SgRppA QR-DLVLRLIQNTGVQTR----------HL ---QVYEAEAK VSCTGFMMPSL PIAQLGCAAGGAAI VSCEFCSLCYQ
ScSrsA ADRALLHRVHASAGVRTR----------HL ---GAWLETGL ASITGVSAPSL PLFGLGCVAGAAGL LTVELCSLTLQ
SgSrsA ADRRVLDRLHENARVRAR----------HT ---DVFIRSAV TSVTGIAAPSV PVFGLGCVAGAAGT LSVELCSLTFQ
Mtpks11 YE-DIVRQLHASAKVNSR----------HL ---KIFIEKAV ATVTGLAVPSL PLFGLGCVAGAAGV VSVELCSLTYP
Mtpks10 HE-EIIRRLHAAAKVNGR----------HL ---EIFIEKAV ATVTGVAVPSL PLFGLGCVAGAAGV VSVELCSLTYP
MxCHS FNLERLEDLHRAVQVGGR----------HL ---DAWIRSAT VTVTGIATPSI PLFGLGCVAGAAGV IAVELCSLTLQ
AoDpgA IEDPKIRSVFLNSAIDRR----------FL -LLDKHKKLAV VTSTGFLTPGL DIVGMGCNAGLNAL LCSEACSAAY-
AbDpgA VEDPKIRSVFLNSAIDRR----------FL -LLDKHKELAV VTSTGFLTPGL DIVGMGCNAGLNAL LCSEACSAAY-
AocsyA PAVRKTLYVNEKSRIQTR----------R ECDVLFRKYGV VTCTNTANPGL LLHGVGCAGGAAAL VACEICMIFFR
AocsyC VALLKIIEINERTKIENR----------Y ECDDLFKKYGI MTCTSASSPGF LLSGVTCAGSVAAL IAAETMTVYVR
AocsyB PSLEKMLEINRKTRIDHR----------Y ECDSLFKEYGI VTCTNTAHPGF LLHGIGCGGGISAM VACEVPTVFAR
AocsyD PAVSETLDRIERSGIKQL----------C DADALFKEHAI VTSTNTANPGF LLYGLGCAELHRIN ----------
NcORAS PAMKKVLAINRYTGIDQR----------S ELHEVFMSDGV TTCTDSANPGY LLHGIGCSGGLAAL LALEVSTTMVR
DdstlA ETNEKVKRIFEQSQIKTR----------HLV DVNNQFKKVVP VTSTGIIIPDV SLNLMGCLAGLSSL VCTEVCSLHFS
ScGcs PGGGGDLGAAREALVRDG----------FTEQ ERTAPAWEAVQ SNSTTPALPGL PATQWACVAGTRSL VISEALSTTYQ
Mtpks18 GQRERIPRVYQKSRITTR----------RMA DRMHLFYEHAV ATSTGFIAPGV VVNFMGCAAAMNAL VCIELCSVNAV
SoceCHS2 RSRVLFGRMADKAQISHR---------FSVLAPSAEVG DRMRLYEKHAP TSCTGFSAPGI MIGFMGCYAAINAL VNVELCTLHLQ
Ddstlb PHKEFIDNIYKKSKIRSR----------Y ERVALFREQTY VTSTGIMAPSF MINFMGCGAAVNSM VAVEASATCMK
206 211 215 254 256 293 294 303 308 336 338 375
MsCHS HLDSLVG---QALFGDGAA SEGAIDGHLR GIS----DYN---SIFWIAHPGGPAILDQ EYGNMSSACVLF EWGVLFGFGPGLTIET
AhSTS DMDSLVG---QALFADGAA SHGAIGGLLR GIS----DYN---SIFWIAHPGGRAILDQ NYGNMSSACVFF DWGVLFGFGPGLTIET
RpBAS HLDSMIG---QAILGDGAA SHGAIEGHLL NIS----DWN---SLFWIAHPGGPAILDQ DYGNMSSATVFF EWGVLFGFGPGITVET
Gh2-PS HLDSLVA---QALFGDGAA TEKAMKLHLR GIT----DWN---SVFWMVHPGGRAILDQ EYGNLISACVLF DCGVLFGFGPGMTVET
AaPCS HLDNAIG---ISLFGDGAA TEDVIHLHLR GITPPE-DWN---SLFWIPHPGGRAILDQ DYGNMVSASVGY EWGVLLGFGPGITVET
AaOKS HLDNAIG---NSLFGDGAA SEDVIHLHMR GMTPPE-DWN---SLFWIPHPGGRAILDQ DCGNMVSACVLY EWGVLLGFGPGMTVET
OsCUS CFRTLLV---QGLFGDGAA SDHALNMRFT LGGDGGGGWN---DLFWAVHPGSSTIMDQ DYGNMSGATVIF ELGVMMAFGPGMTVDA
AvArsB RAENIIT---MTLFADGLI SADEMTWVPG GIDYERERL----ELTFAIHPGGPKIVEH ENGNMSSATIPH TRVLCLGFGPGLTVTG
AvArsC RVDNIIS---ATLFSDGLI SADTMTGVPG GMDFERDKD----ALSFAIHPGGPKIVDH ENGNMSSSTIPH TRIACLGFGPGLTAAG
Pf-5PhlD KLHAFIS---AALFGDAVS SEHYIKYDVK EQNCAHND-------FFIFHTGGRKILDE EAGNIASVVVFD DIGLLAAFGPGFTAEM
Q287PhlD KLHAFIS---AALFGDAVS SEHYIKYDVK NQHCAQND-------FFIFHTGGRKILDE EAGNIASVVVFD ATGMLAAFGPGFTAEM
ScRppA GVGSLLC---NGLFGDGIA TEDWIMYDVK GWDASDLD-------FYIVHAGGPRILDD EYGNIASAVVLD ARGLLAGFGPGITAEM
SoceCHS1 DVGSLLS---DGLFGDAVG TEEWISYAVK GWDVGNLD-------FYIVHAGGPRILND EYGNIASAVVLD ARGVIAGFGPGITAEM
SgRppA GVGSLLS---NGLFGDALS TEDWISYAVR GWSVPNMD-------FFIVHAGGPRILDD ERGNIASSVVFD AQGLIAGFGPGITAEV
ScSrsA SRANLVA---GALFGDGAA TEGLLGWEIG ELTVDDIG-------TWICHPGGPRILSA AVGNMSSVSVLH TWGLLLAMGPGFCSEL
SgSrsA SPANLVA---TALFGDGAA TEHVMGWDVG GLKPKDVA-------HWVCHPGGPKVLEA DVGNLSSSSVLH TPGLLLAMGPGFCCEL
Mtpks11 TLPGLVG---SALFADGAA SLRTMGYDVG GLSTTDVG-------AWVTHPGGPKIINA EIGNLSSASVLH SPGLMIAMGPGFCSEL
Mtpks10 TVSSLVG---TALFGDGAA SLHIMGWDVG RLTKDDIG-------AWVSHPGGPKVIDA EIGNLSSASILH SAGLMLAMGPGFCTEL
MxCHS SIPNIIA---SGLFGDGAA TERIMGWDVV GLKRGDIR-------HWVAHTGGPKVLEA EVGNLSSASVLF EWGVMMAMGPGFGAEM
AoDpgA TMRTAVV---NSLFGDGSA AVDAMRYDWD GLRRSDIG-------HWLVHSGGKKVVDA DYGNLSSGSFLF DYGVLMTMGPGSTIEM
AbDpgA TMRTAVV---NSLFGDGSA ALDAMRYDWD GLRRSDIG-------HWLVHSGGKKVIDS DYGNLSSGSFLF DYGVLMTMGPGSTIEM
AocsyA IVKAQEANVAMTLFGDGAG SASSIQFNIR PSLYSDKSNFDPSSYDWALHPGGYSIAVL SRGNTSSSTVIS DKVIVAAFGPGITMEL
AocsyC IVKESIPNIGPTLFGDGAC PG-CVGIRWM PSLRLEESNFNPSTYDWALHPGGAAILTG SAGNTSSATVLS SKVIGAAFGADITVEM
AocsyB MDKTQDVNVAMCLFGDCAA TEDIAHFNVH PSLALEEKNYVPSNYGWAVHPGGYAVLVA DGGNTISTTIIR DKLVLAAIGHGITLET
AocsyD EVGFGMS-----RFSDGAS SEGYSELSIG PGLRSDNTRDLPNSYNWASDFPPGRSALKF SRGSTISCSVLN NKVIALAIGPGVILEM
NcORAS IDALQETRIGIALFSDCAS SEHDLGFDVD QDQLPSSYQKP-ADFDWAMHPGGATILSG NHGNSSSATIFS EYVVGCAFGPGINVEM
DdstlA GDQMVAS----SIFADGSA TENAMVWDLE KLQTSTAISAK--DCEFLIHTGGKSILMN AYGNMSSASVIF TYSISLAFGPGLAFEG
ScGcs LESLIVR----LLFADTAV TRDLHRLETR EDDPSSWHPD-----VLLAHPGGTRVLEY TSGNRGGAAVFD SRAVLYAAAPGLTATA
Mtpks18 NDVVIHS-----LFGDGCA TEDGIVLGVN GLQISDID--------LWAIHPGGPKIIEQ RFGNMLSVSLIF STGVAFAFGPGVTVEG
SoceCHS2 AMDQMLS---FLLFGDGCA TRELITWNIR APGEID---------LWAVHPGGRSVLDA RFGNMSSATIMF MRGCAMSFGPGLTAET
Ddstlb RSDLLSQ----AIFTDGCV SRDALNLFIG SLQKSDIE-------FFATHPGGAKIISA RYGNMIGVSTYY NYGMAMAFSPGASIEA

numbering with smaller amino acids resulted in an increase in the number of malonyl-CoA condensation. The importance of Tyr224 was also suggested by site-directed mutagenesis of RppA from *S. coelicolor* A3(2).[30] The residues Tyr224 and Ala305 are conserved in all RppA homologues (**Figure 4**).

The crystal structure of an RppA homologue, 1,3,6,8-tetrahydroxynaphthalene synthase (THNS), from *S. coelicolor* A3(2), was solved by Austin *et al.*[29] The overall structure of THNS exhibits an $\alpha\beta\alpha\beta\alpha$-fold core domain and a homodimeric interface, as observed for plant type III PKSs. The crystal structure of THNS contained a polyethylene glycol (PEG) heptamer and glycerol molecules in a tunnel, called a PEG-binding tunnel, which was not observed in plant PKSs. Although the PEG-binding tunnel seems to be not important for the synthesis of a THN scaffold, it perhaps accommodates an alkyl chain of long-chain fatty acids. A similar tunnel was observed in the crystal structure of PKS18, a type III PKS from *M. tuberculosis*,[31] the detail of which will be described later. The tunnel of this enzyme was found to accommodate myristic acid. Tyr224, one of the residues near the active site of THNS, probably contributes to a steric constriction along the horizontal direction of the active site cavity and limits the conformational flexibility of the malonyl moiety attached to Cys135, one of the residues forming the catalytic triad of RppA. C106S mutants also lost the ability to synthesize THN and yielded only a TAL, indicating that the thiol at Cys106 is also important for THN formation. In this TAL formation reaction, the acetyl moiety of TAL is provided by decarboxylation of the malonyl moiety probably catalyzed by THNS. Whether Cys106 contributes to the inhibition of decaboxylation or lactonization is not clear. This residue is conserved in other RppA homologues (132 in CHS numbering) (**Figure 4**). A feeding experiment with [1,2-^{13}C]acetic acid, which is a precursor of malonyl-CoA, in *Escherichia coli* expressing THNS revealed that RppA catalyzes U-shaped cyclization but not S-shaped cyclization (**Figure 5(d)**).[29] The THN synthesized via the S-shaped cyclization harbors a central carbon–carbon bond derived from the polyketide chain itself, but the THN synthesized via the U-shaped cyclization harbors a central carbon–carbon bond formed by aldol or Claisen condensation of the resulting polyketide chain. Although the crystal structure of THNS, a homologue of RppA, provided an insight into the reaction mechanisms, the details of the THN cyclization pathway are still unclear.

RppA also accepts acyl-CoAs as a starter substrate *in vitro* when an excess amount of acyl-CoAs is present.[32] RppA accepts aliphatic acyl-CoAs with carbon lengths from C_4 to C_8 as starter substrates and synthesizes pyrones and phloroglucinols. When acetoacetyl-CoA is accepted, RppA synthesizes TAL. Interestingly, RppA produces a hexaketide from octanoyl-CoA. In addition, incubation of RppA with methylmalonyl-CoA and acetoacetyl-CoA results in the production of methylated TALs. These observations suggest that bacterial type III PKSs possess promiscuous substrate specificity as observed for plant type III PKSs and could be used for the production of unnatural polyketides by means of metabolic engineering.[33,34]

Figure 4 Alignment of amino acid sequences of type III PKSs from plants and microorganisms. This alignment shows only amino acid sequences related to the catalytic diversity. MsCHS = chalcone synthase from *Medicago sativa* (DNA database accession number AAA02824); AhSTS = stilbene synthase from *Arachis hypogaea* (BAA78617); RpBAS = benzalacetone synthase from *Rheum palmatum* (AAK82824); Gh2-PS = 2-pyrone synthase from *Gerbera hybrida* (CAA86219); AaPCS = pentaketide chromone synthase from *Aloe arborescens* (AAX35541); AaOKS = octaketide synthase from *Aloe arborescens* (AAT48709); OsCUS = curcuminoid synthase from *Oryza sativa* (AK109558); AvArsB = ArsB from *Azotobacter vinelandii* (EAM05488); AvArsC = ArsC from *Azotobacter vinelandii* (EAM05489); Pf-5PhlD = PhlD from *Pseudomonas fluorescens* Q2-87 (YP_263015); Q287PhlD = PhlD from *Pseudomonas fluorescens* Q2-87 (AAB48106); ScRppA = RppA from *Streptomyces coelicolor* A3 (2) (SCO1206); SoceCHS1 = type III PKS from *Sorangium cellulosum* 1 (CAN92292); SgRppA = RppA from *Streptomyces griseus* (BAE07216); ScSrsA = SrsA homologue from *Streptomyces coelicolor* A3(2) (SCO7671); SgSrsA = SrsA from *Streptomyces griseus* (SGR472); Mtpks11 = pks11 from *Mycobacterium tuberculosis* (Rv1660); Mtpks11 = pks11 from *Mycobacterium tuberculosis* (Rv1665); MxCHS = type III PKS from *Myxococcus xanthus* (ABF87000); AoDpgA = DpgA from *Amycolatopsis orientalis* (CAA11765); AbDpgA = DpgA from *Amycolatopsis mediterranei* DSM5908 (CAC48378); AocsyA = type III PKS from *Aspergillus oryzae* (BAD97390); AocsyB = type III PKS from *Aspergillus oryzae* (BAD97391); AocsyC = type III PKS from *Aspergillus oryzae* (BAD97392); AocsyD = type III PKS from *Aspergillus oryzae* (BAD97393); NcORAS = ORAS from *Neurospora crassa* (NCU04801); DdstlA = type III PKS domain of StlA from *Dictyostelium discoideum* (EAL72032); ScGcs = Gcs from *Streptomyces griseus* (CAC01488); Mtpks18 = pks18 from *Mycobacterium tuberculosis* (A70958); SoceCHS2 = type III PKS from *Sorangium cellulosum* 2 (CAN92341); Ddstlb = type III PKS domain of StlB from *Dictyostelium discoideum* (EAL62021).

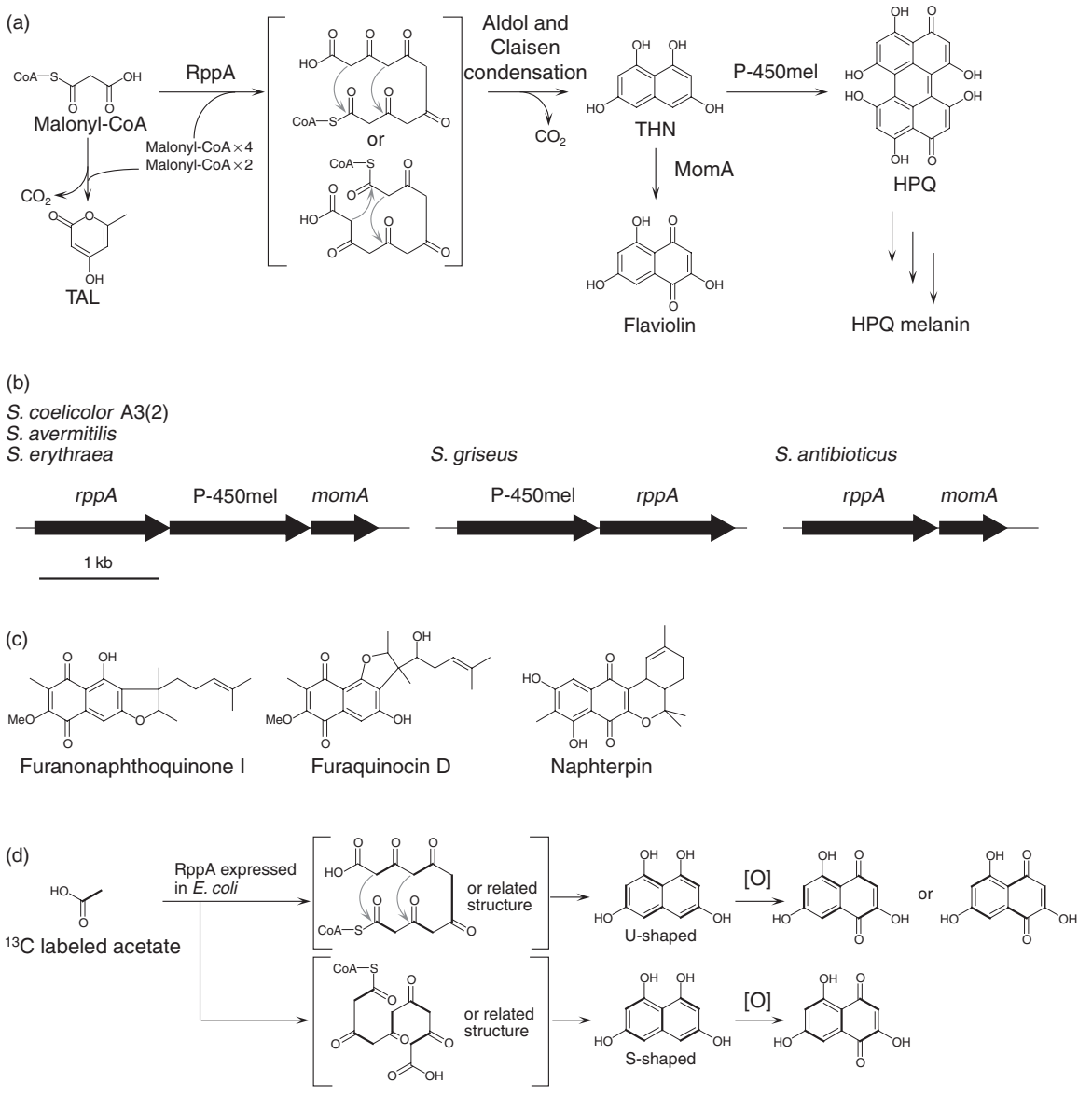


Figure 5 The reaction catalyzed by RppA and subsequent modification of the product, THN, into flaviolin by MomA and into HPQ melanin by P-450mel (a), the gene organizations near *rppA* in various *Streptomyces* strains (b), prenylated polyketides synthesized via the THN synthesized by the action of RppA (c), and the folding of the polyketide chain, as determined by [1,2-^{13}C]acetate feeding (d). RppA catalyzes condensation of five malonyl-CoA and subsequent aldol and Claisen condensations lead to the formation of THN (a).

1.05.3.2 Gcs

Streptomyces coelicolor A3(2) is the first *Streptomyces* species whose genome sequence was uncovered in 2002.[35] In addition to RppA (THNS), there are two type III PKSs, Sco7221 and Sco7671, in this species. Sco7221 was characterized as a germicidin synthase (Gcs) by Song *et al.*[36] Germicidin derivatives were first isolated from *Streptomyces viridochromogenes* NRRL B-1551 and have an inhibitory effect on the germination of spores.[37] Comparison of the metabolic profiles of wild-type *S. coelicolor* M145 and a Sco7221-disrupted mutant predicted that Sco7221 was involved in the biosynthesis of germicidin derivatives, germicidin A, isogermicidin A, germicidin B[7] isogermicidin B, and germicidin C (**Figure 6**). These compounds were probably produced

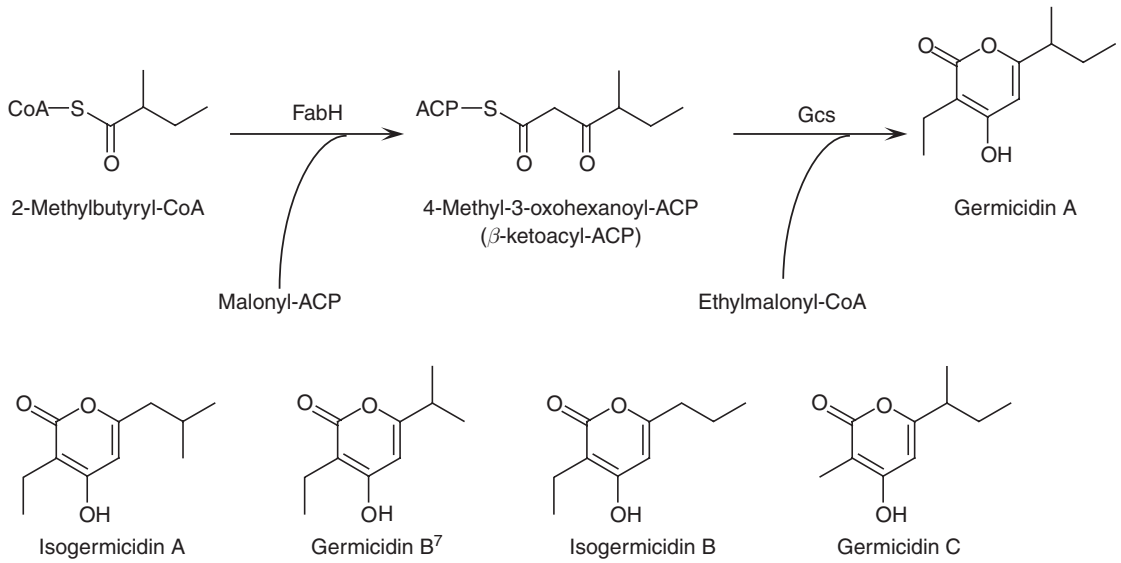


Figure 6 The predicted biosynthesis pathway of germicidin A and the structures of its derivatives. The initial reaction of germicidin synthesis is catalyzed by FabH, synthesizing β-ketoacyl-ACP by condensation of acyl-CoA and malonyl-ACP. Gcs incorporates the β-ketoacyl-ACP and condenses with ethylmalonyl-CoA to synthesize germicidin A.

from 2-methylbutyryl-CoA, isovaleryl-CoA, isobutyryl-CoA, or butyryl-CoA condensed with malonyl-CoA, followed by condensation with ethylmalonyl-CoA or methylmalonyl-CoA. In this novel polyketide synthesis pathway (**Figure 6**), the initial condensation is catalyzed by FabH, a β-ketoacyl carrier protein synthase III involved in the fatty acid biosynthesis. FabH synthesizes 4-methyl-3-oxohexanoyl-ACP from 2-methylbutyryl-CoA and malonyl-ACP. The resultant 4-methyl-3-oxohexanoyl-ACP would be incorporated in Gcs and condensed with ethylmalonyl-CoA or methylmalonyl-CoA to form germicidin A. This prediction was supported by an experiment using *S. coelicolor* M511, in which the *fabH* gene was replaced by the orthologue from *E. coli*. This strain produced predominantly straight-chain fatty acids, whereas wild-type *S. coelicolor* A3(2) produced predominantly branched-chain fatty acids. Reflecting the fatty acids produced, the *S. coelicolor* A3(2) strain carrying the *E. coli fabH* gene lost the ability to produce germicidin A, isogermicidin A, or germicidin C but still produced germicidin B^7 and isogermicidin B. FabH is therefore involved in germicidin biosynthesis and Gcs uses the β-ketoacyl-ACP, not acyl-CoA esters, as the starter substrate that is an intermediate of the fatty acid synthesis (**Figure 6**).[36] Plant type III PKSs so far reported all incorporate CoA esters as starter substrates. Therefore, incorporation of ACP thioesters by Gcs is surprising and the mechanism of ACP recognition by type III PKSs is interesting for future study. Arginine at position 308 (in CHS numbering) was revealed to be important for the interaction between ACP and FabH from *E. coli*.[38] Therefore, this residue also appears to be important for the interaction between Gcs and ACP. In addition, incorporation of ethylmalonyl-CoA as the extender substrate is also interesting because no such activity of type III PKSs has ever been observed.

Interaction between Gcs and ACP was demonstrated *in vitro* by Grüschow *et al.*[39] Four hexanoyl-ACPs, including MmcB responsible for mitomycin biosynthesis, TcmM for tetracenomycin biosynthesis, Otc for oxytetracycline biosynthesis, and functionally unknown SCOACP (Sco0549), were synthesized *in vitro* by using promiscuous substrate specificity of sfp, the PPT involved in surfactin biosynthesis.[40] The hexanoyl-ACPs thus synthesized were each incubated with malonyl-CoA and either Gcs or Sco7671. As a result, Gcs incorporated hexanoyl-ACPs as starter substrates and yielded triketide pyrones in amounts of approximately 25–50%, when compared to the hexanoyl-CoA-primed reaction. Sco7671 incorporated hexanoyl-ACPs with relatively low efficiency. Although hexanoyl-ACPs used in this study are not the native substrate of Gcs and Sco7671, these results support the hypothesis that Gcs and Sco7671 could interact with ACPs.

1.05.3.3 SrsA

Streptomyces griseus contains two type III PKSs.[41] In addition to RppA (SGR6620), SGR472 showing 61% identity in amino acid sequence with Sco7671 is present.[41] SGR472 was named SrsA (*Streptomyces* resorcinol synthesis).[42] SrsA is a member of the *srs* operon, which is composed of *srsA* encoding a type III PKS, *srsB* encoding a methyltransferase, and *srsC* encoding a flavoprotein hydroxylase (**Figure 7(b)**). The heterologous

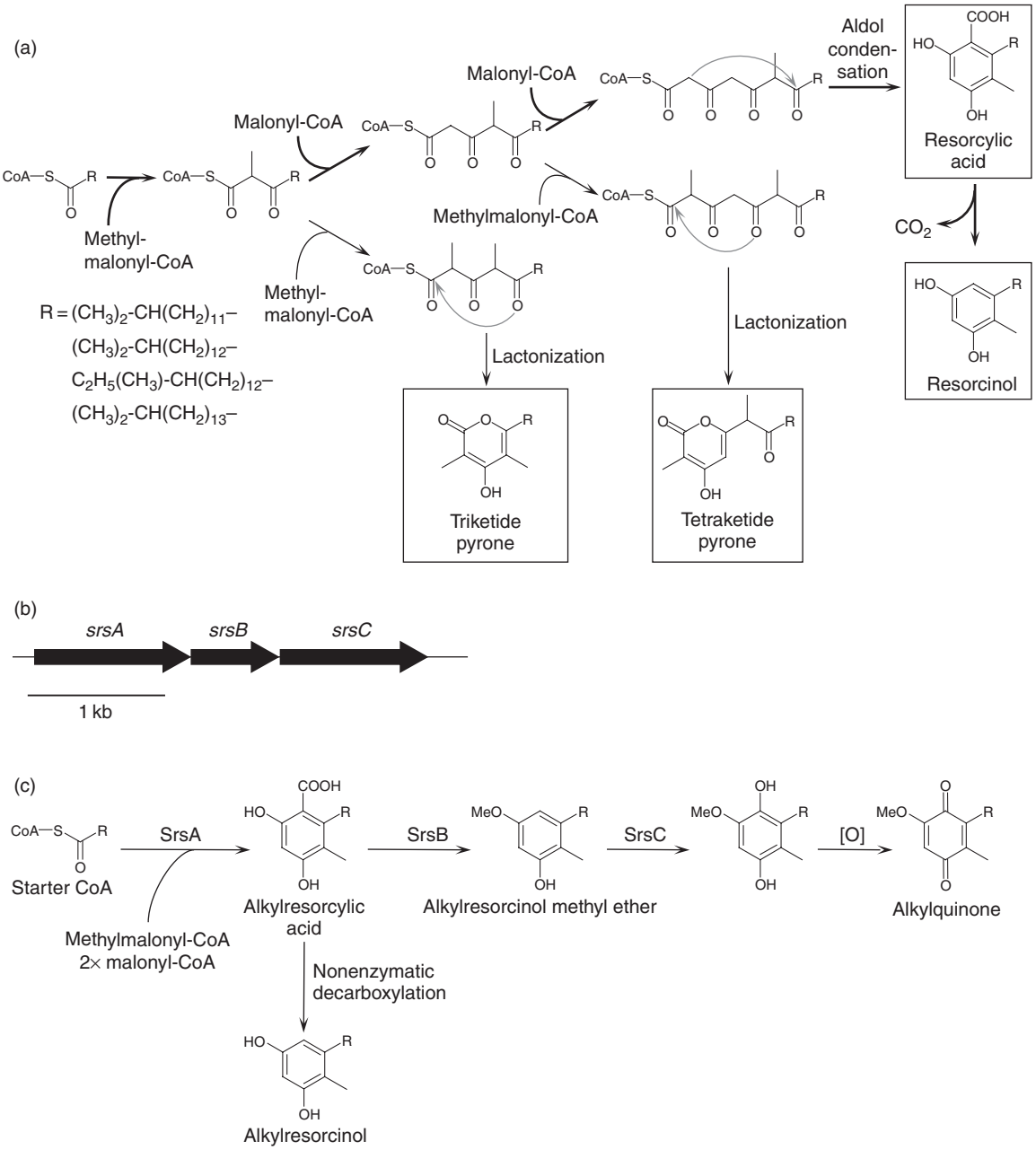


Figure 7 The reaction catalyzed by SrsA to form resorcylic acid and pyrones (a), the gene organization of the *srs* operon (b), and the reaction catalyzed by SrsABC to form hydroxylated alkylresorcinol methyl esters, which are then nonenzymatically oxidized to alkylquinones (c). SrsA catalyzes the formation of alkylresorcylic acids containing a *C*-methyl group by condensing methylmalonyl-CoA and two malonyl-CoAs. The alkylresorcylic acid formed is readily converted to alkylresorcinol nonenzymatically. SrsB catalyzes the methylation of the SrsA products and SrsC catalyzes oxidation of the SrsAB products. The products of SrsABC reactions are nonenzymatically converted to alkylquinones.

expression of SrsA in *S. lividans* resulted in the production of branched-chain alkylresorcinols as major products and tetraketide pyrones (**Figure 7(b)**). Since these alkylresorcinols possessed a C-methyl group on the aromatic ring, SrsA was predicted to incorporate methylmalonyl-CoA as an extender substrate. *In vitro* study of recombinant SrsA produced in *S. lividans* revealed that SrsA synthesizes alkylresorcinols from a branched-chain fatty acid CoA ester, one methylmalonyl-CoA, and two malonyl-CoAs (**Figure 7(a)**). The triketide and tetraketide pyrones produced by the action of SrsA also contained a *C*-methyl group. Of the type III PKSs reported to date, Gcs,[36] a CHS cloned from *Pinus strobus* seedlings,[43] and SrsA are examples that incorporate methylmalonyl-CoA as an extender. Interestingly, SrsA prefers to incorporate one methylmalonyl-CoA molecule; it produces pyrones, instead of resorcinols, when two or more methylmalonyl-CoA molecules are incorporated. The order of incorporation of methylmalonyl-CoA and malonyl-CoA is presumably regulated by a subtle difference in the active site, as observed for the catalysis of many other type III PKSs. The regulatory mechanism of methylmalonyl-CoA incorporation is interesting for future study.

Coexpression of either SrsB or SrsC and SrsB, with SrsA revealed that both enzymes modify the resorcinol scaffold (**Figure 7(c)**).[42] SrsB catalyzes methylation of the hydroxyl moiety on the phenol ring produced by the action of SrsA. SrsC is a hydroxylase responsible for regio-specific hydroxylation of the alkylresorcinol methyl ethers produced by the actions of SrsA and SrsB. Because of the instability of the resulting hydroxylated alkylresorcinols, they are nonenzymatically converted into alkylquinones.

Further *in vitro* analysis of SrsA and SrsB revealed that SrsA catalyzes the formation of alkylresorcylic acids and SrsB catalyzes decarboxylative methylation of alkylresorcylic acids (C. Nakano and S. Horinouchi, unpublished data). Alkylresorcylic acids could be nonenzymatically decarboxylated quickly. Therefore, alkylresorcylic acids could not be detected in the above described *in vivo* experiment; alkylresorcylic acids are readily decarboxylated during the incubation and extraction. By carefully optimizing the reaction and analysis conditions *in vitro*, we found that SrsA synthesizes alkylresorcylic acids as the major products and alkylresorcinols were synthesized via nonenzymatic decarboxylation of alkylresorcylic acids. The assay using stable isotope-labeled malonyl-CoA clearly showed that methylmalonyl-CoA is incorporated at the first condensation (**Figure 7(a)**) (our unpublished data). In addition, SrsB, which was predicted to catalyze methylation of alkylresorcinol, did not synthesize alkylresorcinol methyl ether from alkylresorcinol. However, SrsB synthesized an alkylresorcinol methyl ether when coincubated with SrsA, malonyl-CoA, methylmalonyl-CoA, and long-chain fatty acyl-CoA. Therefore, SrsB catalyzes the methylation of the precursor of alkylresorcinol, probably alkylresorcylic acids. As no alkylresorcylic acid methyl ether was detected in this assay, SrsB was supposed to catalyze methylation of alkylresorcylic acid coupled with decarboxylation. From all these findings, it is clear that Srs enzymes presumably catalyze the formation of alkylquinone by the reactions depicted in **Figure 7(c)**.

Disruption of the *srs* operon in *S. griseus* resulted in loss of alkylresorcinols and alkylquinones formation. At the same time, the *srs* disruptants exhibited higher sensitivity to penicillin G and cephalexin, both of which belong to the β-lactam antibiotics. The alkylresorcinols and alkylquinones are probably associated with or integrated in the membrane because of their amphiphilic properties. In fact, they were fractionated in the membrane/cell wall fraction. All these findings suggest that these phenolic lipids synthesized by the *srs* operon confer penicillin resistance by influencing the overall constitution of the cell wall and membranes. As *srs*-like operons exist in many Gram-negative and Gram-positive bacteria, such as *B. subtilis*, *Rhodobacter sphaeroides*, *Rhodospirillum centenum*, *Rhizobium etli*, and *Mycobacterium* spp.[42] In fact, the *srs*-like operon in *B. subtilis* directs the synthesis of alkylpyrones, although no detectable phenotypic changes of an *srs* mutant were observed.[44]

1.05.3.4 DpgA

Balhimycin is a vancomycin-type antibiotic isolated from *Amycolatopsis mediterranei*.[45] As the vancomycin group glycopeptides are used as powerful antibiotics for methicillin-resistant *Staphylococcus aureus* (MRSA), they are one of the most important groups of antibiotics. Balhimycin is a nonribosomal peptide synthesized by a nonribosomal peptide synthetase and requires several unusual amino acids, (*S*)-3,5-dihydroxyphenylglycine (DPG) and (*S*)-4-hydroxyphenylglycine, for its biosynthesis. DpgA was the firstly characterized as a key enzyme responsible for the biosynthesis of DPG (**Figure 8**). *dpgA* is a member of the *dpgABCD* gene cluster, which is involved in balhimycin biosynthesis in *A. mediterranei* DSM5908 (**Figure 8(b)**).[46] DpgA, sharing 26%

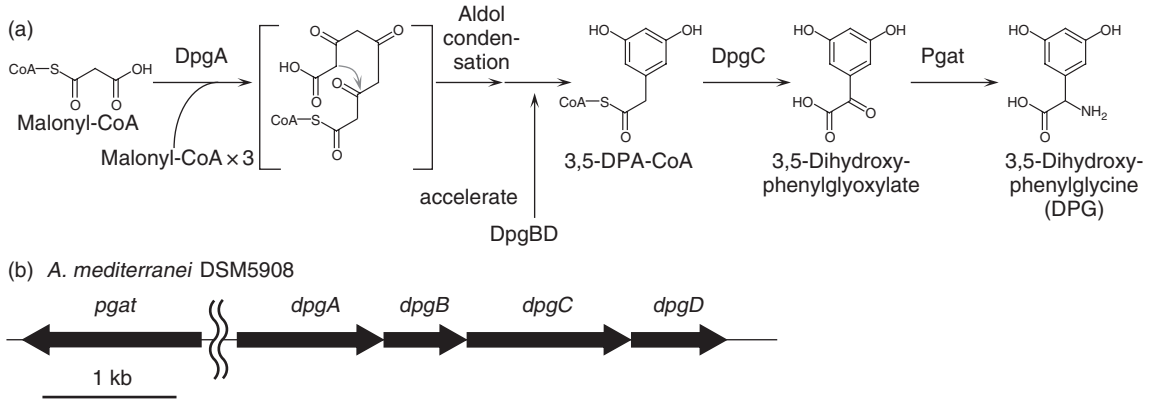


Figure 8 The reaction catalyzed by DpgABCD and Pgat (a) and the gene organization of the *dpg* operon (b). DpgA synthesizes DPA-CoA by catalyzing condensation of three malonyl-CoAs and aldol condensation-type cyclization. DpgBC accelerates the aromatization and dehydration, followed by aldol condensation catalyzed by DpgA. DpgC catalyzes the oxidative cleavage of the thioester bond of DPA-CoA and synthesizes 3,5-dihydroxyphenylglyoxylate. Pgat catalyzes the amidation of 3,5-dihydroxyphenylglyoxylate, yielding DPG.

identity in amino acid sequence with a type III PKS GHCHS2 from the plant *Gerbera hybrida*, was found to be involved in DPG biosynthesis and catalyze the formation of 3,5-dihydroxyphenylacetic acid or its precursor 3,5-dihydroxyphenylacetyl-CoA (DPA-CoA). *In vitro* study of DpgA encoded by the chloroeremomycin biosynthesis gene cluster in *A. orientalis*[47] showed that DpgA catalyzes the formation of DPA-CoA via condensation of four malonyl-CoAs and cyclization of aldol condensation type. Interestingly, DpgA releases the product as a CoA thioester and incorporates malonyl-CoA as the starter substrate, as does RppA. DpgB and DpgD, both belonging to the crotonase superfamily, accelerate the DPA-CoA formation catalyzed by DpgA in some unknown way.[48] Chen *et al.*[47] detected a weak enoyl-CoA hydratase activity of DpgBD and predicted that DpgBD accelerated the reaction by catalyzing the dehydration and aromatization followed by aldol condensation. DpgC, which is an oxygenase, converted DPA-CoA to 3,5-dihydroxyphenylglyoxylate, which was then converted to DPG by pgat (HpgT). Surprisingly, DpgA still retained DPA-CoA synthesis activity when the catalytic Cys160 was substituted by alanine or serine (C160A and C160S mutants), although DpgA lost the ability to bind malonyl-CoA by these substitutions.[48] Substitution of His296, forming the catalytic triad with Cys160 and Asn329, caused only partial loss of the activity. DpgA is the only type III PKS enzyme thus far reported whose activity is not completely eliminated by mutations of the catalytic residues. These findings suggest that the thioester bond between the starter substrate and a polyketide intermediate is not absolutely necessary for decarboxylative condensation of malonyl-CoA. As most type III PKSs show cryptic TE activity or cleave a thioester bond by Claisen condensation, DpgA is unique in that it releases the product as a CoA thioester.

The crystal structure of DpgC was reported recently.[49] This enzyme is an oxidase, catalyzing the formation of 3,5-dihydroxyphenylglyoxylate from DPA-CoA without requiring any cofactors or metal ions.[50] This oxidation couples with the release of a CoA thioester. During polyketide synthesis, the products are released by cleavage of a thioester bond of the final intermediate by a TE activity. Therefore, the oxidative, nonhydrolytic cleavage of the thioester bond catalyzed by DpgC is one of the interesting reactions in DPG biosynthesis. Although an oxidative cleavage of the thioester bond by type III PKSs is rare, a reductive cleavage of the thioester bond is reported for some nonribosomal peptide synthetases and type I PKSs.[51,52]

1.05.3.5 Type III PKSs from *Mycobacterium*

Mycobacterium tuberculosis is the causative agent of tuberculosis.[53] Fifty million people are now infected with drug-resistant strains of the tuberculosis pathogen, as estimated by the World Health Organization (WHO). The cell wall of *Mycobacterium* contains various unusual lipids, such as mycolic acid, methyl-branched fatty acids, phthiocerol, and phenolphthiocerol, and these unusual fatty acids are believed to provide resistance to

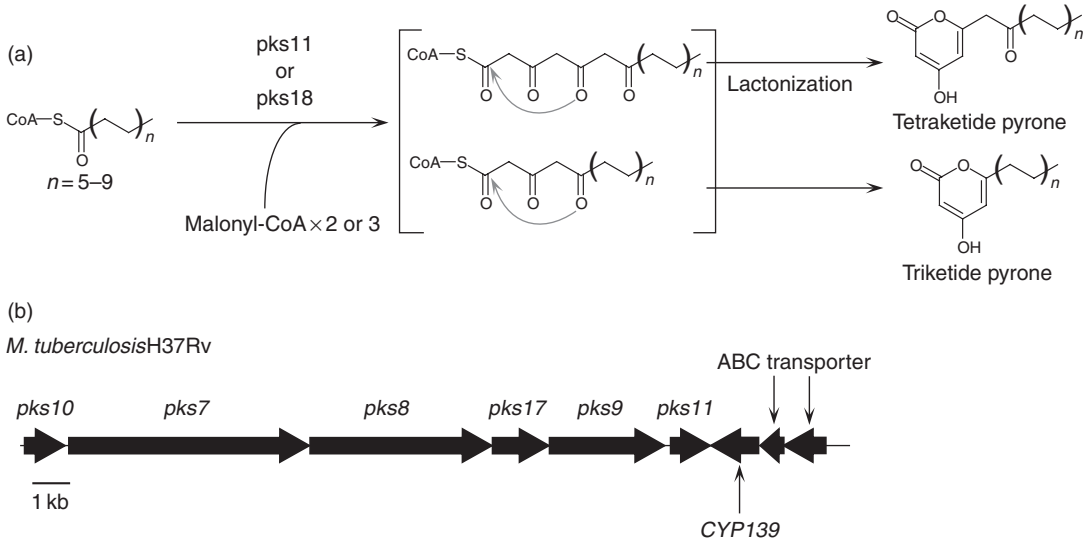


Figure 9 The reaction catalyzed by pks11 and pks18 (a) and the genes located near *pks10* and *pks11* (b). Both pks11 and pks18 incorporate long-chain fatty acid CoA esters and synthesize triketide and tetraketide pyrones via condensation of malonyl-CoA and lactonization (a). pks7, pks8, pks9, and pks17 are similar to modular type I PKSs.

antibacterial agents. Therefore, a better understanding of their biosynthesis pathways should provide novel drug targets. Because PKS-like enzymes are involved in the biosynthesis of some of these unusual lipids, studying PKS in *Mycobacterium* is important and useful in seeking novel drug targets.[54]

The genome sequences of *M. tuberculosis* strains predicted the presence of a large number of PKSs, including three putative type III PKSs, pks10, pks11, and pks18.[55] *pks10* and *pks11* are clustered in an unusual organization with other PKS genes (**Figure 9(b)**). In contrast, there are no PKS-related genes near *pks18*. Both pks11 and pks18 use fatty acid CoA esters as starter substrates *in vitro* (**Figure 9(a)**).[56] pks18 favors long-chain fatty acid CoA esters, such as lauroyl-CoA, palmitoyl-CoA, and arachidoyl-CoA, as starter substrates and produces triketide and tetraketide pyrones. pks11 possesses similar activity. Although both pks11 and pks18 synthesize alkylpyrones *in vitro*, no such compounds have been isolated *in vivo* from *M. tuberculosis*. Therefore, the *in vivo* functions of these enzymes are still unclear.

The crystal structure of pks18 was solved as a complex with myristic acid by Sankaranarayanan *et al.*[31] Myristic acid was buried in a tunnel, which had never been observed in plant type III PKSs. This tunnel was 20 Å long and extended from the active site to the surface. A similar tunnel was observed in THNS (RppA in *S. coelicolor* A3(2)).[29] This tunnel was predicted to accommodate a long alkyl chain of the starter substrate. As predicted, substitution of the amino acids near this tunnel by a bulkier amino acid, phenylalanine (C205F and A209F mutant; 194 and 197 respectively in CHS numbering), inhibited polyketide formation from long-chain fatty acids.

1.05.4 Type III PKSs from Proteobacteria

1.05.4.1 PhlD

PhlD was discovered by Bangera and Thomashow in 1999[10] in the biosynthesis gene cluster of 2,4-diacetylphloroglucinol (2,4-DAPG) in *P. fluorescens* Q2-87 (**Figure 10**). 2,4-DAPG is the major determinant of the ability of *P. fluorescens* to suppress *Gaeumannomyces graminis*, a fungal pathogen that causes the take-all disease of wheat.[57] The 2,4-DAPG biosynthesis gene cluster contains six genes *phlABCDEF*, four products (PhlABCD) of which appear to be involved in the biosynthesis of 2,4-DAPG. PhlD shows similarity in amino acid sequence to type III PKSs. PhlA shows 26% identity to FabH in *E. coli* but lacks the active site cysteine. This feature is similar to CLF, which is the KS lacking the active site cysteine in type II PKSs. PhlC shows similarity to the

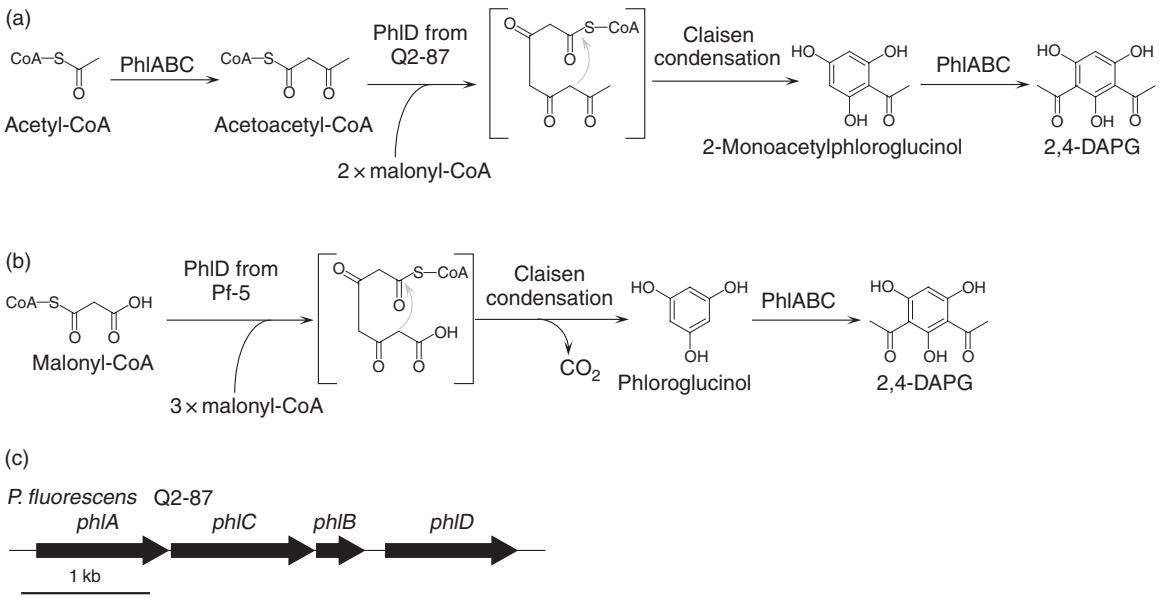


Figure 10 The 2,4-DAPG biosynthesis pathway predicted by Bangera and Thomashow[10] (a) and by Achkar *et al*.[58] (b), and the organization of the *phlABCD* operon (c). PhlD was predicted to synthesize 2-monoacetylphloroglucinol from acetoacetyl-CoA and malonyl-CoA (a) and phloroglucinol from three malonyl-CoAs (b).

thiolase domain of rat sterol carrier protein. PhlB shows no similarity to known proteins. Overexpression of *phlABCD* in *E. coli* resulted in the production of 2-monoacetylphloroglucinol and 2,4-DAPG. Gene disruption and feeding experiments established that PhlD catalyzes the formation of 2-monoacetylphloroglucinol by condensation of two malonyl-CoAs and one acetoacetyl-CoA, followed by Claisen condensation (**Figure 10(a)**).[10] Furthermore, PhlABC catalyze acetoacetyl-CoA synthesis and the conversion of monoacetylphloroglucinol to 2,4-DAPG (**Figure 10(a)**).[10]

Achkar *et al*.[58] characterized PhlD from *P. fluorescens* Pf-5. *In vitro* study of recombinant PhlD and overexpression of PhlD in *E. coli* clearly showed that this enzyme is a phloroglucinol synthase and catalyzes condensation of three malonyl-CoAs and Claisen condensation (**Figure 10(a)**). The substrate specificity of PhlD from *P. fluorescens* Pf-5 was examined by Zha *et al*.[59] It showed rather wide substrate specificity. However, it did not use acetoacetyl-CoA as a favorable starter substrate and produced mainly phloroglucinol from malonyl-CoAs and a trace amount of TAL from acetoacetyl-CoA and malonyl-CoA. Monoacetylphloroglucinol was not produced in this reaction. PhlD incorporated various aliphatic fatty acid CoA esters and phenylacetyl-CoA, similar to RppA, when these CoA esters were present in excess. Most starter substrates were converted to triketide and tetraketide pyrones. Interestingly, PhlD produced a heptaketide from octanoyl-CoA. The incorporation of a long-chain fatty acid CoA ester decreased in M21I, H24V, L59M mutants of PhlD, while the productivity of phloroglucinol was not affected.[59] These amino acid residues therefore seem to slightly reduce the volume of the tunnel that accommodates the acyl moiety.

There are two hypotheses for the route of 2,4-DAPG biosynthesis. Experimental data so far obtained suggest that the pathway proposed by Achkar *et al*.[58] is more likely. As PhlD from *P. fluorescens* Pf-5 turned out to be a phloroglucinol synthase, PhlD from *P. fluorescens* Q2-87, showing end-to-end similarity to that in strain Pf-5 (**Figure 4**), conceivably catalyzes the formation of phloroglucinol from malonyl-CoAs. Furthermore, the production of a red pigment, perhaps a phloroglucinol derivative, by *E. coli* carrying *phlD* from *P. fluorescens* Q2-87[10] supports the idea that PhlD from *P. fluorescens* strains Pf-5 and Q2-87 is a phloroglucinol synthase catalyzing the formation of phloroglucinol from three malonyl-CoA molecules. Both PhlDs conserve Tyr224 and Cys106 (256 and 132 in CHS numbering), which are important for malonyl-CoA-primed reaction.

1.05.4.2 ArsABCD

Azotobacter vinelandii is the Gram-negative, nitrogen-fixing soil bacterium. *Azotobacter vinelandii* forms a metabolically dormant cyst that resembles an endospore during its life cycle. The cyst is known to contain twice as much lipids as a vegetative cell does.[60,61] During the cyst formation, phospholipids in the membrane are replaced by 5-alkylresorcinols and other phenolic lipids.[61] This unique membrane matrix is presumed to contribute to the physiology and desiccation resistance of the cyst. There are tandem type III PKS genes in the genome of *A. vinelandii*. These genes, named *arsB* (*Azotobacter* resorcinol synthesis) and *arsC*, form an *ars* operon, with a set of type I FASs, *arsA* and *arsD* (**Figure 11(b)**).[62] Disruption of this operon eliminated the ability to produce the phenolic lipids, which proved that this operon was indeed responsible for phenolic lipids synthesis. In addition, the *ars* mutants formed only impaired cysts, suggesting that the phenolic lipids produced by the action of Ars enzymes are crucial for cyst formation. *Rhodospirillum centenum* is another bacterium that forms a resting cell cyst and possesses a type III PKS homologue.[63] This type III PKS homologue constitutes an operon with a methyltransferase homologue similar to SrsB. Because the expression of this type III PKS increases upon induction of the cyst cell development, alkylresorcinols produced by the action of this type III PKS are presumably necessary for cyst membrane formation even in this microorganism.

In vitro experiments with ArsB and ArsC[62] revealed that ArsB incorporates long-chain fatty acid CoA esters, such as arachidoyl-CoA and behenyl-CoA, and synthesizes alkylresorcinols via aldol condensation (**Figure 11(a)**). Similarly, ArsC incorporates long-chain fatty acid CoA esters and synthesizes triketide and tetraketide pyrones via lactonization (**Figure 11(a)**). Although ArsB and ArsC share high similarity, their cyclization specificity is quite different. Therefore, these enzymes are useful materials for the study of cyclization mechanism. In addition, the N-termini of ArsB and ArsC are quite different from other type III PKSs (**Figure 4**), suggesting some role of this region in the interaction with ArsA (see below).

ArsA and ArsD are both type I FASs having a domain architecture different from fungal and mammalian type I FASs.[64] ArsA harbors ER, KS, malonyl/acetyl transferase (MAT), ACP, and KR domains, and ArsD harbors DH and PPT domains (**Figure 11(c)**). The most striking feature of ArsAD is the lack of a TE domain, which catalyzes hydrolysis of the thioester bond between ACP and alkyl chains,[3,65] and a malonyl/palmitoyl transferase (MPT) domain, which catalyzes transfer of the acyl moiety to CoA.[66] These two domains are present in mammalian and fungal type I FASs to catalyze the termination of the fatty acid synthesis. Taken together, the fact that *arsAD* consists of the operon with *arsBC* suggested that the phenolic lipids (alkylresorcinols and pyrones) are synthesized by direct transfer of an intermediate alkyl chain from ArsAD (type I FASs) to ArsBC (type III PKSs).[64] ArsAD synthesized C_{22}–C_{26} fatty acids from malonyl-CoAs *in vitro* and the fatty acids synthesized were attached to ArsA, probably at the phosphopantetheinyl moiety in the ACP domain. The direct transfer of alkyl chains from ArsA to ArsBC was observed by *in vitro* radiolabeling study when ArsBC was added to the alkyl chain-attached ArsA enzyme. Further addition of malonyl-CoA yielded alkylresorcinols and alkyl triketide and tetraketide pyrones. These findings showed that the Ars enzymes synthesize alkyl resorcinols and pyrones by the reactions depicted in **Figure 11(a)**. This was the first demonstration of a type I FAS directly interacting with a type III PKS. There are several type III PKSs and type I FASs that are bioinformatically predicted to interact with each other. Steely 1 and 2 from *Dictyostelium discoideum* are examples to show a similar interaction between a type I FAS and a type III PKS (see below).[67]

1.05.4.3 Type III PKSs from Myxobacteria

Myxobacteria are the proteobacteria known to produce interesting and biologically active secondary metabolites as actinobacteria.[51,68] The secondary metabolites from Myxobacteria often target cellar structures that are rarely the targets of other compounds and because of this feature biosynthesis of secondary metabolites in this group of bacteria has been studied extensively. For example, epothilone isolated from *Sorangium cellulosum* interacts with the eukaryotic cytoskeleton and is about to be approved for breast cancer treatment.[51] Recently, the genome sequences of *Myxococcus xanthus* and *S. cellulosum* were determined.[69,70] *Myxococcus xanthus* possesses a single type III PKS homologue, mxCHS, and *S. cellulosum* possesses two type III PKS homologues, soceCHS1 and soceCHS2.[71] Heterologous expression of soceCHS1 in *Pseudomonas putida* revealed that soceCHS1

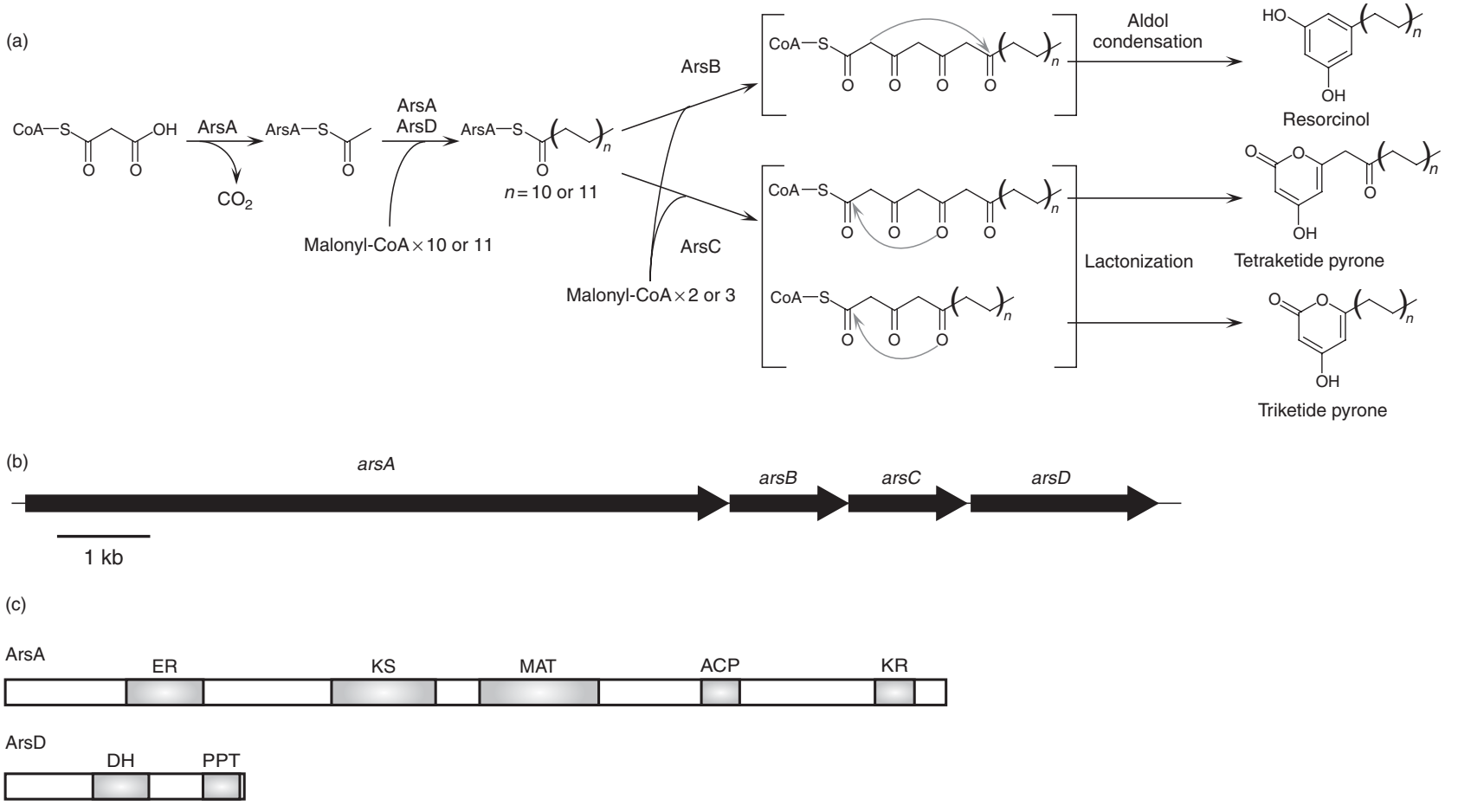


Figure 11 The reaction catalyzed by ArsABCD to produce resorcinols and pyrones (a), the organization of the *ars* operon (b), and the domain organizations of ArsA and ArsD (c). ArsAD synthesizes an acyl chain attached to ACP of ArsA. The direct transfer of this acyl moiety to type III PKSs, ArsBD, followed by malonyl-CoA condensation results in the production of alkylresorcinols and alkylpyrones.

catalyzes the formation of THN. The functions of the other two enzymes remain unsolved because disruption of the genes exhibited no difference in the metabolic profile, compared to that of the parental strain.

1.05.5 Type III PKSs from Eukaryotic Microbes

1.05.5.1 Type III PKSs from *Dictyostelium discoideum* (Steely 1, 2 (StlA, B))

Dictyostelium discoideum is a social amoeba and has been studied as a model organism to understand cellular motility, signaling, and interaction. Its genome sequence was determined.[72] *Dictyostelium discoideum*, which shows an interesting life cycle, usually grows by predation on soil bacteria.[73] Starvation induces cell aggregation and formation of a mound. The mound differentiates into a slug and moves to an optimal environment. Finally, the slug reaches an appropriate environment and differentiates into spore-containing fruiting bodies. In this development, cyclic AMP (cAMP) plays an important role in signaling. In addition, several polyketides, probably synthesized by type III PKSs, seem to play important roles in signaling of this differentiation. Differentiation-inducing factors (DIFs) and 4-methyl-5-pentylbenzene-1,3-diol (MPBD) are the polyketides important for differentiation and are predicted to be synthesized by type III PKSs (**Figure 12(a)**).[74,75]

The genome sequence of *D. discoideum* predicted the presence of two genes encoding a type III PKS. Surprisingly, the two type III PKSs, named Steely 1 and Steely 2, were fused to the C-terminus of a type I FAS.[67] Steely 1 contains KS, AT, DH, MT, ER, KR, and ACP domains, in addition to a type III PKS domain (**Figure 12(b)**). Steely 2 contains KS, AT, DH, ER, KR, and ACP domains, in addition to a type III PKS domain (**Figure 12(b)**). The domain architectures of these type I FASs resemble that of the mammalian type I FAS (**Figure 12(b)**).[76] The MT domain on Steely 1 is located at a position similar to that of the pseudo-methyltransferase domain of the mammalian FAS, which has recently been discovered by analyzing the crystal structure.[76] A difference between Steely 1 and mammalian type I FASs is that the TE domain at the C-terminus of the latter is replaced by a type III PKS domain in the former. Such architecture suggests that these enzymes are involved in the synthesis of DIF and MPBD; the type I FAS domain could synthesize the alkyl chain moiety and the type III PKS domain could synthesize the phenyl ring moiety by using the alkyl chain attached to the ACP domain as a starter substrate (**Figure 12(a)**).

The activity of Steely 2 and its *in vivo* role were investigated by Austin *et al.*[67] The C-terminal type III PKS domain, which was expressed in and purified from *E. coli*, catalyzed the formation of phlorocaprophenone (PCP), the predicted precursor of DIFs, from hexanoyl-CoA and malonyl-CoA. In addition, disruption of Steely 2 resulted in the loss of the ability to synthesize DIF-1, showing that Steely 2 is involved in the biosynthesis of DIF-1. Supplementation of PCP to the Steely 2 disruptant restored the ability to synthesize DIF-1. Hence Steely 2 synthesizes PCP by the reaction depicted in **Figure 12(a)**. Although the Steely 2 disruptant develops into the slug stage, its slug is thin and tends to break up. Furthermore, the Steely 2 disruptant exhibits a phenotype representing a fruiting body stage.[77] However, the outer basal disk of the fruiting body is missing and the stalk lies on the substratum. The crystal structure of the C-terminal type III PKS domain of Steely 1 was also solved by Austin *et al.*[67] Although Steely 1 has an overall structure similar to typical type III PKSs, its enzyme backbone is slightly different from that of alfalfa CHS2. These subtle differences should affect the positions and orientations of the amino acid residues near the active site, which may affect the catalytic property of Steely 1.

Ghosh *et al.*[78] further analyzed Steely 1, which was called DiPKS1 by this group. *In vitro* study of the C-terminal type III PKS domain of Steely 1 revealed that this PKS domain is an alkylresorcinol synthase synthesizing alkylresorcinols from hexanoyl-CoA, octanoyl-CoA, dodecanoyl-CoA, and lauloyl-CoA with by-products of triketide pyrones. They also found that the type III PKS activity was affected by the presence of the ACP domain of Steely 1. An ACP/type III PKS fused domain or didomain was over-expressed in *E. coli* and purified to examine the activity. The product ratio of resorcinol/pyrone increased in the reaction catalyzed by the ACP/type III PKS didomain; the type III PKS produced resorcinols and pyrones in almost equal amounts but the ACP/type III PKS didomain produced almost only resorcinols (**Figure 12(c)**).[78] Therefore, an interaction between the ACP and type III PKS domains is important for the catalysis. Homology modeling by using the modeled ACP domain in the crystal structure (code 1t8k) as a

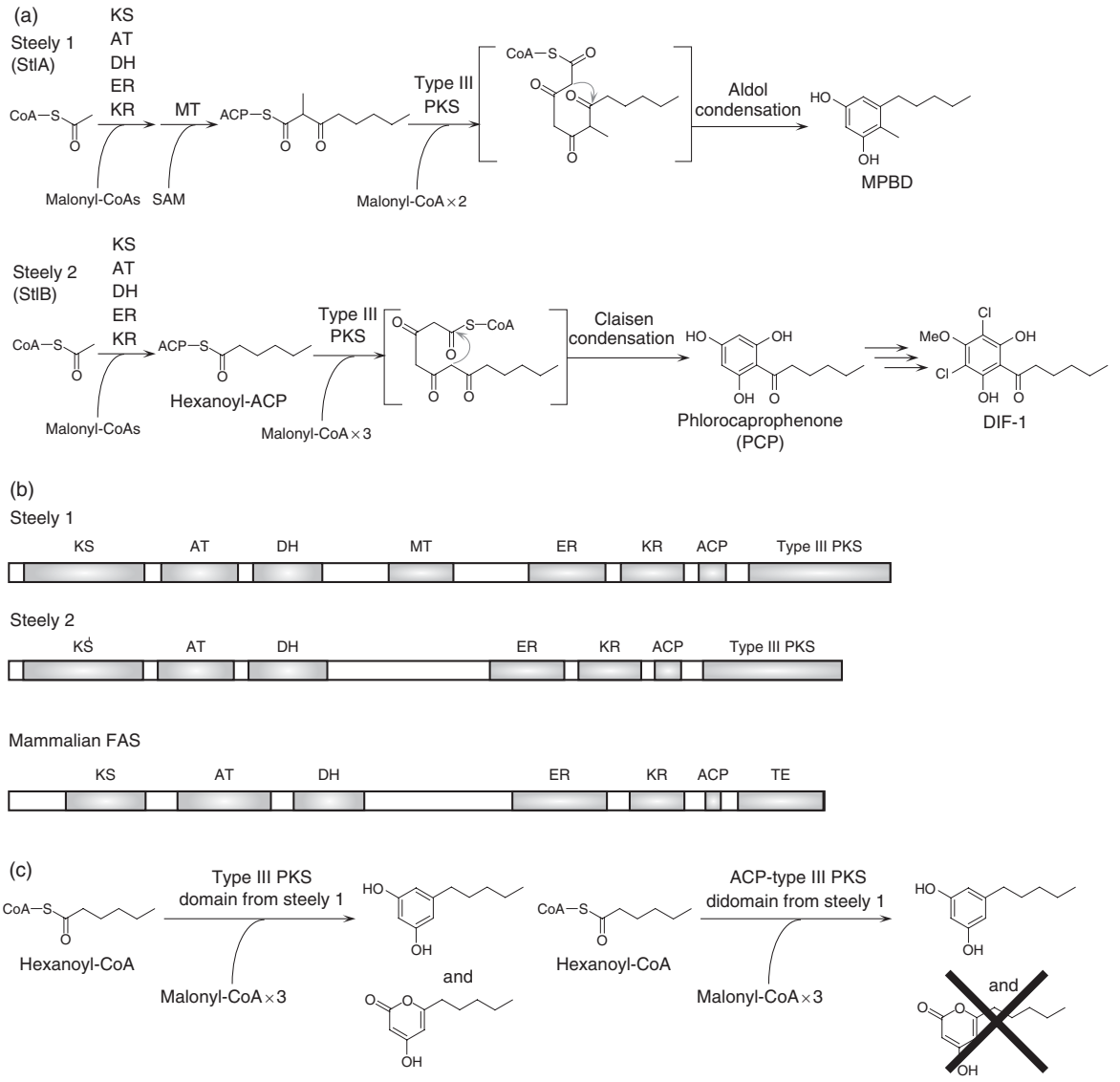


Figure 12 The reactions catalyzed by Steely 1 to produce MPBD and by Steely 2 to produce DIF-1 (a), the domain organizations of Steely 1, Steely 2, and mammalian fatty acid synthase (b), and the difference between the reactions catalyzed by the type III PKS domain of Steely 1 and the ACP/type III PKS didomain of Steely 1(c). Steely 1 catalyzes the formation of MPBD; the FAS domain of Steely 1 synthesizes 3-keto-2-methyl-octanoyl-ACP and the type III PKS domain catalyzes the formation of the resorcinol scaffold (a). The FAS domain of Steely 2 synthesizes hexanoyl-ACP and the type III PKS domain synthesizes the phloroglucinol scaffold (a). The domain architectures of Steely 1 and 2 resemble that of mammalian FAS (b). The reactions catalyzed by the type III PKS domain and the ACP/type III PKS didomain are different; the ACP/type III PKS didomain synthesizes resorcinols as the almost sole product, whereas the type III PKS domain produces resorcinols and triketide pyrones (c).

template and docking study with the crystal structure of the type III PKS domain of Steely 1[78] suggested two potential electrostatic interactions, between Lys2612 of ACP and Glu2829 of type III PKS and between Ser2613 of ACP and Arg2826 of type III PKS, that stabilize the complex. These residues are distinct from the amino acids that were reported to be important for the interaction with ACP.[38] On the basis of these findings, Ghosh *et al.*[78] proposed that Steely 1 catalyzes the formation of MPBD from acetyl-CoA, malonyl-CoA, and *S*-adenosylmethionine according to the reactions shown in **Figure 12(a)**.

1.05.5.2 Fungal Type III PKS

Fungi are also known to produce various bioactive compounds that belong to polyketides, including toxic aflatoxins, melanin pigments, and clinically important lovastatin. The majority of fungal polyketides are synthesized by iterative type I PKSs, which are multifunctional enzymes containing functional domains.[51] Recently, type III PKS homologues were discovered by comparative genome analysis.[79] *Aspergillus oryzae* possesses four type III PKS homologues, CsyA, CsyB, CsyC, and CsyD. Transcription of *csyA*, *csyB*, and *csyD* was observed by reverse transcription polymerase chain reaction (RT-PCR), suggesting that the products of these genes may play some *in vivo* role. Type III PKS homologues also exist in other species of fungi. *Neurospora crassa* and *Fusarium graminearum* possess a single copy of a type III PKS homologue gene, *Magnaporthe grisea* and *Podospora anserine* possess two copies, and *Phanerochaete chrysosporium* possesses three copies.[80] Most of these genes still remain uncharacterized.

The first fungal type III PKS characterized *in vitro* was 2′-oxoalkylresorcylic acid synthase (ORAS) (NCU04801.1) in *N. crassa*.[81] This enzyme is also the first fungal type III PKS whose crystal structure was solved. *In vitro* experiments of ORAS showed that ORAS accepts fatty acid CoA esters of various lengths as starters and synthesizes triketide and tetraketide pyrones and tetraketide and pentaketide resorcinols. In addition, ORAS synthesizes pentaketide resorcylic acids from C_{18} to C_{20} fatty acid CoA esters and a tetraketide resorcylic acid from C_{20} fatty acid CoA ester. Careful analysis of the reaction revealed that ORAS produces pentaketide resorcylic acid from stearoyl-CoA; the experimentally observed pentaketide resorcinol is formed via nonenzymatic decarboxylation of the enzymatically synthesized pentaketide resorcylic acid (**Figure 13(a)**). As described above, on the other hand, ArsB from *A. vinelandii* yields only arkylresorcinols.[61] This difference might be due to the timing of hydrolysis. In the ArsB reaction, the hydrolysis occurs before the decarboxylative condensation (**Figure 13(b)**). Thus, ArsB produces alkylresorcinols. In the ORAS reaction, the aldol condensation occurs before the hydrolysis. Thus, ORAS produces only alkyl resorcylic acid (**Figure 13(c)**).[81]

The crystal structure of ORAS, which was determined by Goyal *et al.*,[82] showed the presence of a hydrophobic tunnel, which can accommodate an alkyl chain, as suggested for THNS and pks18. Curiously, the recombinant protein obtained by Goyal *et al.* did not produce pentaketide resorcylic acid. This may be due to the difference in the preparation procedure of recombinant ORAS. Goyal *et al.*[82] also performed site-directed mutagenesis on amino acids near the aliphatic tunnel and constructed mutant S186F (194 in CHS numbering). The serine residue at this position has been determined to be important for association with the aliphatic tunnel in pks18 and decreased the incorporation of long-chain fatty acid CoA esters.[31] However, this mutant did not show any differences in substrate specificity, suggesting that the phenyl group of Ser186 does not serve as a lid of the tunnel. Therefore, they constructed S186F/S340L double mutant. This mutant lost the ability to incorporate long-chain fatty acid CoA esters. These results suggest that an additional mutation changes the position and orientation of the phenyl residue of Ser186 and blocks the tunnel.

Transcription of *ORAS* occurs at all stages of growth, as determined by RT-PCR and Western blot analysis using polyclonal antibodies generated against the type III PKS domain of Steely 1. In addition, *ORAS* was expressed in vegetative mycelia and hyphae, asexually developed conidia, and sexually generated ascospores. The *in vivo* function of ORAS still remains unclear, however.

1.05.6 Conclusions and Future Perspectives

Many type III PKSs have so far been discovered from a variety of microorganisms. Since RppA was first discovered in *S. griseus*, more than 15 bacterial type III PKSs have been functionally characterized and the crystal structures of three bacterial PKSs have been solved. They possess distinct enzymatic properties different from the type III PKS in plants. For instance, RppA, PhlD, and DpgA incorporate malonyl-CoA as a starter substrate, which is usually an extender substrate in plant type III PKSs.[9,29,46,47,58] The RppA homologue (THNS) from *S. coelicolor* A3(2), pks18, and ORAS possess a buried tunnel in their structures that can accommodate an alkyl chain.[29,31,82] Several type III PKSs discovered from bacteria appear to interact with ACPs, which are involved in fatty acid biosynthesis.[36,62,64,67,78] Many type III PKSs are also accompanied by various enzymes that are involved in the synthesis of substrates and post

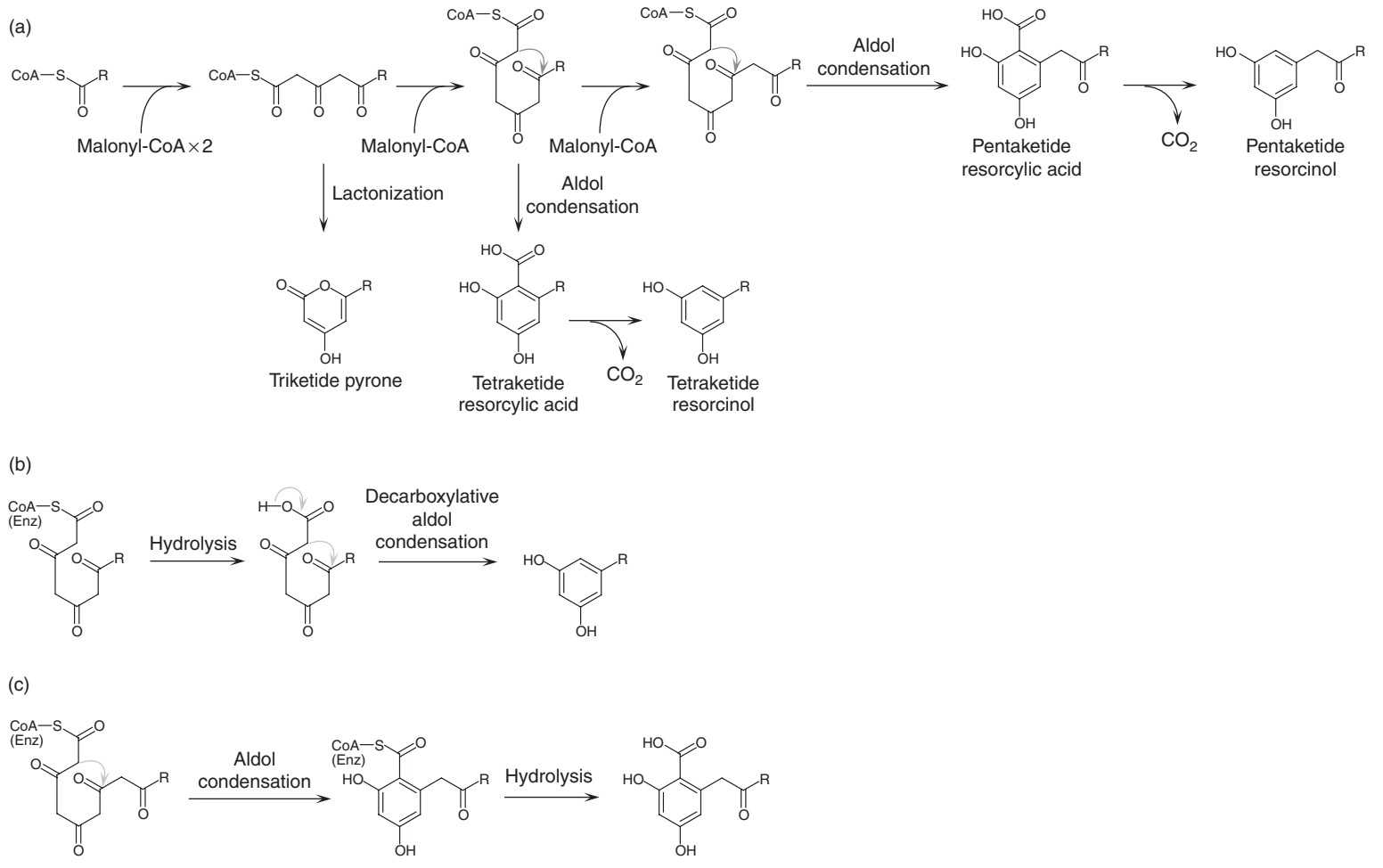


Figure 13 The reaction catalyzed by ORAS (a) and the proposed mechanism of ArsB (b) and ORAS (c) reactions. ORAS synthesizes pentaketide and tetraketide resorcylic acids via aldol condensation (a). In the ArsB reaction, hydrolysis occurs before aldol condensation and the aldol condensation couples with decarboxylation (b). In the ORAS reaction, aldol condensation occurs before hydrolysis, releasing resorcylic acid as the major product (c).

polyketide modification. ArsA and ArsD provide long-chain fatty acids to ArsB and ArsC via a direct interaction.[64] The study of the interaction between a type III PKS and a type I FAS would provide a useful insight into protein–protein interaction. DpgC is a novel enzyme catalyzing the oxidative cleavage of the thioester bond in an unusual manner, in which the oxidation requires no cofactors or metal ions.[49,50] NphB discovered from *Streptomyces* sp. CL190 adds a geranyl moiety to polyketide scaffolds.[25,28] The genes for the fascinating modification enzymes usually form a gene cluster with a type III PKS gene in bacteria. Therefore, exploring type III PKS genes on bacterial genomes may lead to the discovery of enzyme genes whose products are involved in substrate synthesis and modification. Recent progress in combinatorial biosynthesis enables the production of unnatural polyketides by using type III PKSs.[83–86] Type III PKSs and modification enzymes discovered from microorganisms could be utilized to produce unnatural polyketides. The genome sequencing of microorganisms revealed that there are many type III PKSs whose functions are unknown. Analysis of type III PKSs in bacteria should lead to the discovery of fascinating and interesting enzymes and use of them in combinatorial biosynthesis should lead to the production of natural and unnatural polyketides, some of which show useful biological activity.

Abbreviations

2-PS	2-pyrone synthase
2,4-DAPG	2,4-diacetylphloroglucinol
ACP	acyl carrier protein
ars	*Azotobacter* resorcinol synthesis
cAMP	cyclic AMP
CHS	chalcone synthase
CLF	chain length factor
CoA	coenzyme A
CTAL	coumaroyl triacetic acid lactone
CTAS	coumaroyl triacetic acid synthase
CUS	curcuminoid synthase
DH	dehydratase
DIF	differentiation-inducing factor
DPA-CoA	dihydroxyphenylacetyl-CoA
DPG	(*S*)-3,5-dihydroxyphenylglycine
ER	enoylreductase
FAS	fatty acid synthase
Gcs	germicidin synthase
HPQ	1,4,6,7,9,12-hexahydroxyperylene-3,10-quinone
KR	ketoreductase
KS	ketosynthase
MAT	malonyl/acetyl transferase
MPBD	4-methyl-5-pentylbenzene-1,3-diol
MPT	malonyl/palmitoyl transferase
MRSA	methicillin-resistant *Staphylococcus aureus*
OKS	octaketide synthase
ORAS	2′-oxoalkylresorcylic acid synthase
PCP	phlorocaprophenone
PCS	pentaketide chromone synthase
PEG	polyethylene glycol
PKS	polyketide synthase
PPT	phosphopantethinyl transferase
rpp	red pigment production
srs	*Streptomyces* resorcinol synthesis

STS	stilbene synthase
TAL	triacetic acid lactone
TE	thioesterase
THN	1,3,6,8-tetrahydroxynaphthalene
THNS	1,3,6,8-tetrahydroxynaphthalene synthase

References

1. B. Shen, *Curr. Opin. Chem. Biol.* **2003**, *7*, 285–295.
2. S. W. White; J. Zheng; Y. M. Zhang; C. O. Rock, *Annu. Rev. Biochem.* **2005**, *74*, 791–831.
3. S. Smith; A. Witkowski; A. K. Joshi, *Prog. Lipid Res.* **2003**, *42*, 289–317.
4. M. B. Austin; J. P. Noel, *Nat. Prod. Rep.* **2003**, *20*, 79–110.
5. S. Smith; S. C. Tsai, *Nat. Prod. Rep.* **2007**, *24*, 1041–1072.
6. C. Hertweck; A. Luzhetskyy; Y. Rebets; A. Bechthold, *Nat. Prod. Rep.* **2007**, *24*, 162–190.
7. B. S. Moore; J. N. Hopke, *ChemBioChem* **2001**, *2*, 35–38.
8. C. Bender; V. Rangaswamy; J. Loper, *Annu. Rev. Phytopathol.* **1999**, *37*, 175–196.
9. N. Funa; Y. Ohnishi; I. Fujii; M. Shibuya; Y. Ebizuka; S. Horinouchi, *Nature* **1999**, *400*, 897–899.
10. M. G. Bangera; L. S. Thomashow, *J. Bacteriol.* **1999**, *181*, 3155–3163.
11. J. M. Jez; M. B. Austin; J. Ferrer; M. E. Bowman; J. Schröder; J. P. Noel, *Chem. Biol.* **2000**, *7*, 919–930.
12. I. Abe; Y. Utsumi; S. Oguro; H. Morita; Y. Sano; H. Noguchi, *J. Am. Chem. Soc.* **2005**, *127*, 1362–1363.
13. I. Abe; Y. Utsumi; S. Oguro; H. Noguchi, *FEBS Lett.* **2004**, *562*, 171–176.
14. I. Abe; S. Oguro; Y. Utsumi; Y. Sano; H. Noguchi, *J. Am. Chem. Soc.* **2005**, *127*, 12709–12716.
15. Y. Katsuyama; M. Matsuzawa; N. Funa; S. Horinouchi, *J. Biol. Chem.* **2007**, *282*, 37702–37709.
16. J. L. Ferrer; J. M. Jez; M. E. Bowman; R. A. Dixon; J. P. Noel, *Nat. Struct. Biol.* **1999**, *6*, 775–784.
17. M. B. Austin; M. E. Bowman; J. L. Ferrer; J. Schröder; J. P. Noel, *Chem. Biol.* **2004**, *11*, 1179–1194.
18. H. Morita; S. Kondo; S. Oguro; H. Noguchi; S. Sugio; I. Abe; T. Kohno, *Chem. Biol.* **2007**, *14*, 359–369.
19. K. Ueda; K. M. Kim; T. Beppu; S. Horinouchi, *J. Antibiot. (Tokyo)* **1995**, *48*, 638–646.
20. N. Funa; M. Funabashi; Y. Ohnishi; S. Horinouchi, *J. Bacteriol.* **2005**, *187*, 8149–8155.
21. J. Córtes; J. Velasco; G. Foster; A. P. Blackaby; B. A. Rudd; B. Wilkinson, *Mol. Microbiol.* **2002**, *44*, 1213–1224.
22. N. Funa; Y. Ohnishi; Y. Ebizuka; S. Horinouchi, *Biochem. J.* **2002**, *367*, 781–789.
23. M. Izumikawa; P. R. Shipley; J. N. Hopke; T. O'Hare; L. Xiang; J. P. Noel; B. S. Moore, *J. Ind. Microbiol. Biotechnol.* **2003**, *30*, 510–515.
24. N. Funa; M. Funabashi; E. Yoshimura; S. Horinouchi *J. Biol. Chem.* **2005**, *280*, 14514–14523.
25. T. Kuzuyama; J. P. Noel; S. B. Richard, *Nature* **2005**, *435*, 983–987.
26. T. Kawasaki; Y. Hayashi; T. Kuzuyama; K. Furihata; N. Itoh; H. Seto; T. Dairi, *J. Bacteriol.* **2006**, *188*, 1236–1244.
27. J. M. Winter; M. C. Moffitt; E. Zazopoulos; J. B. McAlpine; P. C. Dorrestein; B. S. Moore, *J. Biol. Chem.* **2007**, *282*, 16362–16368.
28. T. Kumano; S. B. Richard; J. P. Noel; M. Nishiyama; T. Kuzuyama, *Bioorg. Med. Chem.* **2008**, *16*, 8117–8126.
29. M. B. Austin; M. Izumikawa; M. E. Bowman; D. W. Udwary; J. L. Ferrer; B. S. Moore; J. P. Noel, *J. Biol. Chem.* **2004**, *279*, 45162–45174.
30. S. Li; S. Grüschow; J. S. Dordick; D. H. Sherman, *J. Biol. Chem.* **2007**, *282*, 12765–12772.
31. R. Sankaranarayanan; P. Saxena; U. B. Marathe; R. S. Gokhale; V. M. Shanmugam; R. Rukmini, *Nat. Struct. Mol. Biol.* **2004**, *11*, 894–900.
32. N. Funa; Y. Ohnishi; Y. Ebizuka; S. Horinouchi, *J. Biol. Chem.* **2002**, *277*, 4628–4635.
33. R. Schüz; W. Heller; K. Hahlbrock, *J. Biol. Chem.* **1983**, *258*, 6730–6734.
34. H. Morita; Y. Takahashi; H. Noguchi; I. Abe, *Biochem. Biophys. Res. Commun.* **2000**, *279*, 190–195.
35. S. D. Bentley; K. F. Chater; A. M. Cerdeño-Tárraga; G. L. Challis; N. R. Thomson; K. D. James; D. E. Harris; M. A. Quail; H. Kieser; D. Harper; A. Bateman; S. Brown; G. Chandra; C. W. Chen; M. Collins; A. Cronin; A. Fraser; A. Goble; J. Hidalgo; T. Hornsby; S. Howarth; C. H. Huang; T. Kieser; L. Larke; L. Murphy; K. Oliver; S. O'Neil; E. Rabbinowitsch; M. A. Rajandream; K. Rutherford; S. Rutter; K. Seeger; D. Saunders; S. Sharp; R. Squares; S. Squares; K. Taylor; T. Warren; A. Wietzorrek; J. Woodward; B. G. Barrell; J. Parkhill; D. A. Hopwood, *Nature* **2002**, *417*, 141–147.
36. L. Song; F. Barona-Gomez; C. Corre; L. Xiang; D. W. Udwary; M. B. Austin; J. P. Noel; B. S. Moore; G. L. Challis, *J. Am. Chem. Soc.* **2006**, *128*, 14754–14755.
37. F. Petersen; H. Zähner; J. W. Metzger; S. Freund; R. P. Hummel, *J. Antibiot. (Tokyo)* **1993**, *46*, 1126–1138.
38. Y. M. Zhang; M. S. Rao; R. J. Heath; A. C. Price; A. J. Olson; C. O. Rock; S. W. White, *J. Biol. Chem.* **2001**, *276*, 8231–8238.
39. S. Grüschow; T. J. Buchholz; W. Seufert; J. S. Dordick; D. H. Sherman, *ChemBioChem* **2007**, *8*, 863–868.
40. L. E. Quadri; P. H. Weinreb; M. Lei; M. M. Nakano; P. Zuber; C. T. Walsh, *Biochemistry* **1998**, *37*, 1585–1595.
41. Y. Ohnishi; J. Ishikawa; H. Hara; H. Suzuki; M. Ikenoya; H. Ikeda; A. Yamashita; M. Hattori; S. Horinouchi, *J. Bacteriol.* **2008**, *190*, 4050–4060.
42. M. Funabashi; N. Funa; S. Horinouchi, *J. Biol. Chem.* **2008**, *283*, 13983–13991.
43. J. Schröder; S. Raiber; T. Berger; A. Schmidt; J. Schmidt; A. M. Soares-Sello; E. Bardshiri; D. Strack; T. J. Simpson; M. Veit; G. Schröder, *Biochemistry* **1998**, *37*, 8417–8425.
44. C. Nakano; H. Ozawa; G. Akanuma; N. Funa; S. Horinouchi, *J. Bacteriol.* **2009**, *191*, 4916–4923.

45. S. Pelzer; R. Süssmuth; D. Heckmann; J. Recktenwald; P. Huber; G. Jung; W. Wohlleben, *Antimicrob. Agents Chemother.* **1999**, *43*, 1565–1573.
46. V. Pfeifer; G. J. Nicholson; J. Ries; J. Recktenwald; A. B. Schefer; R. M. Shawky; J. Schröder; W. Wohlleben; S. Pelzer, *J. Biol. Chem.* **2001**, *276*, 38370–38377.
47. H. Chen; C. C. Tseng; B. K. Hubbard; C. T. Walsh, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14901–14906.
48. C. C. Tseng; S. M. McLoughlin; N. L. Kelleher; C. T. Walsh, *Biochemistry* **2004**, *43*, 970–980.
49. P. F. Widboom; E. N. Fielding; Y. Liu; S. D. Bruner, *Nature* **2007**, *447*, 342–345.
50. C. C. Tseng; F. H. Vaillancourt; S. D. Bruner; C. T. Walsh, *Chem. Biol.* **2004**, *11*, 1195–1203.
51. H. B. Bode; R. Müller, *J. Ind. Microbiol. Biotechnol.* **2006**, *33*, 577–588.
52. R. J. Cox, *Org. Biomol. Chem.* **2007**, *5*, 2010–2026.
53. P. E. Kolattukudy; N. D. Fernandes; A. K. Azad; A. M. Fitzmaurice; T. D. Sirakova, *Mol. Microbiol.* **1997**, *24*, 263–270.
54. R. S. Gokhale; P. Saxena; T. Chopra; D. Mohanty, *Nat. Prod. Rep.* **2007**, *24*, 267–277.
55. S. T. Cole; R. Brosch; J. Parkhill; T. Garnier; C. Churcher; D. Harris; S. V. Gordon; K. Eiglmeier; S. Gas; C. E. Barry, 3rd; F. Tekaia; K. Badcock; D. Basham; D. Brown; T. Chillingworth; R. Connor; R. Davies; K. Devlin; T. Feltwell; S. Gentles; N. Hamlin; S. Holroyd; T. Hornsby; K. Jagels; A. Krogh; J. McLean; S. Moule; L. Murphy; K. Oliver; J. Osborne; M. A. Quail; M. A. Rajandream; J. Rogers; S. Rutter; K. Seeger; J. Skelton; R. Squares; S. Squares; J. E. Sulston; K. Taylor; S. Whitehead; B. G. Barrell, *Nature* **1998**, *393*, 537–544.
56. P. Saxena; G. Yadav; D. Mohanty; R. S. Gokhale, *J. Biol. Chem.* **2003**, *278*, 44780–44790.
57. H. Afsharmanesh; M. Ahmadzadeh; A. Sharifi-Tehrani; M. Javan-Nikkhah; K. Ghazanfari, *Commun. Agric. Appl. Biol. Sci.* **2007**, *72*, 941–950.
58. J. Achkar; M. Xian; H. Zhao; J. W. Frost, *J. Am. Chem. Soc.* **2005**, *127*, 5332–5333.
59. W. Zha; S. B. Rubin-Pitel; H. Zhao, *J. Biol. Chem.* **2006**, *281*, 32036–32047.
60. C. J. Su; R. Reusch; H. L. Sadoff, *J. Bacteriol.* **1979**, *137*, 1434–1436.
61. R. N. Reusch; H. L. Sadoff, *Nature* **1983**, *302*, 268–270.
62. N. Funa; H. Ozawa; A. Hirata; S. Horinouchi, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 6356–6361.
63. J. E. Berleman; B. M. Hasselbring; C. E. Bauer, *J. Bacteriol.* **2004**, *186*, 5834–5841.
64. A. Miyanaga; N. Funa; T. Awakawa; S. Horinouchi, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 871–876.
65. B. Chakravarty; Z. Gu; S. S. Chirala; S. J. Wakil; F. A. Quiocho, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15567–15572.
66. F. Lynen, *Eur. J. Biochem.* **1980**, *112*, 431–442.
67. M. B. Austin; T. Saito; M. E. Bowman; S. Haydock; A. Kato; B. S. Moore; R. R. Kay; J. P. Noel, *Nat. Chem. Biol.* **2006**, *2*, 494–502.
68. K. Gerth; S. Pradella; O. Perlova; S. Beyer; R. Muller, *J. Biotechnol.* **2003**, *106*, 233–253.
69. B. S. Goldman; W. C. Nierman; D. Kaiser; S. C. Slater; A. S. Durkin; J. A. Eisen; C. M. Ronning; W. B. Barbazuk; M. Blanchard; C. Field; C. Halling; G. Hinkle; O. Iartchuk; H. S. Kim; C. Mackenzie; R. Madupu; N. Miller; A. Shvartsbeyn; S. A. Sullivan; M. Vaudin; R. Wiegand; H. B. Kaplan, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15200–15205.
70. S. Schneiker; O. Perlova; O. Kaiser; K. Gerth; A. Alici; M. O. Altmeyer; D. Bartels; T. Bekel; S. Beyer; E. Bode; H. B. Bode; C. J. Bolten; J. V. Choudhuri; S. Doss; Y. A. Elnakady; B. Frank; L. Gaigalat; A. Goesmann; C. Groeger; F. Gross; L. Jelsbak; J. Kalinowski; C. Kegler; T. Knauber; S. Konietzny; M. Kopp; L. Krause; D. Krug; B. Linke; T. Mahmud; R. Martinez-Arias; A. C. McHardy; M. Merai; F. Meyer; S. Mormann; J. Muñoz-Dorado; J. Perez; S. Pradella; S. Rachid; G. Raddatz; F. Rosenau; C. Rückert; F. Sasse; M. Scharfe; S. C. Schuster; G. Suen; A. Treuner-Lange; G. J. Velicer; F. J. Vorhölter; K. J. Weissman; R. D. Welch; S. C. Wenzel; D. E. Whitworth; S. Wilhelm; C. Wittmann; H. Blöcker; A. Pühler; R. Müller, *Nat. Biotechnol.* **2007**, *25*, 1281–1289.
71. F. Gross; N. Luniak; O. Perlova; N. Gaitatzis; H. Jenke-Kodama; K. Gerth; D. Gottschalk; E. Dittmann; R. Müller, *Arch. Microbiol.* **2006**, *185*, 28–38.
72. L. Eichinger; J. A. Pachebat; G. Glockner; M. A. Rajandream; R. Sucgang; M. Berriman; J. Song; R. Olsen; K. Szafranski; Q. Xu; B. Tunggal; S. Kummerfeld; M. Madera; B. A. Konfortov; F. Rivero; A. T. Bankier; R. Lehmann; N. Hamlin; R. Davies; P. Gaudet; P. Fey; K. Pilcher; G. Chen; D. Saunders; E. Sodergren; P. Davis; A. Kerhornou; X. Nie; N. Hall; C. Anjard; L. Hemphill; N. Bason; P. Farbrother; B. Desany; E. Just; T. Morio; R. Rost; C. Churcher; J. Cooper; S. Haydock; N. van Driessche; A. Cronin; I. Goodhead; D. Muzny; T. Mourier; A. Pain; M. Lu; D. Harper; R. Lindsay; H. Hauser; K. James; M. Quiles; M. Madan Babu; T. Saito; C. Buchrieser; A. Wardroper; M. Felder; M. Thangavelu; D. Johnson; A. Knights; H. Loulseged; K. Mungall; K. Oliver; C. Price; M. A. Quail; H. Urushihara; J. Hernandez; E. Rabbinowitsch; D. Steffen; M. Sanders; J. Ma; Y. Kohara; S. Sharp; M. Simmonds; S. Spiegler; A. Tivey; S. Sugano; B. White; D. Walker; J. Woodward; T. Winckler; Y. Tanaka; G. Shaulsky; M. Schleicher; G. Weinstock; A. Rosenthal; E. C. Cox; R. L. Chisholm; R. Gibbs; W. F. Loomis; M. Platzer; R. R. Kay; J. Williams; P. H. Dear; A. A. Noegel; B. Barrell; A. Kuspa, *Nature* **2005**, *435*, 43–57.
73. D. C. Mahadeo; C. A. Parent, *Curr. Top. Dev. Biol.* **2006**, *73*, 115–140.
74. C. R. Thompson; R. R. Kay, *Mol. Cell* **2000**, *6*, 1509–1514.
75. H. Kikuchi; Y. Oshima; A. Ichimura; N. Gokan; A. Hasegawa; K. Hosaka; Y. Kubohara, *Life Sci.* **2006**, *80*, 160–165.
76. T. Maier; M. Leibundgut; N. Ban, *Science* **2008**, *321*, 1315–1322.
77. T. Saito; A. Kato; R. R. Kay, *Dev. Biol.* **2008**, *317*, 444–453.
78. R. Ghosh; A. Chhabra; P. A. Phatale; S. K. Samrat; J. Sharma; A. Gosain; D. Mohanty; S. Saran; R. S. Gokhale, *J. Biol. Chem.* **2008**, *283*, 11348–11354.
79. Y. Seshime; P. R. Juvvadi; I. Fujii; K. Kitamoto, *Biochem. Biophys. Res. Commun.* **2005**, *331*, 253–260.
80. P. R. Juvvadi; Y. Seshime; K. Kitamoto, *J. Microbiol.* **2005**, *43*, 475–486.
81. N. Funa; T. Awakawa; S. Horinouchi, *J. Biol. Chem.* **2007**, *282*, 14476–14481.
82. A. Goyal; P. Saxena; A. Rahman; P. K. Singh; D. P. Kasbekar; R. S. Gokhale; R. Sankaranarayanan, *J. Struct. Biol.* **2008**, *162*, 411–421.
83. J. A. Chemler; Y. Yan; E. Leonard; M. A. Koffas, *Org. Lett.* **2007**, *9*, 1855–1858.
84. W. Zhang; Y. Tang, *J. Med. Chem.* **2008**, *51*, 2629–2633.
85. Y. Katsuyama; N. Funa; S. Horinouchi, *Biotechnol. J.* **2007**, *2*, 1286–1293.
86. Y. Katsuyama; N. Funa; I. Miyahisa; S. Horinouchi, *Chem. Biol.* **2007**, *14*, 613–621.

Biographical Sketches

Yohei Katsuyama graduated from the Department of Agricultural Chemistry, The University of Tokyo in 2005. His research field is basic and applied microbiology. Currently, he is working on combinatorial biosynthesis of unnatural plant polyketides by microorganisms and analysis of catalytic mechanisms of type III polyketide synthase.

Sueharu Horinouchi graduated from the Department of Agricultural Chemistry, The University of Tokyo in 1974. He received his Ph.D. in 1979 from The University of Tokyo. He spent 2 years from 1979 to 1981 as Project Associate in Bernard Weisblum's laboratory, Pharmacology Department, University of Wisconsin, where he discovered the mechanism of posttranslational regulation of the macrolide resistance genes in pathogenic bacteria. He got a position as Assistant Professor in 1981, as Associate Professor in 1987, and Professor in 1994 at the Department of Biotechnology, The University of Tokyo. He was the Director of Biotechnology Research Center, The University of Tokyo from 2003 to 2005. His research field is basic and applied microbiology.

1.06 Plant Type III PKS

Hiroyuki Morita and Ikuro Abe, The University of Tokyo, Hongo, Tokyo, Japan

Hiroshi Noguchi, University of Shizuoka, Yada, Shizuoka, Japan

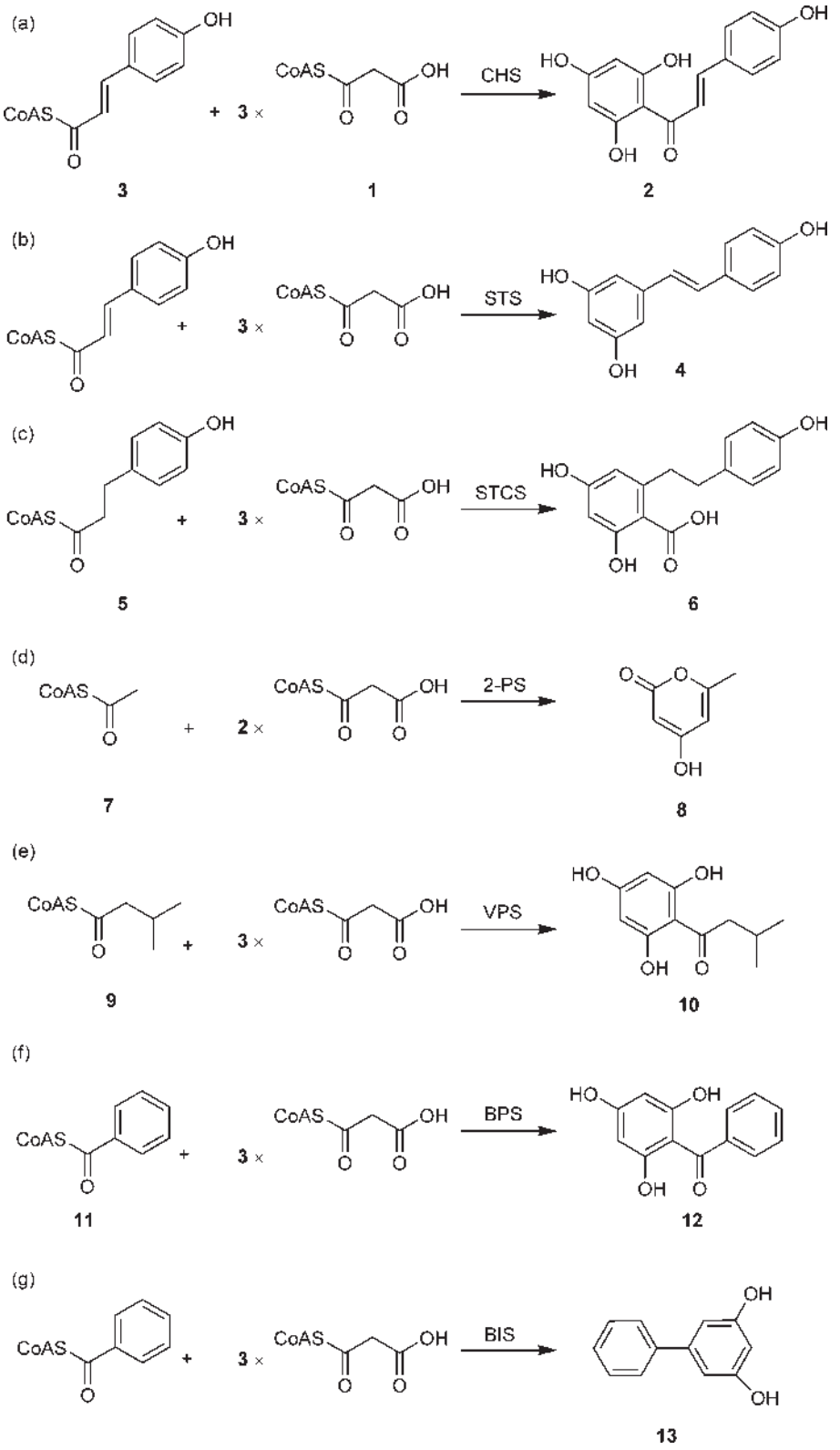
© 2010 Elsevier Ltd. All rights reserved.

1.06.1	**Introduction**	171
1.06.2	**Functional Diversity and Catalytic Potential**	174
1.06.2.1	Chalcone Synthase	178
1.06.2.2	Homoeriodictyol/Eriodictyol Synthase	181
1.06.2.3	Benzophenone Synthase	181
1.06.2.4	Acridone Synthase	187
1.06.2.5	Phlorisovalerophenone Synthase and Isobutyrophenone Synthase	187
1.06.2.6	Stilbene Synthase	187
1.06.2.7	Bibenzyl Synthase and Biphenyl Synthase	191
1.06.2.8	Coumaroyl Triacetic Acid Lactone Synthase and Stilbenecarboxylate Synthase	191
1.06.2.9	*Cannabis sativa* Polyketide Synthase-1	192
1.06.2.10	Benzalacetone Synthase	195
1.06.2.11	Curcuminoid Synthase	195
1.06.2.12	2-Pyrone Synthase	195
1.06.2.13	Long-chain 2-Pyrone Synthase	199
1.06.2.14	Pentaketide Chromone Synthase	199
1.06.2.15	Aloesone Synthase	199
1.06.2.16	Hexaketide Synthase and Octaketide Synthase	201
1.06.2.17	*C*-Methylchalcone Synthase	205
1.06.3	**Enzyme Structure and Site-Directed Mutagenesis**	205
1.06.4	**Protein Engineering**	214
1.06.5	**Combinational Biosynthesis**	215
1.06.6	**Conclusions**	220
References		221

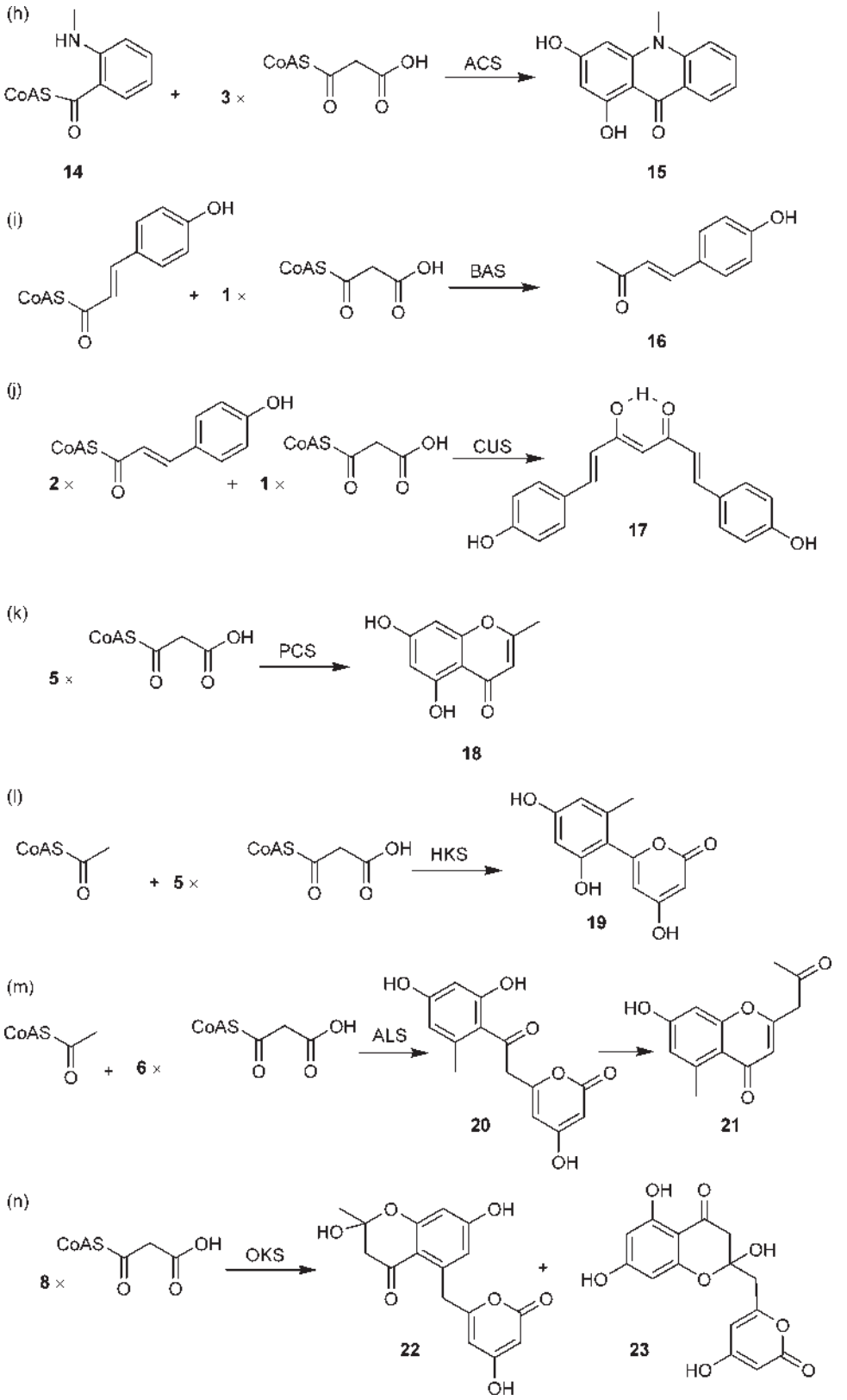
1.06.1 Introduction

Plant polyketides such as chalcone, stilbene, phloroglucinol, chromone, and curcuminoid are one of the largest and important families of natural secondary products. They have served as antibiotics, anticancer drugs, antifungal agents, immunosuppressants, and insecticides.[1–3] Their chemical structures are rich in variety as well as pharmaceutical activity; however, recent studies have begun to reveal that they are biosynthesized by a group of chalcone synthase (CHS) superfamily of type III polyketide synthases (PKSs).[4–6]

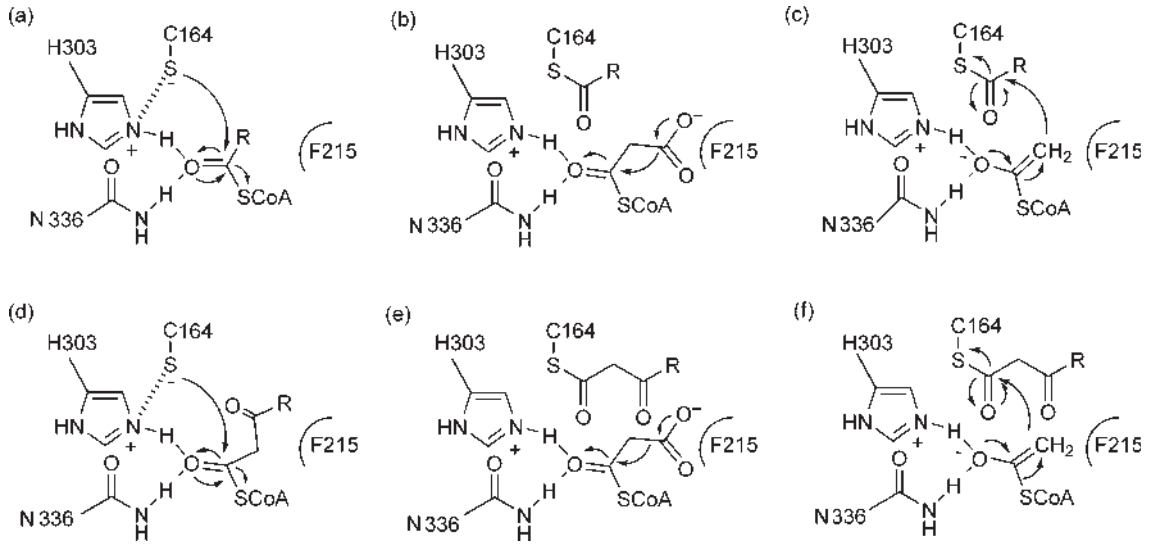
The type III PKSs are structurally and mechanistically distinct from the type I (modular type) and type II (subunit type) PKSs of bacterial origin, using free CoA thioesters as substrates without the involvement of acyl carrier protein. The enzymes catalyze iterative decarboxylative condensations of malonyl-CoA (**1**) with a CoA-linked starter molecule. For example, CHS, the pivotal enzyme in flavonoid biosynthesis and the most-studied enzyme in the type III PKSs, produces naringenin chalcone (**2**) through the sequential condensation of 4-coumaroyl-CoA (**3**) with three C_2 units from malonyl-CoA (**1**) (**Scheme 1a**).[4,5] A growing number of functionally diverse plant type III PKSs, such as stilbene synthase (STS),[7–10] acridone synthase (ACS),[11,12] and 2-pyrone synthase (2-PS),[13,14] sharing 60–75% amino acid sequence identity with each other, have been cloned and characterized (**Scheme 1**). In addition, bacterial type III PKSs, such as 1,3,6,8-tetrahydroxynaphthalene synthase (THNS)[15] and 2′-oxoalkylresorcylic acid synthase (ORAS),[16] have also been reported.



Scheme 1 (Continued)



Scheme 1 Typical biosynthesis reactions catalyzed by plant type III PKSs. Abbreviations: ACS, acridone synthase; ALS, aloesone synthase; BAS, benzalacetone synthase; BBS, bibenzyl synthase; BPS, benzophenone synthase; CHS, chalcone synthase; CUS, curcuminoid synthase; HKS, hexaketide synthase; OKS, octaketide synthase; PCS, pentaketide chromone synthase; 2-PS, 2-pyrone synthase; STS, stilbene synthase; STCS, stilbenecarboxylate synthase; VPS, phlorisovalerophenone synthase.



Scheme 2 Proposed common reaction pathway for the enzyme type III PKS reaction. (a) Starter substrate loading, (b) malonyl-CoA decarboxylation, (c) formation of diketide intermediate, (d) stabilization of diketide intermediate, (e) malonyl-CoA decarboxylation, and (f) formation of triketide intermediate.

The CHS-superfamily enzymes are homodimers of 40–45 kDa proteins, catalyzing the assembly of complex natural products by successive decarboxylative condensations of malonyl-CoA (**1**) in a biosynthetic process that closely parallels fatty acid biosynthesis. The type III PKS reaction consists of starter molecule loading, malonyl-CoA decarboxylation, polyketide chain elongation, and subsequent cyclization and aromatization of the enzyme-bound intermediate, the combinational difference of which results in the functional diversity of the type III PKS enzymes. Plant and bacterial type III PKSs share a common three-dimensional overall fold with a conserved Cys–His–Asn catalytic triad. A common active site machinery of the polyketide formation reactions proceeds through starter molecule loading at the active-site Cys, malonyl-CoA decarboxylation, polyketide chain elongation, and subsequent cyclization and aromatization of the enzyme-bound intermediate (**Scheme 2**).[4,5] Recent crystallographic and site-directed studies revealed that only a small modification of the active site architecture generates the functional diversity of the type III PKSs.[4,17–25] Principally, the volume and shape of the initiation/elongation/cyclization cavity govern starter molecule selectivity, polyketide chain length, and the folding and cyclization pathways of the type III PKSs.

1.06.2 Functional Diversity and Catalytic Potential

Since the first isolation of *chs* gene from parsley (*Petroselium hortense*) in 1983,[26] more than 862 type III pks genes have been reported in databases (NCBI: http://www.ncbi.nlm.gov/). The plant type III PKSs consist of approximately 400 amino acid long polypeptide chain (41–44 kDa) of identical ketoacyl synthase (KAS) domain and form a homodimer. The plant type III PKSs share 38–95% amino acid sequence identity with each other, and at least 20 functionally different plant type III PKSs have been isolated (**Scheme 1** and **Table 1**). Initially, the type III PKS enzymes were regarded as plant-specific enzymes and hence were called CHS-superfamily enzymes. However, CHS-related enzymes sharing 29–31% amino acid sequence identity were discovered in bacteria,[15,27,28] and hence the enzymes were called type III PKS enzymes. Deduced amino acid comparison and phylogenetic tree of the type III PKSs are show in **Figures 1** and **2**, respectively. It should be noted that STS, grouping with CHS from the same or related plants, has been proposed to have evolved independently several times from CHS.[29]

The type III PKS catalyzes repetitive condensation reaction of acetate unit derived from malonyl-CoA (**1**) initialized with a starter substrate as well as that of fatty acid synthase (FAS) in preliminary metabolite

Table 1 Examples of plant type III PKSs, their preferred substrates, and reaction products

Enzyme (type of ring closure, ring type)	*Substrates (starter, extender*[a]*, no. condensation)*	*Product*	*Plant species*	*Ref.*
None cyclization				
Benzalacetone synthase (BAS), EC 2.3.1.-	4-Coumaroyl-CoA (**3**), (**1**) (1X)	4-Hydroxybenzalacetone (**16**)	*Rheum palmatum*	40,41
	Feruloyl-CoA (**25**), (**1**) (1X)	3-Methoxy-4-hydroxybenzalacetone (**43**)	*Rubus idaeus,*	43
	5-Hydroxyferuloyl-CoA (**26**), (**1**) (1X)	3-Methoxy-4,5-hydroxybenzalacetone (**44**)		44
Curcuminoid synthase (CUS)	4-Coumaroyl-CoA (**3**), (**1**) (1X)	Bisdemethoxycurcumin (**17**)	*Oryza sativa*	42
	4-Coumaroyl-CoA (**3**) (1X)			45
	Cinnamoyl-CoA (**27**), (**1**) (1X)	Cinnamoyl(hexanoyl)methane (**45**)		
	3-Oxo-octanoic acid (**28**) (1X)	(gingerol (**46**) analog)		
One cyclization (Heterocyclic)				
Benzalacetone synthase (BAS), EC 2.3.1.-	*N*-Methylanthraniloyl-CoA (**14**) or anthraniloyl-CoA (**29**), (**1**) or (**24**)	4-Hydroxy-2(1*H*)quinolones (**47–50**)	*R. palmatum*	46
CTAS - type (Lactonization, heterocyclic)	(1X)			
C-Methylchalcone synthase (PstrCHS2)	Diketide NAC (**30**), (**24**) (1X)	Methylstyrylpyrone (**51**)	*Pinus strobes*	47
2-Pyrone synthase (2-PS)	Acetyl-CoA (**7**), (**1**) (2X)	Triacetic acid lactone (TAL) (**8**)	*Gerbera hybrida*	13,14
Long-chain-2-pyrone synthase (PKS-A, PKS-B)	C_6-C_{20} acyl-CoAs (**31–38**), (**1**) (2X)	C_5-C_{19} arkyl-2-pyrones (**52–59**) and C_5-C_{19} arkylacetyl-2-pyrone (**60–67**)	*Arabidopsis thaliana*	48
4-Coumaroyltriacetic acid synthase (CTAS)	4-Coumaroyl-CoA (**3**), (**1**) (3X)	4-Coumaroyltriacetic acid lactone (CTAL) (**68**)	*Hydrangea macrophylla Var.*	49
CHS - type (Claisen, aromatic)			thunbergii	
Chalcone synthase (CHS), EC 2.3.1.74	4-Coumaroyl-CoA (**3**), (**1**) (3X)	Naringenin chalcone (**2**)	*Medicago sativa*	4,17
Phlorisovalerophenone synthase (VPS), EC 2.3.1.156	Isovaleryl-CoA (**9**), (**1**) (3X)	Phlorisovalerophenone (**10**)	*Humulus lupulus*	30,31
Isobutyrophenone synthase (BUS)	Isobutyryl-CoA (**39**), (**1**) (3X)	Phlorisobutyrophenone (**69**)	*Hypericum calycinum* *Centaurium erythraea*	50
Benzophenone synthase (BPS), EC 2.3.1.151	3-Hydroxybenzoyl-CoA (**40**), (**1**) (3X) Benzoyl-CoA (**11**), (**1**) (3X)	2,3′,4,6-Tetrahydroxybenzophenone 2,4,6-Trihydroxybenzophenone (**12**)	*Hypericum androsaemum*	51,52
Acridone synthase (ACS), EC 2.3.1.159	*N*-Methylanthraniloyl-CoA (**14**), (**1**) (3X)	1,3-Dihydroxy-*N*-methylacridone (**15**)	*Ruta graveolens* *Huperzia serrata*	53,54 55
Homoeriodictyol/ eriodictyol synthase (HEDS or HvCHS)	Feruloyl-CoA (**25**), (**1**) (3X) Caffeoyl-CoA (**41**), (**1**) (3X)	Homoeriodictyol chalcone (**71**) Eriodictyol chalcone (**72**)	*Hordeum vulgare*	56

(*Continued*)

Table 1 (Continued)

Enzyme (type of ring closure, ring type)	*Substrates (starter, extender*[a]*, no. condensation)*	*Product*	*Plant species*	*Ref.*
STS - type				
(Aldol with -CO_2, aromatic)				
Stilbene synthase (STS), EC 2.3.1.95	4-Coumaroyl-CoA (**3**), (**1**) (3X)	Resveratrol (**4**)	*Arachis hypogaea*	4,5,7, 57
Pinosylvin synthase, EC 2.3.1.146	Cinnamoyl-CoA (**27**), (**1**) (3X)	Pinosylvin (**73**)	*Pinus sylvestris, Pinus strobes*	8,58,59
Bibenzyl synthase (BBS)	3-Hydroxyphenylpropionyl-CoA (**42**), (**1**) (3X)	3,3′,5-Trihydroxybibenzyl (**74**)	*Phalaenopsis* sp., *Bletilla striata*	33,34
Biphenyl synthase (BIS)	Benzoyl-CoA (**11**), (**1**) (3X)	3,5-Dihydroxybiphenyl (**13**)	*Sorbus aucuparia H. macrophylla,*	35
(Aldol aromatic)				
Stilbenecarboxylate synthase (STCS)	Dihydro-4-coumaroyl-CoA (**5**), (**1**) (3X)	5-Hydroxylunularic acid (**6**)	*Marchantia polymorpha*	36
More than 2 cyclization				
Miscellaneous type				
(Aldol, aromatic, heterocyclic)				
Pentaketide chromone synthase (PCS)	Malonyl-CoA (**1**), (**1**) (4X)	5,7-Dihydroxy-2-methylchromone (**18**)	*Aloe arborescens*	32,60
Hexaketide synthase (HKS)	Acetyl-CoA (**7**), (**1**) (5X)	6-(2′,4′-Dihydroxy-6′-methylphenyl)-4-hydroxy-2-pyrone (**19**)	*D. lusitanicum Plumbago indica*	61,62
Aloesone synthase (ALS)	Acetyl-CoA (**7**), (**1**) (6X)	Aloesone (**21**)	*R. palmatum, A. arborescens*	37,63
Octaketide synthase (OKS)	Malonyl-CoA (**1**), (**1**) (7X)	SEK4 (**22**) and SEK4b (**23**) (octaketides)	*A. arborescens Hypericum perforatum*	38,39

[a] malonyl-CoA (**1**), methylmalonyl-CoA (**24**)

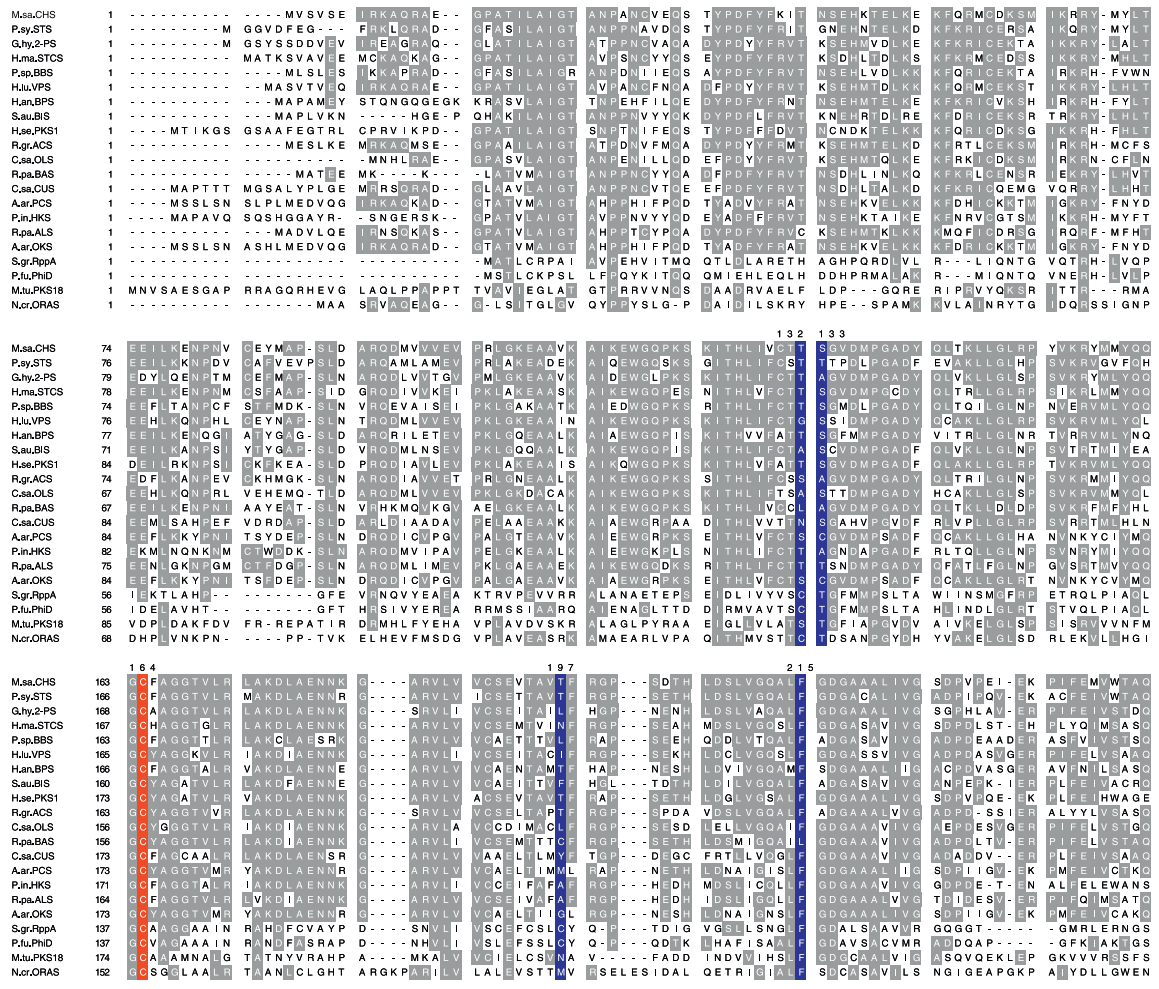


Figure 1 (Continued)

(**Scheme 2**). Indeed, recent structural evidence suggests that the type III PKSs emerged by gain of function from the structurally similar homodimeric KAS III, an enzyme of the initial step of type II fatty acid biosynthesis.[4] FAS performs reduction and dehydration reactions, whereas PKSs omit or modify these later reactions, thereby preserving varying degrees of polar chemical reactivity along portions of the growing linear polyketide chain.

The functional diversity of the CHS-superfamily type III PKSs is derived from the differences in their selection of starter substrate, number of acetate unit condensation, and mechanisms of cyclization and aromatization reactions (**Scheme 1** and **Table 1**). They utilize a variety of different CoA-linked starter substrates ranging from aliphatic-CoA to aromatic-CoA, from small (acetyl-CoA (**7**)) to bulky (4-coumaroyl (**3**) and *N*-methylanthraniloyl-CoA (**14**)) substrates, or from polar (malonyl-CoA (**1**)) to nonpolar substrates (isovaleryl-CoA (**9**)) giving the plants an extraordinary functional diversification. For the number of chain extension, although many of the type III PKSs condense up to three molecules of malonyl-CoA (**1**), they vary between one and eight. Cyclization reaction is basically classified into three types: Claisen and aldol cyclizations and lactonization. Claisen cyclization is catalyzed by enzymes such as CHS, ACS,[11] phlorisovalerophenone synthase (VPS),[30,31] and pentaketide chromone synthase (PCS).[32] The enzymes form a C-C bond from C6 position carbon to C1 position carbon of the catalytic cysteine-bound intermediate. STS,[7] bibenzyl synthase (BBS),[33,34] biphenyl synthase (BIS),[35] stilbenecarboxylate synthase (STCS),[36] aloesone

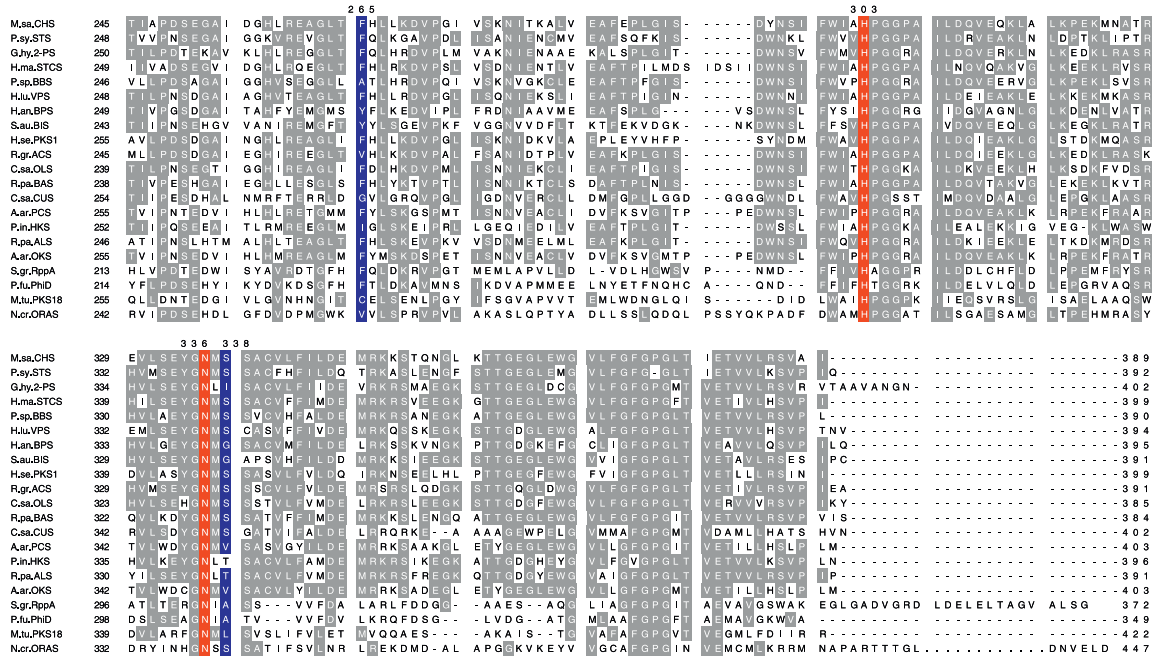


Figure 1 Example of sequence alignment of plant type III polyketide synthases (PKSs) with bacterial and fungal type III PKSs. The catalytic triad of Cys-His-Asn is colored red. The residues thought to be crucial for the functional diversity of type III PKSs are highlighted in blue (numbering in *Medicago sativa* CHS2). Abbreviations (GenBank accession numbers): A.ar.OKS, *Aloe arborescens* OKS (AAT48709); A.ar.PCS, *A. arborescens* PCS (AAX35541); C.sa.OLS, *Cannabis sativa* OLS (BAG14339); H.an.BPS, *Hypericum androsaemum* BPS (AAL79808); H.lu.VPS, *Humulus lupulus* VPS (BAA29039); H.ma.STCS, *Hydrangea macrophylla* L. STCS (AAN76182); H.se.PKS1, *Huperzia serrata* PKS1 (ABI94386); M.sa.CHS, *Medicago sativa* CHS2 (P30074); M.tu.PKS18, *Mycobacterium tuberculosis* PKS18 (AAK45681); N.cr.ORAS, *Neurospora crassa* ORAS (XP_960427); O.sa.CUS, *Oryza sativa* CUS (Os07g17010.1_ORYZA); P.fu.PhlD, *Pseudomonas fluorescens* PhlD (AAB48106); P.in.HKS, *Plumbago indica* HKS (BAF44539); P.sp.BBS, *Phalaenopsis* sp. BBS (CAA56276); P.sy.STS, *Pinus sylvestris* STS (AAB24341); *Gerbera hybrida* 2-PS (P48391); R.gr.ACS, *Ruta graveolens* ACS (CAC14058); R.pa.ALS, *Rheum palmatum* ALS (AAS87170); R.pa.BAS, *Rheum palmatum* BAS (AAK82824); S.au.BIS, *Sorbus aucuparia* BIS (ABB89212); S.gr.RppA, *Streptomyces griseus* RppA (BAA33495).

synthase (ALS),[37] and octaketide synthase (OKS)[38,39] represent the aldol-type cyclized enzymes, but the cyclized pattern is basically classified into two types. For example, STS, BBS, and BIS produce the polyketides through a C-C bond formation from C2 position carbon to C7 position carbon with an additional decarboxylative loss of the C1 as CO_2. STCS also cyclizes the same position, although the C1 position carbon is maintained. On the contrary, enzymes such as ALS and OKS make a new C-C bond between carbonyl carbon of the starter substrate and methylene carbon of the third extender on the intermediate without a decarboxylation. 2-PS[13,14] is a typical example of enzymes involved in pyrone ring formation reactions. However, such pyrone ring formation has remained unclear irrespective of whether the lactonization is enzymatically carried out or not. Together with the different types of cyclization mentioned above, some plant PKSs such as benzalacetone synthase (BAS)[40,41] and curcuminoid synthase (CUS)[42] only catalyze condensation reactions without a cyclization.

1.06.2.1 Chalcone Synthase

CHS is the most-studied prototype type III PKS that catalyzes condensation of the C_6–C_3 unit of 4-coumaroyl-CoA (**3**) with three C_2 units from malonyl-CoA (**1**) to form a new aromatic ring system (**Scheme 3**).[4,5] The reaction is initiated by binding of 4-coumaroyl-CoA (**3**) followed by formation of a thioester at the active site

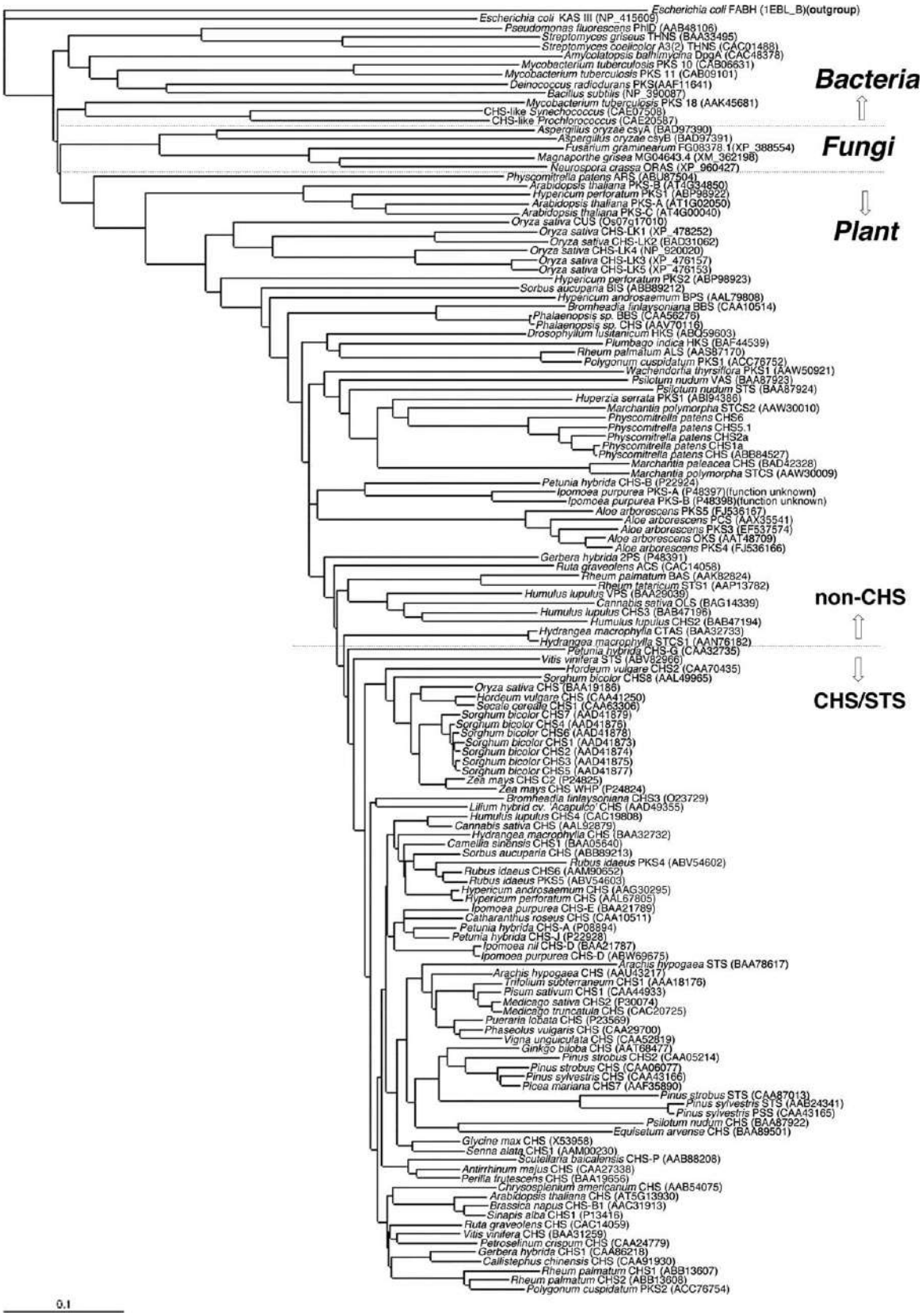
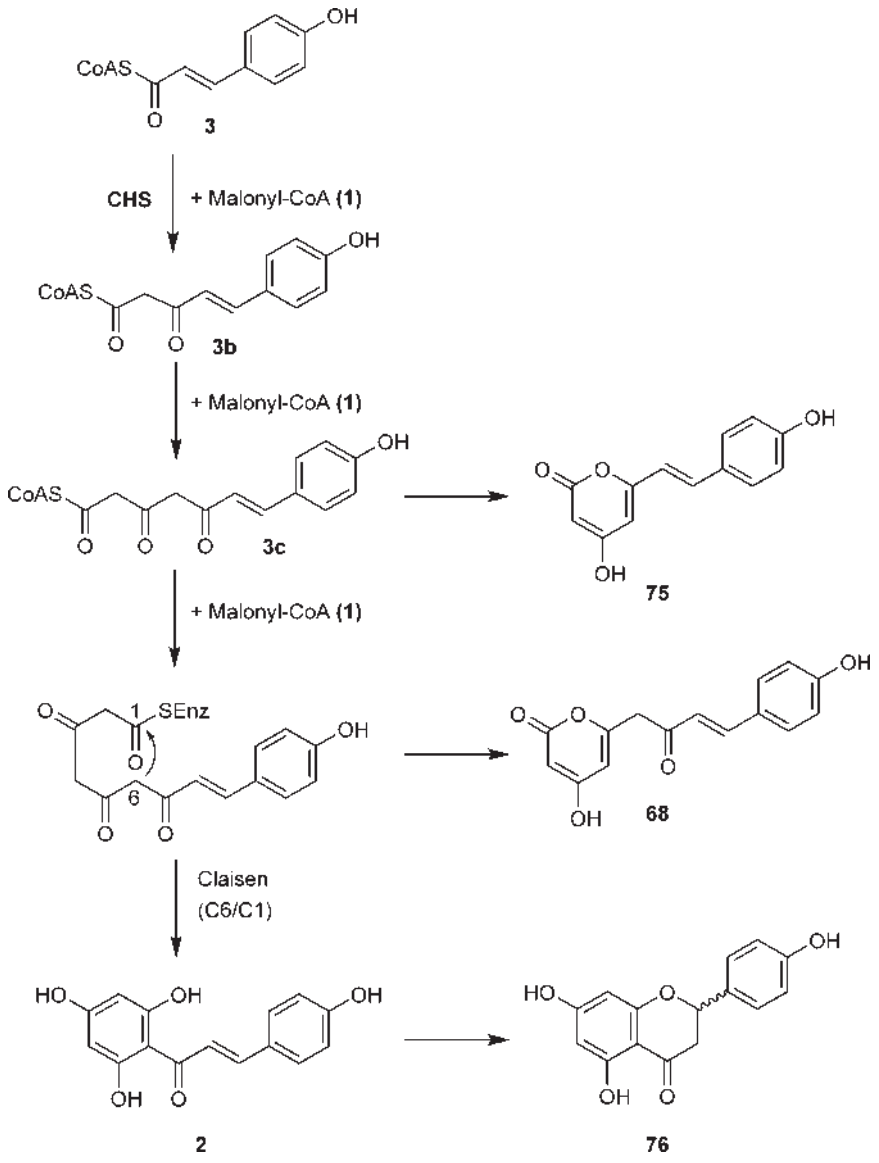


Figure 2 Phylogenetic tree analysis of plant and bacterial type III PKSs. A total of 139 amino acid sequences of type III PKS were aligned, and phylogenetic tree was developed with the CLUSTAL W (1.8) program (DNA Data Bank of Japan, URL: http://www.ddbj.nig.ac.jp). The β-ketoacyl carrier protein synthase III (FABH) of *Escherichia coli* was used as an outgroup. The indicated scale represents 0.1 amino acid substitutions per site.



Scheme 3 Proposed mechanism for the conversion of 4-coumaroyl-CoA (**3**) to naringenin chalcone (**2**), by-products bisnoryangonin (BNY) (**75**) and 4-coumaroyltriacetic acid lactone (CTAL) (**68**), and the tetraketide intermediate (**3c**).

cysteine of the enzyme. After three rounds of sequential decarboxylative Claisen condensation, cyclization and aromatization of the enzyme-bound tetraketide intermediate lead to the formation of naringenin chalcone (**2**), the key intermediate in the biosynthesis of flavonoids. In addition to chalcone, two lactone derivatives, bisnoryangonin (BNY) (**75**)[64] and 4-coumaroyltriacetic acid lactone (CTAL) (**68**),[65] are formed as early-released derailment by-products when the enzyme reaction is carried out *in vitro*.

The first demonstration of CHS activity *in vitro* was reported in 1972 with extracts from parsley, but initially this enzyme was identified as 'flavanone synthase', which was the result of flavanone naringenin (**76**) formation from 4-coumaroyl-CoA (**3**) and malonyl-CoA (**1**).[65] However, later experiments in 1980 using purified enzymes from parsley[66] and tulip[67] revealed that the rapid isomerization from chalcone to flavanone occurs spontaneously and nonstereospecifically in aqueous solutions through Michael-like ring closure, and the CHS function was corrected as described above. Now, it is known that the flavanone is biosynthesized by chalcone isomerase (CHI) *in vivo*. Further characterization of the enzyme confirmed dimer formation of 40–45 kDa

proteins, the fatty acid biosynthesis similarity such as CO_2 exchange at the malonate moiety, inhibition by cerulenin[68] and the catalytic cysteine of intermediate binding residue,[69] and the unnecessity of acyl carrier protein (or phosphopantetheine arm) in the reactions.[70] X-ray crystal structures in the absence and presence of ligands are now available.[17]

Interestingly, CHS shows a remarkable broad substrate specificity. The enzyme accepts a variety of nonphysiological substrate analogs to produce a series of unnatural polyketides (**Scheme 4**).[71] For example, when 4-fluorocinnamoyl-CoA (**77a**) was incubated with CHS from *Scutellaria baicalensis*, a fluorinated flavanone (**77e**) was obtained along with lactone by-products (**77b, 77c**), whereas other 4-substituted analogs (X = Cl (**78a**), Br (**79a**), and OCH_3 (**80a**)) afforded only lactones (**78b–80b, 78c–80c**) (**Scheme 4a**). On the contrary, 4-coumaroyl-CoA analogs in which the coumaroyl aromatic ring is replaced by furan (**81a**) or thiophene (**82a**) are efficiently converted into novel unnatural polyketides (**81b–81e, 82b–82e**) with the heteroaromatic ring system. Steric and/or electronic perturbations by the substituents alter the stability of the enzyme-bound intermediate or the optimally folded conformation in the cyclization pocket of the active site of the enzyme (**Scheme 4b**).[71]

In addition, CHS also accepts benzoyl-CoA (**11**), cinnamoyl-CoA (**27**), and phenylacetyl-CoA (**83**) to afford phlorobenzophenone (2,4,6-trihydroxybenzophenone (**12**)), 4′-deoxychalcone (pinocembrin chalcone (**84**)), and phlorobenzylketone (2,4,6-trihydroxyphenylbenzylketone (**85**)), respectively. Furthermore, CHS also accepts aliphatic-CoA starters (isovaleryl (**9**), isobutyryl (**39**), *n*-hexanoyl (**31**), *n*-octanoyl (**32**), *n*-decanoyl (**33**), and *n*-dodecanoyl (**34**)) to produce triketide and tetraketide lactones in the presence or absence of tetraketide phloroglucinols (**Scheme 5**).[37,72,73]

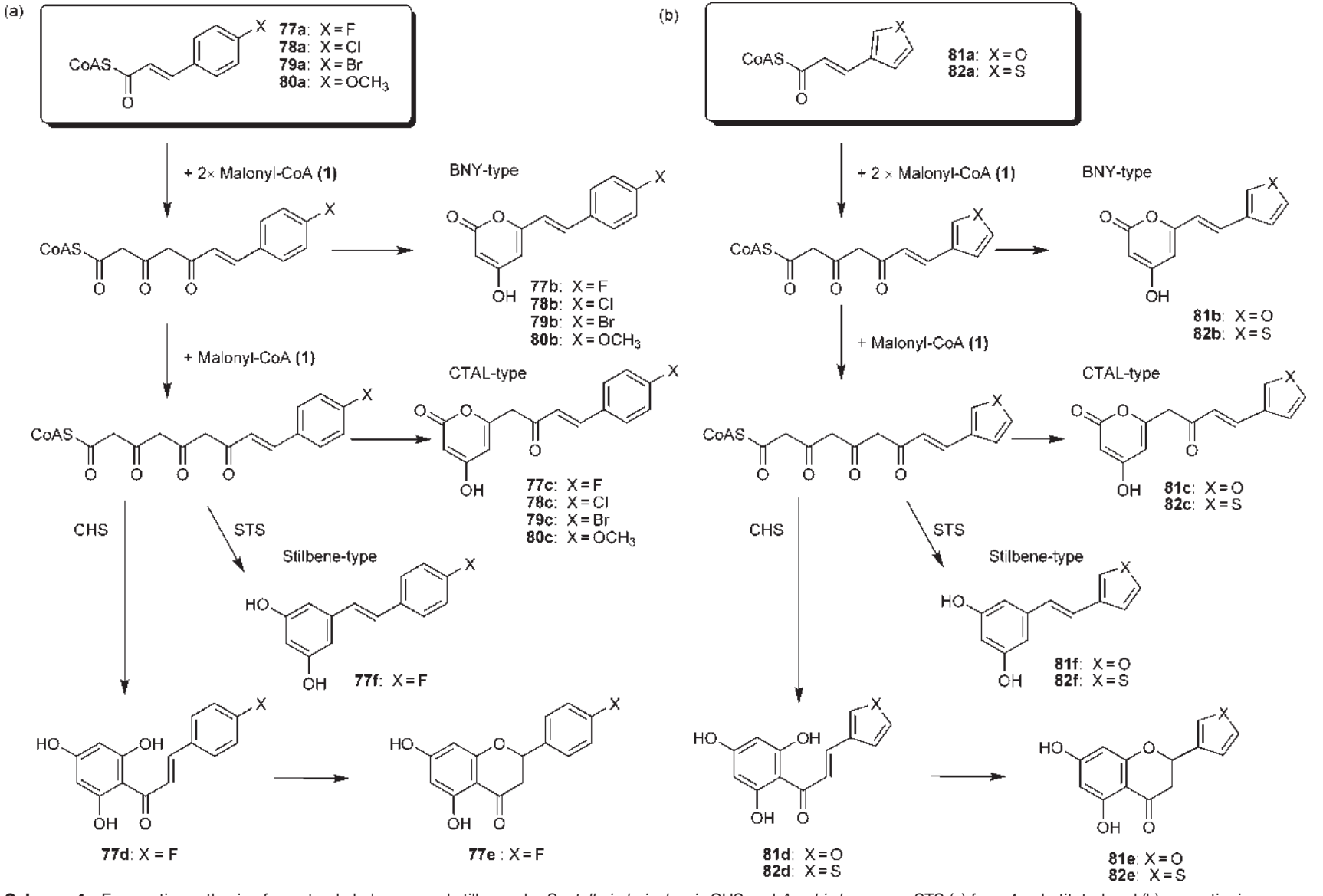
Moreover, it is surprising that CHS even accepts *N*-acetylcysteamine (NAC) thioester of cinnamic acid (cinnamoyl-NAC (**86**)) and an NAC derivative corresponding to the diketide intermediate (cinnamoyl diketide-NAC (**30**)) to produce pinocembrin chalcone (**84**). Detailed analysis of enzyme kinetics as well as full characterization of the enzyme reaction products of NAC-thioesters has been reported (**Scheme 6**).[37]

1.06.2.2 Homoeriodictyol/Eriodictyol Synthase

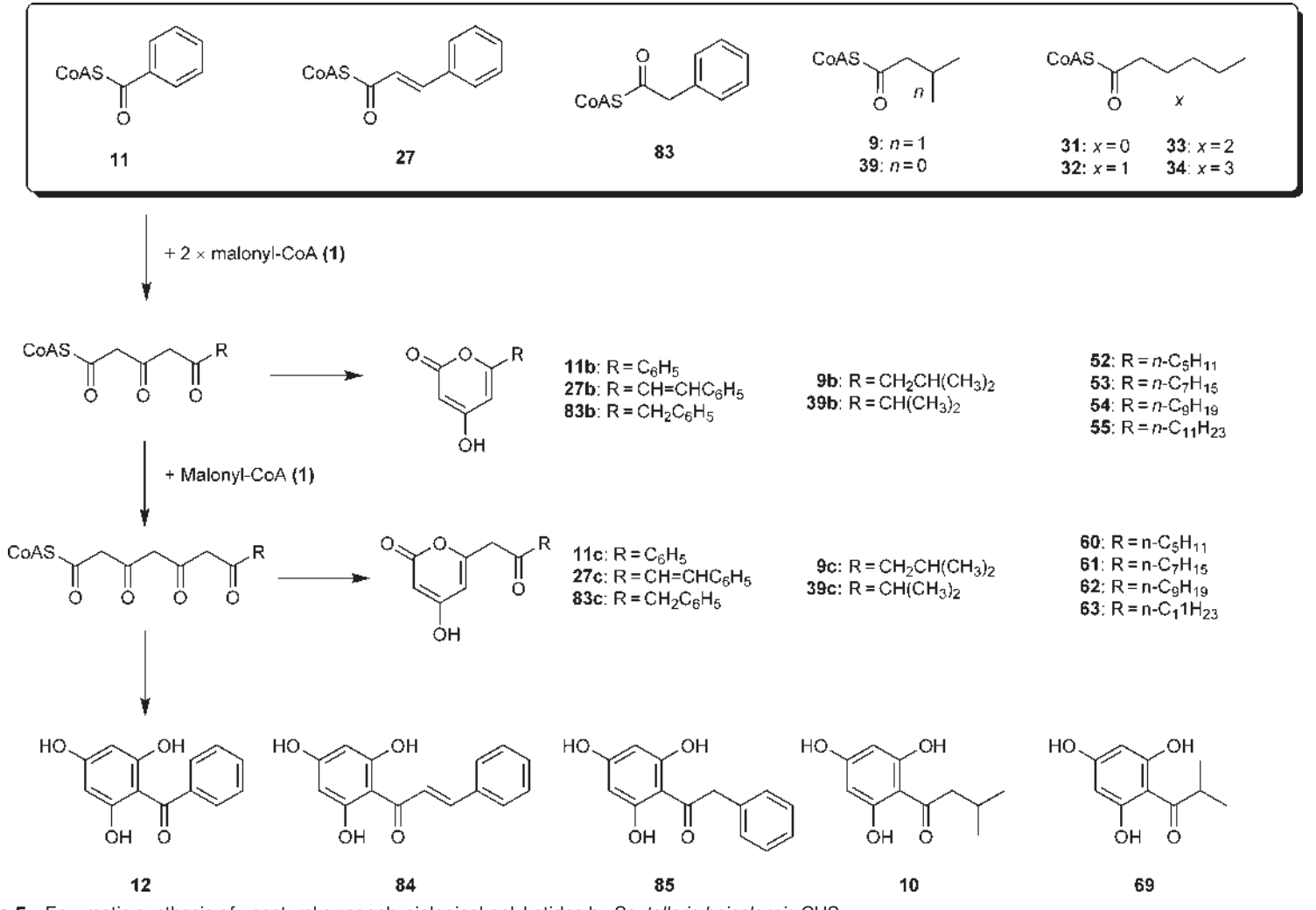
Homoeriodictyol/eriodictyol synthase (HEDS or HvCHS2) has been cloned from barley (*Hormorden vulgare*) leaves inoculated with the fungus *Blumeria graminis* f. sp. *hordei (Bgh)*.[56] The enzyme produces chalcone through the same reaction pathway as prototype CHS. However, both HvCHS2 and prototype CHS differ in their substrate selectivity. HvCHS2 prefers feruloyl-CoA (**25**) and caffeoyl-CoA (**41**) to produce homoeriodictyol chalcone (**71**) and eriodictyol chalcone (**72**), respectively, at the highest rate, whereas 4-coumaroyl-CoA (**3**) and cinnamoyl-CoA (**27**) are poor substrates (**Scheme 7**). In contrast, prototype CHS does not efficiently accept the disubstituted aromatic ring CoA-linked thioesters to produce the corresponding chalcones. Interestingly, eriodictyol-derived phytoalexine lutonarin (**87**) drastically accumulates in barley leaves in response to infection with *Bgh*,[74] which is in good agreement with the expression of HvCHS2 introduced by the infection of the fungus.[56] HvCHS2 is thus labeled as the specific type III PKS enzyme that is directly involved in the biosynthesis of lutonarin in barley.

1.06.2.3 Benzophenone Synthase

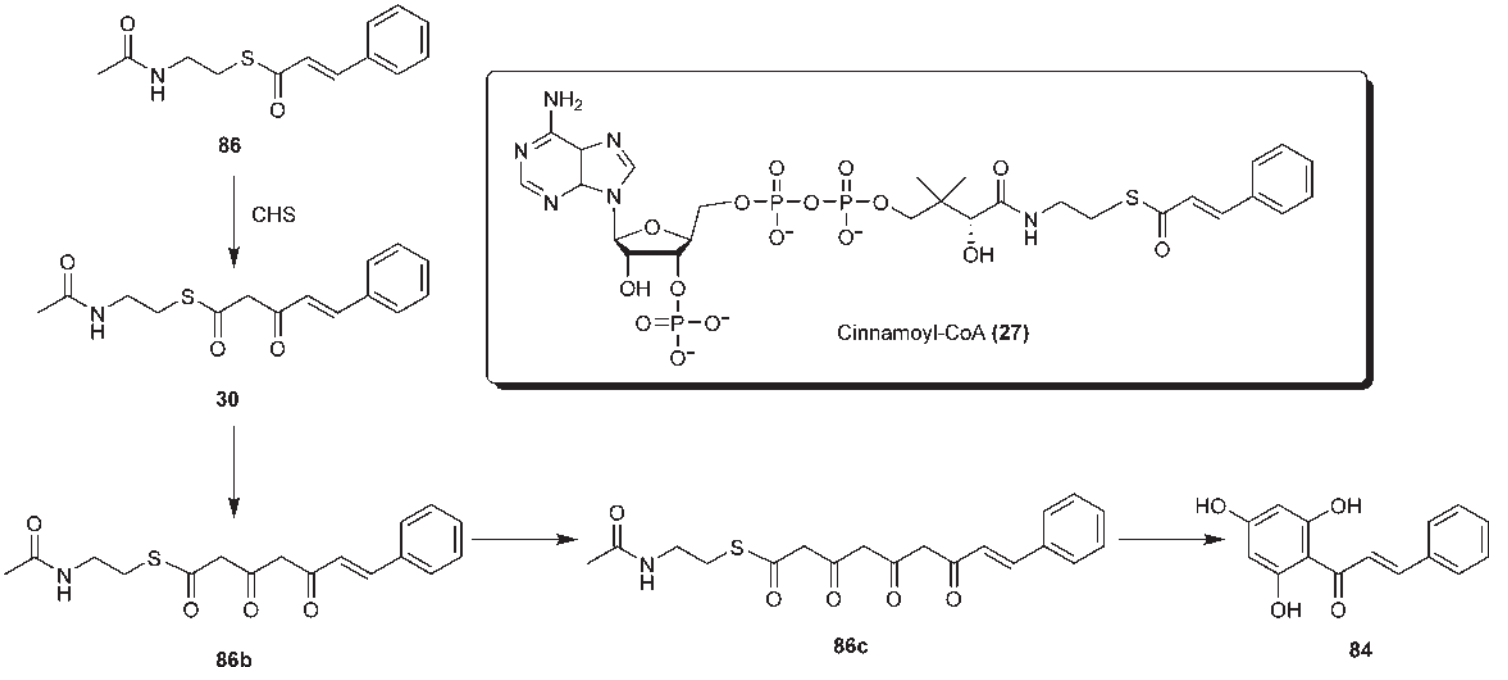
As a key intermediate, benzophenone plays an important role in the biosynthesis of xanthones, a class of compounds widely distributed in plants.[75] Xanthones exhibit significant biological activities; for example, guttiferone F (**88**) and sampsonione A (**89**) show anti-HIV and cytotoxic activities, respectively.[76,77] A related type III PKS is benzophenone synthase (BPS), which was discovered in *Hypericum androsaemum* (Hypericaceae)[52,78] and *Centaurium erythraea* (Gentianaceae).[51] However, in contrast to prototype CHS, the enzyme accepts only CoA-linked thioesters of benzoic acid derivatives and undergoes the Claisen cyclization reaction (**Scheme 8**). For example, BPS from *H. androsaemum* catalyzes the biosynthesis of 2,4,6-trihydroxybenzophenone (**12**) by condensation of benzoyl-CoA (**11**) starter with three molecules of malonyl-CoA (**1**). In Hypericaceae, 2,4,6-trihydroxybenzophenone (**12**) is thought to be then converted into 2,3′,4,6-tetrahydroxybenzophenone (**70**).[78] On the contrary, BBS from *C. erythraea* directly yields 2,3′,4,6-tetrahydroxybenzophenone (**70**) from 3-hydroxylated benzoyl-CoA (**40**). Then, 2,3′,4,6-tetrahydroxybenzophenone



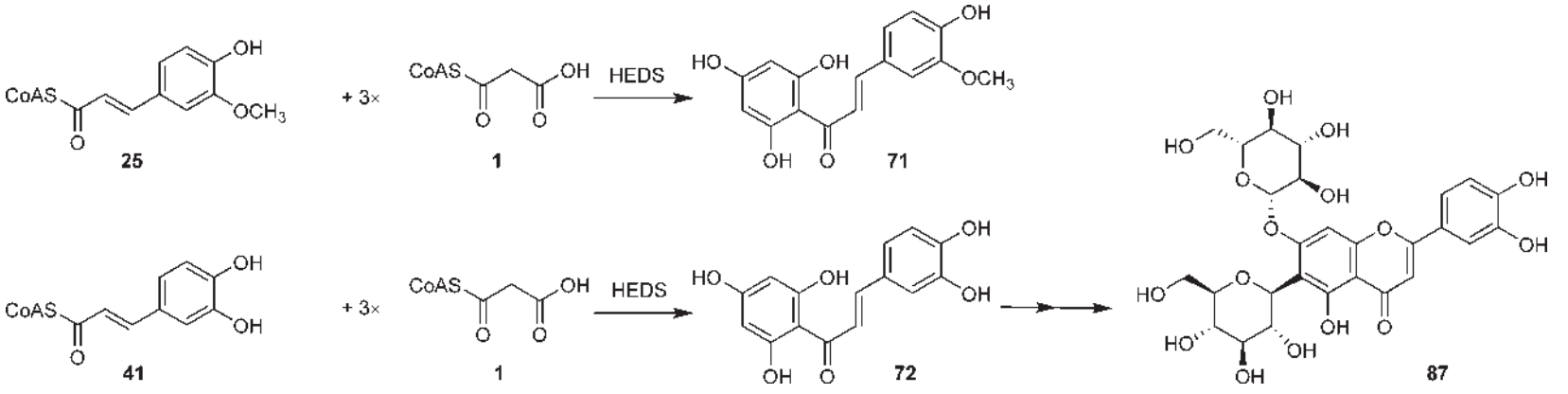
Scheme 4 Enzymatic synthesis of unnatural chalcones and stilbenes by *Scutellaria baicalensis* CHS and *Arachis hypogaea* STS (a) from 4-substituted and (b) aromatic ring-substituted cinnamoyl-CoA analogs.



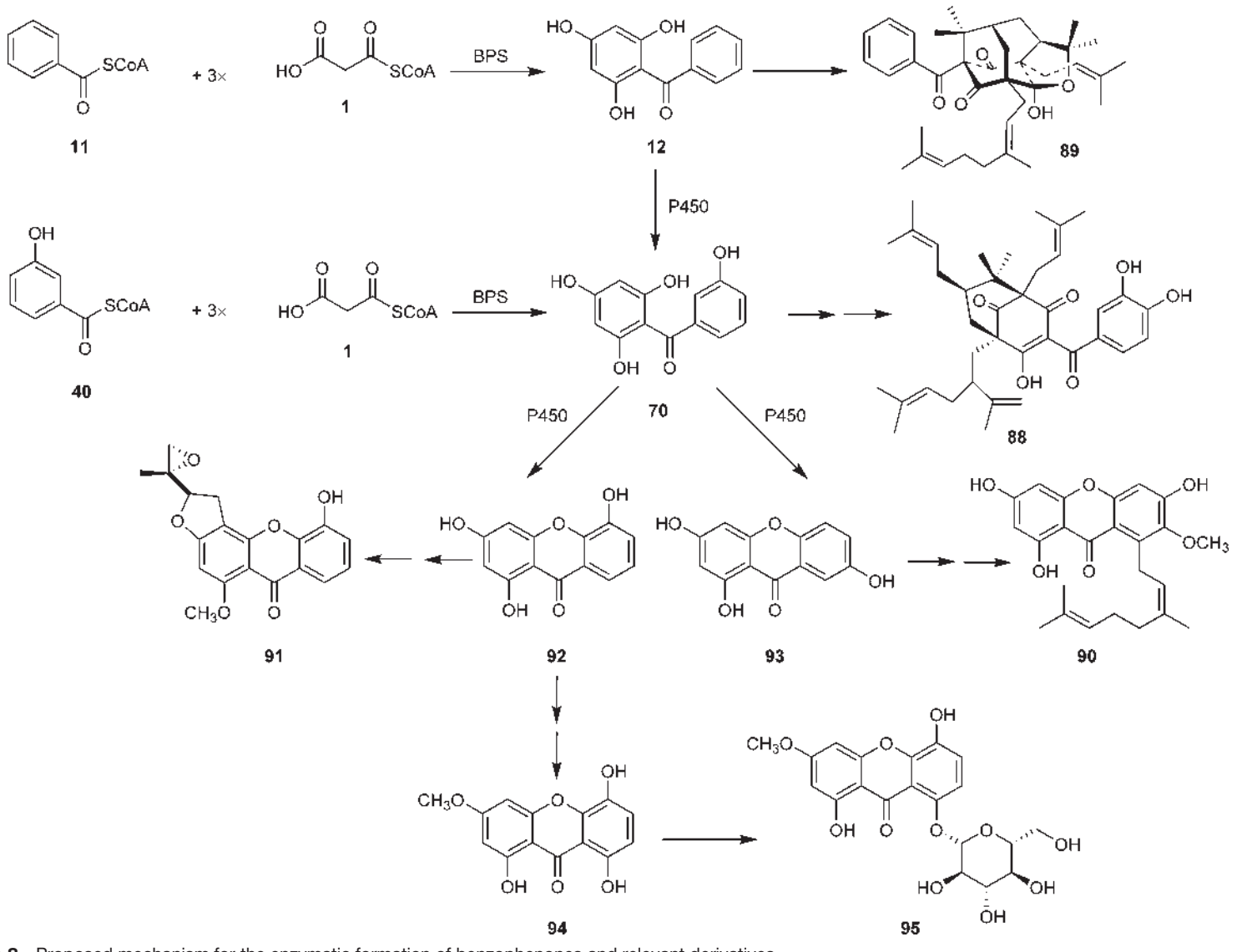
Scheme 5 Enzymatic synthesis of unnatural or nonphysiological polyketides by *Scutellaria baicalensis* CHS.



Scheme 6 Enzymatic synthesis of pinocembrin chalcone (**84**) from cinnamoyl-NAC (**86**) and cinnamoyl diketide-NAC (**30**) thioesters as a startes, and its putative reaction mechanism.



Scheme 7 Enzymatic formation of homoeriodictyol (**71**) and eriodictyol chalcones (**72**), and relative derivative lutonarin (**87**).



Scheme 8 Proposed mechanism for the enzymatic formation of benzophenones and relevant derivatives.

metabolite is further converted into prenylated xanthones such as rubraxanthone (**90**) and psorospermin (**91**), which exhibit significant antibacterial and antitumor activities, respectively, and glycosylated derivatives such as acetylcholinesterase inhibitors swertianolin (**94**) and its precursor bellidifolin (**95**).[79–81]

1.06.2.4 Acridone Synthase

Anthranilic acid has been thought to be a key intermediate in the biosynthesis of acridone and quinolinone alkaloids, which occur in greatest abundance in plants from the family of Rutaceae.[82] In fact, ACS from *Ruta graveolens* is a plant-specific type III PKS that selects *N*-methylanthraniloyl-CoA (**14**) as a starter and performs three condensations with malonyl-CoA (**1**) and CHS-like ring folding to produce 1,3-dihydroxy-*N*-methylacridone (**15**).[11,12,53] 1,3-Dihydroxy-*N*-methylacridone (**15**) is further metabolized to acridone derivatives including melicopicine (**96**), acronycine (**97**), rutacridone (**98**), and rutacridone epoxide (**99**) (**Scheme 9**). Unlike the Michael-like ring closure reaction, formation of the heterocyclic middle ring is likely to occur possibly involving the formation of a Schiff base, which is accompanied by elimination of a water molecule to produce the acridone skeleton.[54] However, it is unclear whether the heterocyclic ring closure is caused by intrinsic activity of the enzyme or by spontaneous cyclization of the enzyme-released product.

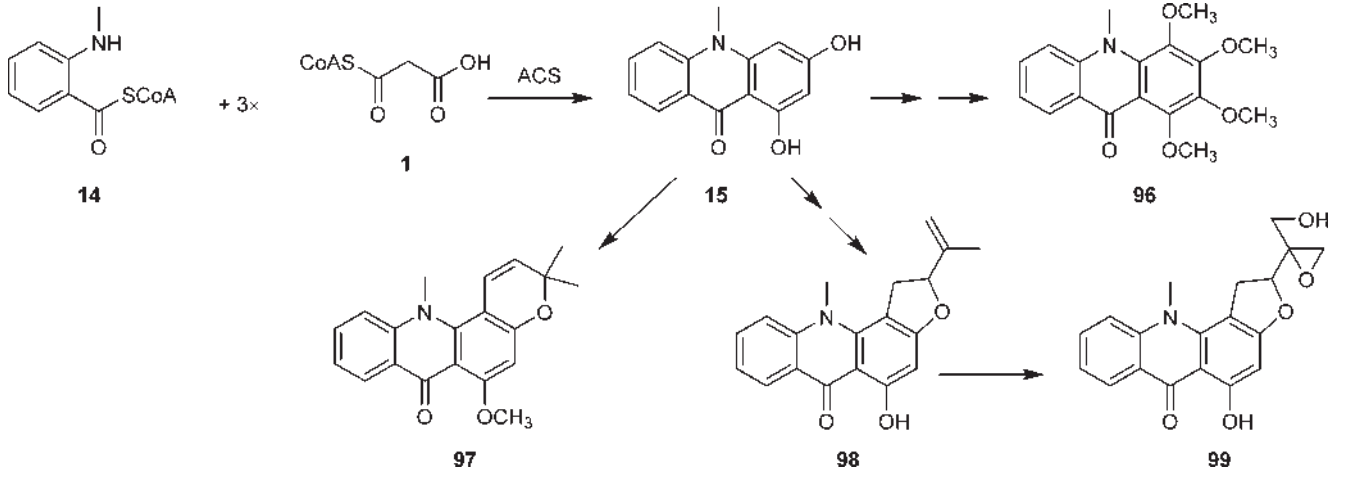
A novel type III PKS (PKS1) has been cloned from a Chinese club moss *Huperzia serrata*.[55] The enzyme catalytic potential is greater than that of prototype CHS. *Huperzia serrata* PKS1 accepts bulky *N*-methylanthraniloyl-CoA (**14**) starter, and carries out three condensations with malonyl-CoA (**1**) to produce 1,3-dihydroxy-*N*-methylacridone (**15**). Interestingly, the chalcone-forming CHS and other type III PKSs, except ACS, usually do not accept the shorter and the bulkier *N*-methylanthraniloyl-CoA (**14**) as a starter substrate in spite of their promiscuous substrate specificity. Although acridone alkaloids have not been isolated from *H. serrata*, this is the first demonstration of the enzymatic production of acridone by a type III PKS from a non-Rutaceae plant. Comprehensive investigations on the intimate structural details of these alkaloid-producing type III PKSs would be helpful to understand the machinery of the biosynthesis of acridones.

1.06.2.5 Phlorisovalerophenone Synthase and Isobutyrophenone Synthase

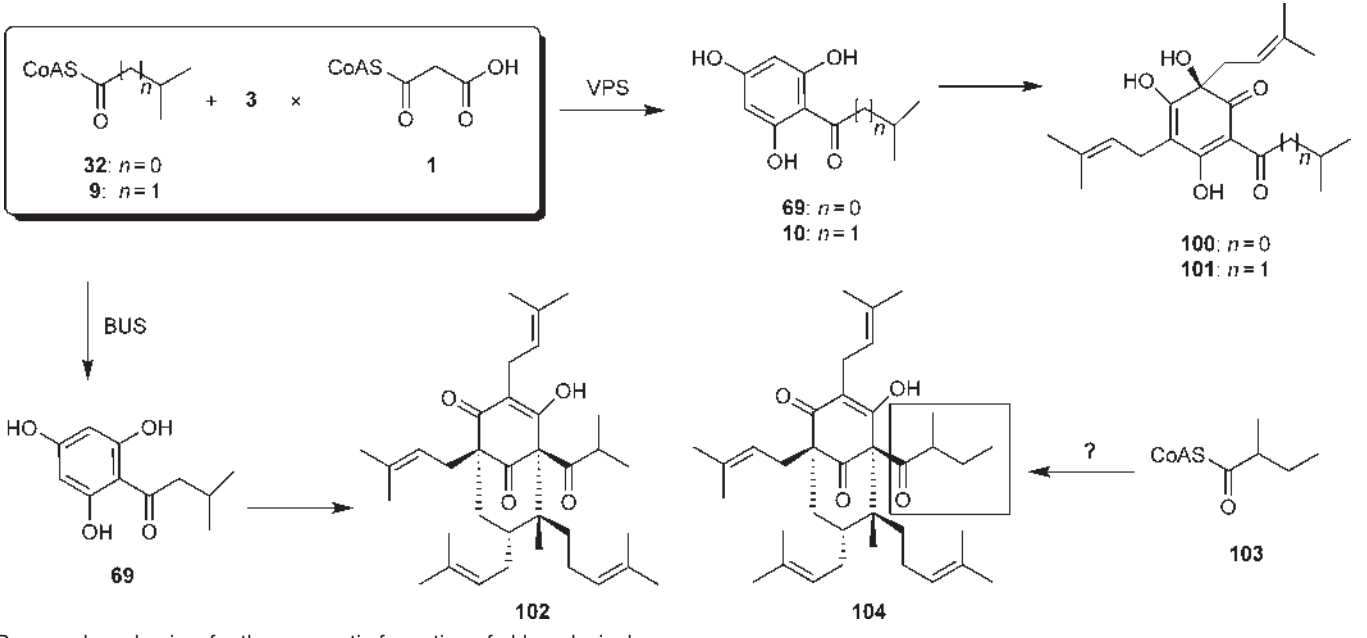
The CHS superfamily of type III PKSs that catalyze the formation of phloroglucinols in plants have also been reported.[30,31] For example, VPS from hop (*Humulus lupulus* L.) selects isovaleryl-CoA (**9**) and isobutyryl-CoA (**39**) as a starter, to produce phlorisovalerophenone (**10**) and phlorisobutyrophenone (**69**), respectively, by condensation of three molecules of malonyl-CoA (**1**) with the Claisen cyclization.[30] The two phloroglucinols are biosynthetic precursors of humlone (**100**) and cohumulone (**101**), which are important components contributing to the bitterness of beer (**Scheme 10**). On the contrary, isobutyrophenone synthase (BUS), which catalyzes the formation of phlorisobutyrophenone (**69**), as in the case of VPS, was also purified from the cell-free extract of *Hypericum calycinum*, suggesting the involvement of the biosynthesis of antidepressant hyperforin (**102**) (**Scheme 10**).[50] In addition, although the detailed mechanism is still unclear, 2-methylbutyryl-CoA (**103**) possibly serves as the adhyperforin (**104**) precursor by a type III PKS, as in the case of hyperforin.

1.06.2.6 Stilbene Synthase

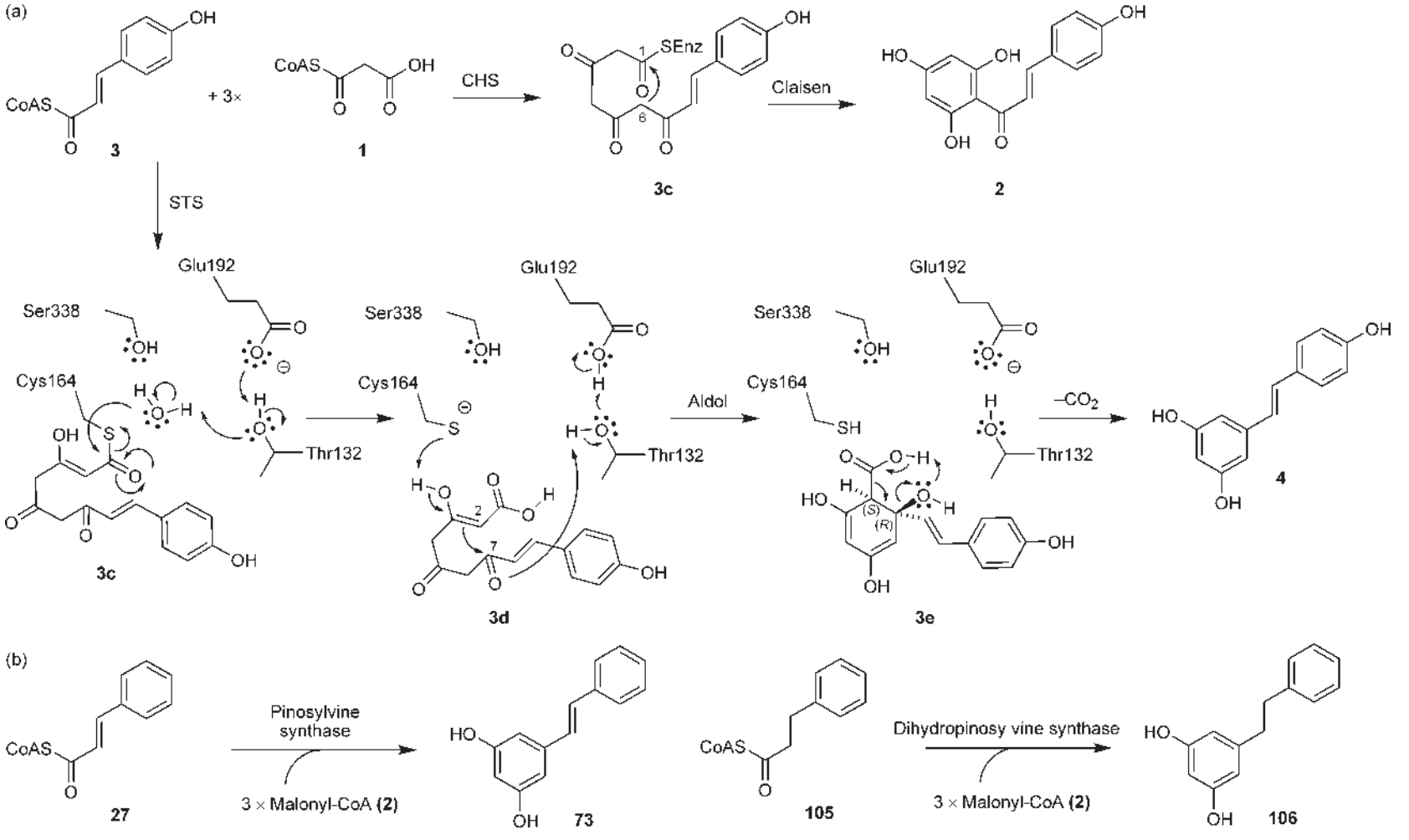
STS is the second identified type III PKS that catalyzes the sequential condensation of 4-coumaroyl-CoA (**3**) with three C_2 units from malonyl-CoA (**1**) to generate resveratrol (**4**) (**Scheme 11a**).[4,5] Although STS and CHS apparently use the same condensation mechanism up to the common tetraketide intermediate, they catalyze different ring-closure reactions. Thus, in STS, the aldol-type cyclization and decarboxylation lead to the formation of resveratrol (**4**), whereas in CHS, the Claisen-type cyclization results in the production of naringenin chalcone (**2**). The X-ray crystal structural analyses of STS led to the proposal that the so-called 'aldol-switch' hydrogen-bonding network plays a critical role in the determination of the stilbene/chalcone product specificity.[20,83] The detailed crystal structure analyses of STS are discussed later. STS is rare in higher plants and occur in distantly related species such as groundnut (*Arachis hypogaea*),[7] grapevine (*Vitis vinifera*),[57] and pines (*Pinus sylvestris* and *P. strobe*).[58,59] In plants, stilbenes and their derivatives are regarded as



Scheme 9 Proposed mechanism for the enzymatic formation of acridone and relevant derivatives.



Scheme 10 Proposed mechanism for the enzymatic formation of phloroglucinols.



Scheme 11 (a) Proposed mechanism for the enzymatic formation of stilbenes based on the crystal structure analyses. (b) Preferred substrates and products of pinosylvin synthase and dihydropinosylvine synthase.

phytoalexins,[84] whereas resveratrol (**4**), a medicinal natural product, is known to be a cancer chemopreventive agent[2] as well as an agonist for the estrogen receptor.[85] Notably, *P. strobe* STS and *P. sylvestris* STS are often known as pinosylvin synthase and dihydropinosylvine synthase, respectively, in their preference for the starter substrate (**Scheme 11b**).[8,58,59]

As in the case of CHS, STS from *A. hypogaea* shows substrate promiscuity to form novel and unusual polyketides from alternative substrates (**Scheme 4**).[86] Three types of products are obtained: (1) complete reaction (stilbene-type) (**77f, 81f, 82f**), (2) three condensations without formation of aromatic ring (CTAL-type) (**77c–82c**), and (3) two condensations without formation of aromatic ring (BNY-type) (**77b–82b**). All product types are obtained from 4-fluorocinnamoyl-CoA (**77a**) and analogs in which the coumaroyl moiety was replaced by furan (**81a**) or thiophene (**82a**). Only type 2 and 3 products were synthesized from other 4-substituted 4-coumaroyl-CoA analogs (-Cl (**78a**), -Br (**79a**), -OCH_3 (**80a**)). Benzoyl-CoA (**11**), phenylacetyl-CoA (**83**), and medium-chain aliphatic-CoA-esters were poor substrates, and the majority of the products was of type 3.[86]

1.06.2.7 Bibenzyl Synthase and Biphenyl Synthase

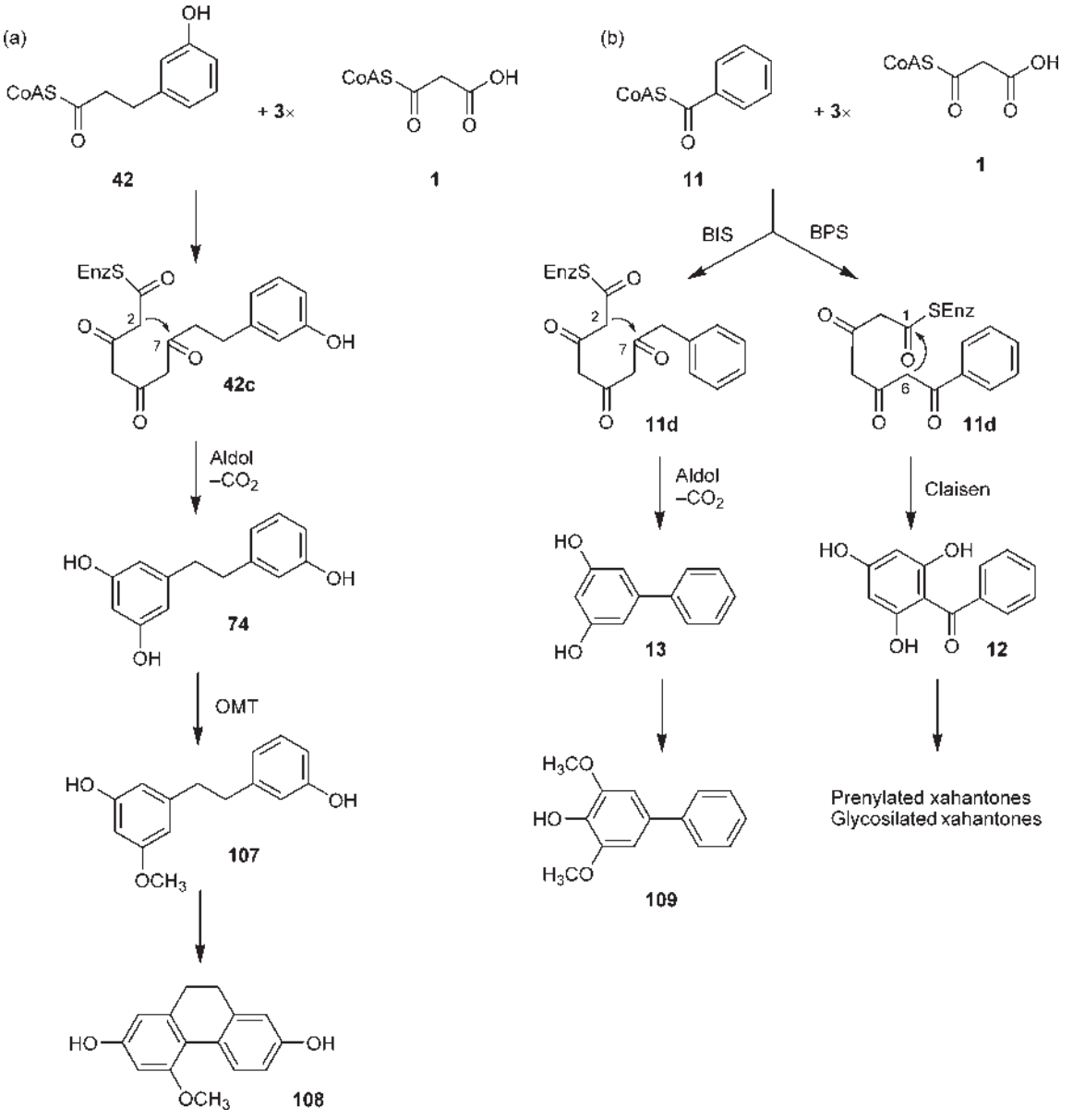
BBS has been cloned from *Phalaenopsis* sp.[34] The enzyme catalyzes STS-like aldol ring closure, but clearly prefers double bond missing cinnamoyl-CoA derivatives such as 3-hydroxyphenylpropionyl-CoA (**42**) as a starter substrate over 4-coumaroyl-CoA (**3**) to yield 3,3′,5-trihydroxybibenzyl (**74**). 3,3′,5-Trihydroxybibenzyl (**74**) is then converted by *O*-methyltransferase (OMT) to prerequisite batatasine III (**107**), which is further metabolized to tricyclic phytoalexine 9,10-dihydrophenanthrene derivatives such as hircinol (**108**) accumulating in stressed or wounded orchid tissues (**Scheme 12a**).[33,34,87]

On the contrary, BIS has been cloned from *Sorbus aucuparia*.[35] Similar to BPS, BIS prefers benzoyl-CoA (**11**) as a starter substrate and catalyzes the condensation of benzoyl-CoA (**11**) with three molecules of malonyl-CoA (**1**) to produce 3,5-dihydroxybiphenyl (**13**), which is a precursor of biphenyl phytoalexins including aucuparin (**109**), isolated from *S. aucuparia* (**Scheme 12b**).[88] Interestingly, as in the case of CHS/STS, BPS and BIS apparently share the same condensation mechanism up to the common tetraketide intermediate; however, in the ring-closure reactions, aldol-type cyclization, instead of Claisen-type cyclization, proceeds in BIS.

1.06.2.8 Coumaroyl Triacetic Acid Lactone Synthase and Stilbenecarboxylate Synthase

Coumaroyl triacetic acid lactone synthase (CTAS) from *Hydrangea macrophylla* var. *thunbergii* produces the CHS by-products pyrone CTAL (**68**) and BNY (**75**), but yields neither chalcones nor stilbenes, including stilbenecarboxylate hydrangenic acid, when 4-coumaroyl-CoA (**3**) was incubated with malonyl-CoA (**1**) (**Scheme 13a**).[49] CTAS has been thought to play a crucial role in the biosynthesis of hydramacroside B (**110**), but evidence for the involvement of this enzyme in the biosynthesis is still lacking.

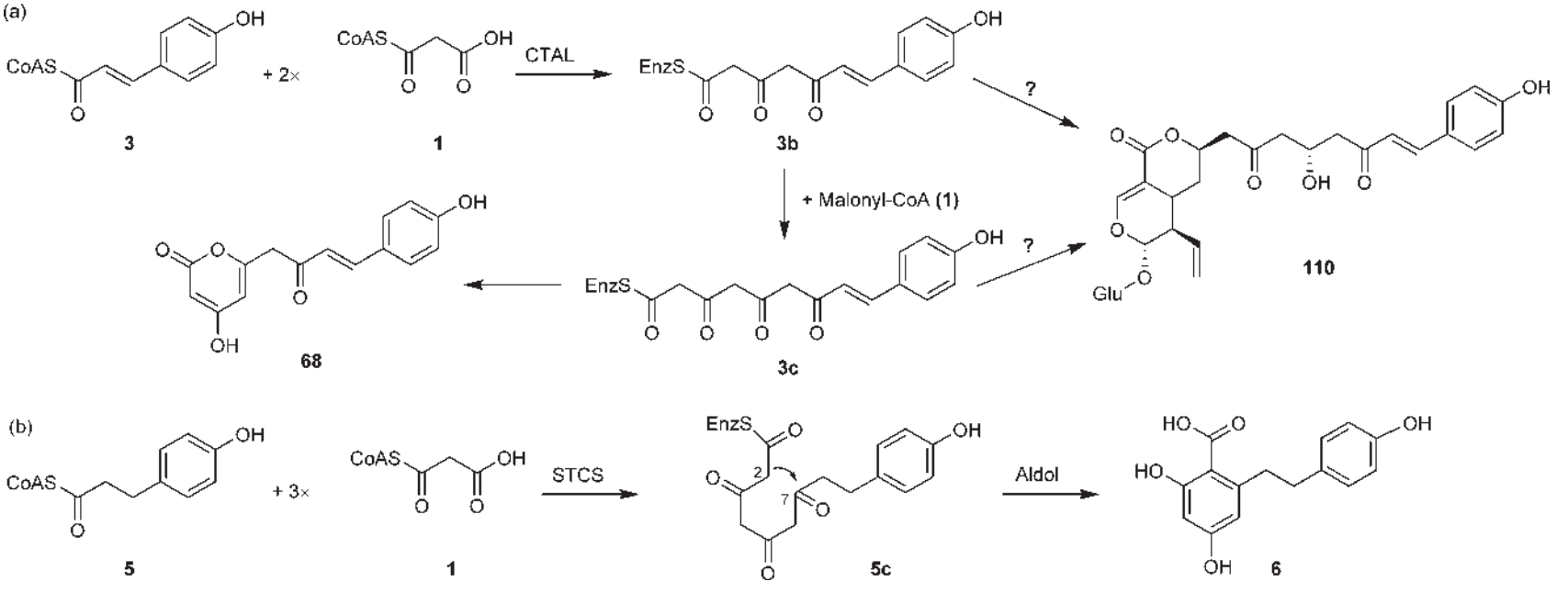
STCS has been cloned from another *Hydranged* variety *H. macrophylla*[36] and *Marchantia polymorpha*. STCS accepts dihydro-4-coumaroyl-CoA (**5**) as a starter and performs three condensation reactions of malonyl-CoA (**1**) (**Scheme 13b**). This is followed by STS-like ring folding, but with retention of terminal carboxyl group that is removed in standard STS-type reaction, and formation of 5-hydroxylunularic acid (**6**). Although aromatic ring formation of stilbenes and that of stilbenecarboxylates have been speculated to occur through the same catalytic machinery of the enzymes, recent structure-based studies have begun to suggest that STCS is structurally close to prototype CHS rather than STS.[20,36] Indeed, when a pine CHS was given dihydro-4-coumaroyl-CoA (**5**) as a starter, a trace amount of 5-hydroxylunularic acid (**6**) was synthesized *in vitro*. Notably, both CTAS and STCS from the *Hydrangea* varieties are now thought to be identical in their function because the enzymes are similar except for five amino acid replacements, and when 4-coumaroyl-CoA is incubated as a starter substrate, STCS produces only BNY and CTAL without the stilbenecarboxylate hydrangenic acid, as in the case of CTAS.[36]



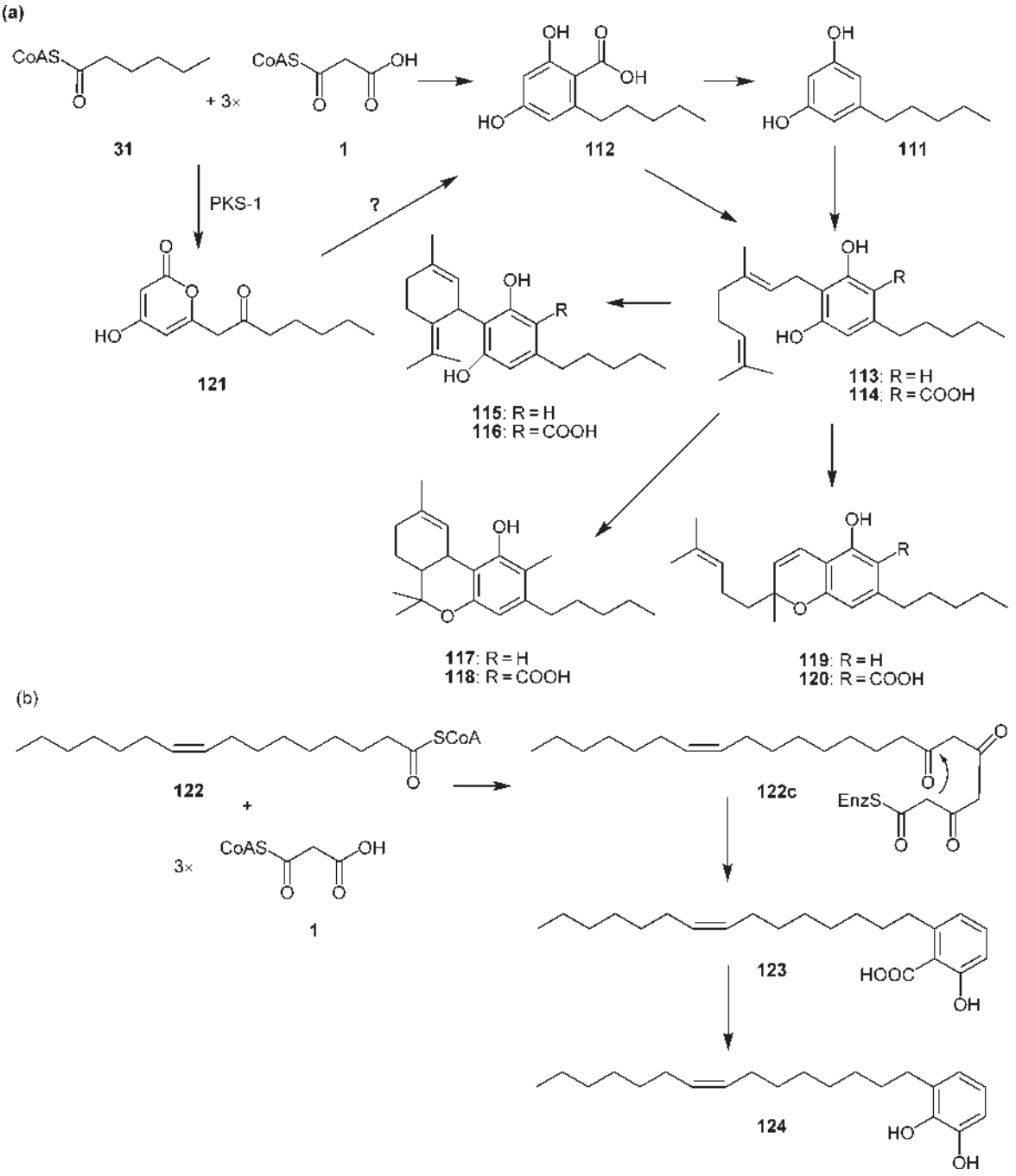
Scheme 12 (a) Proposed mechanism for the enzymatic formation of 3,3′,5-trihydroxybibenzyl (**74**) by BBS and relevant derivatives and (b) Proposed mechanism for the enzymatic formation of 3,5-dihydroxybiphenyl (**13**) by BIS and relevant derivatives.

1.06.2.9 *Cannabis sativa* Polyketide Synthase-1

Olivetol (**111**) and olivetolic acid (**112**) are the precursors of tetrahydrocannabinol (THC), the main psychoactive principle of the cannabis plant (*C. sativa*) (**113–120**).[89] Olivetol (**111**) and olivetolic acid (**112**) are thought to be formed by an STCS-like enzyme through condensation of *n*-hexanoyl-CoA (**31**) starter with three molecules of malonyl-CoA (**1**).[90] The polyketide synthase-1 (PS-1) has been cloned from *C. sativa*[91] and reported in databases (NCBI: http://www.ncbi.nlm.gov/) as olivetol synthase, though the enzyme catalyzes the formation of hexanoyl triacetic acid lactone (**121**) from *n*-hexanoyl-CoA (**31**) and malonyl-CoA (**1**), without generating olivetol and olivetolic acid. As the recent result showed that hexanoyl pyrone (**121**) is likely to be the biosynthetic isoform of



Scheme 13 (a) Proposed mechanism for the enzymatic formation of CTAL (**68**) and (b) proposed mechanism for the enzymatic formation of 5-hydroxylunularic acid (**6**).



Scheme 14 Proposed mechanism for the enzymatic formation of (a) olivetol (**112**), olivetolic acid (**111**), and tetrahydrocannabinol (THC) and (b) ginkgolic acid (**123**) and urushiol (**124**).

olivetolic acid (**112**), the PS-1 is probably involved in the biosynthesis of THC (**Scheme 14a**). However, there remains the possibility of as-yet isolated enzyme in charge of THC biosynthesis.

In addition, it has been postulated that CoA esters of long-chain fatty acids such as palmitoleoyl (C_{16})-CoA (**122**) are accepted as a starter substrate for malonyl-CoA chain extension, which leads to the formation of alkyl polyphenols including ginkgolic acid (anacardic acid) (**123**) and urushiol (**124**), the allergic substances of ginkgo tree (*Ginkgo biloba*) and lacquer tree (*Rhus verniciflua*), respectively (**Scheme 14b**).[92] The discovery of ORAS from the fungus *Neurospora*

crassa, which catalyzes the formation of pentaketide resorcinols with different aliphatic chain length (C_4–C_{20}) by condensation of aliphatic-CoA ester with four molecules of malonyl-CoA (**1**),[16] strongly suggests that the plant alkyl polyphenols are also formed by type III CHS-superfamily type III PKSs. In fact, as mentioned above, the regular CHS also accepts CoA thioesters of long-chain fatty acids (C_4–C_{20}) as a starter substrate and carries out sequential condensations with malonyl-CoA (**1**) to produce a series of triketide and tetraketide lactones.

1.06.2.10 Benzalacetone Synthase

Many of type III PKSs catalyze iterative condensations of malonyl-CoA (**1**) with a starter substrate, whereas BAS catalyzes the one-step decarboxylative condensation of malonyl-CoA (**1**) with 4-coumaroyl-CoA (**3**) to produce 4-hydroxybenzalacetone 4-(4-hydroxyphenyl)-but-3-en-2-one (**16**). BAS plays a crucial role in the biosynthesis of the C_6–C_4 moiety of a variety of pharmaceutically important phenylbutanoids, such as anti-inflammatory glucoside lindleyin (**125**) in the medicinal plant rhubarb (*Rheum palmatum*)[93,94] and the characteristic aroma raspberry ketone (**126**) of raspberry (*Rubus idaeus*) (**Scheme 15a**).[43,95] Indeed, BAS has been cloned from rhubarb[40] and raspberry,[41] although raspberry BAS is identified as a bifunctional enzyme producing both 4-hydroxybenzalacetone (**16**) and naringenin chalcone (**2**).

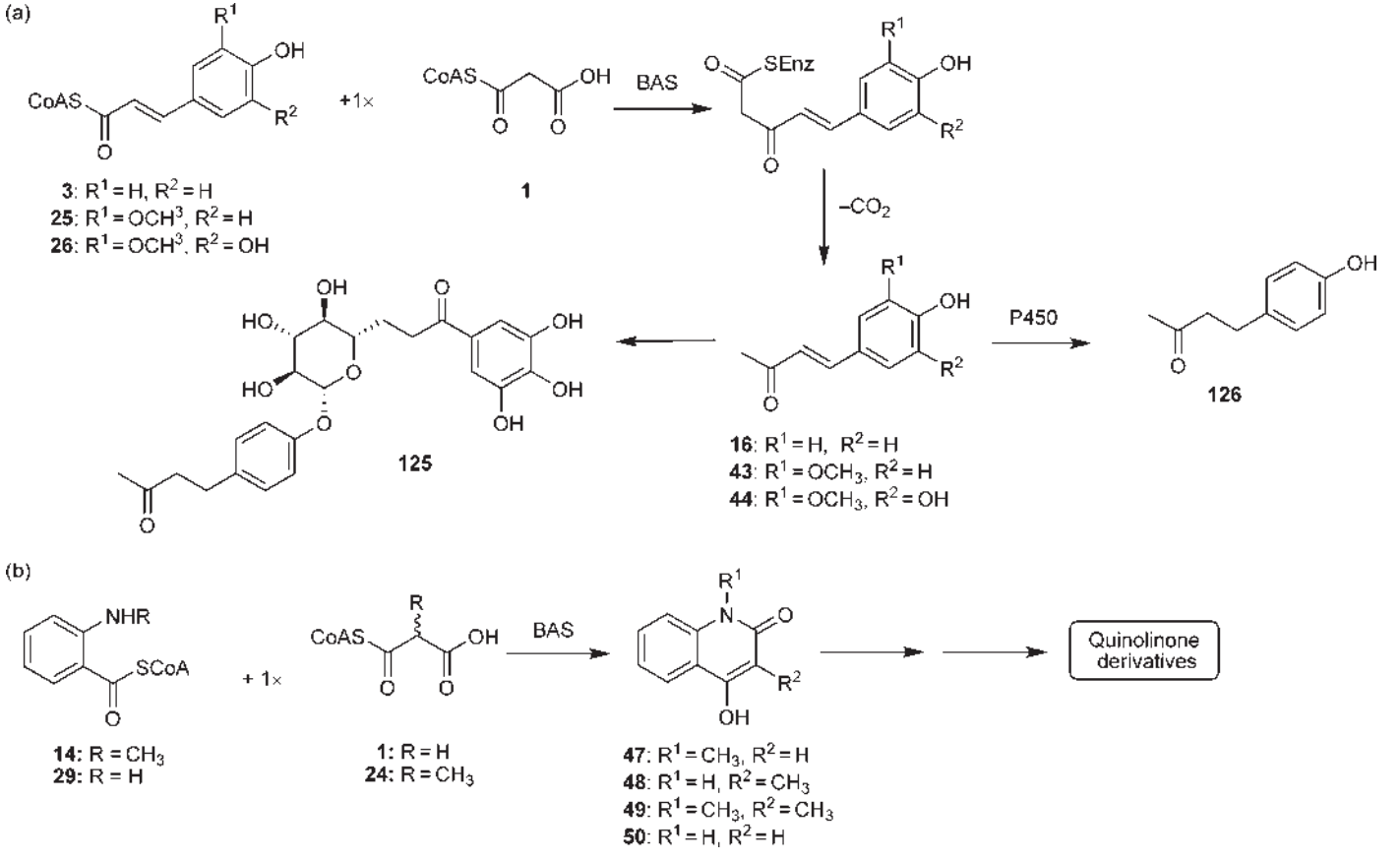
As in the case of CHS, BAS shows a broad substrate specificity, leading to the formation of benzalacetone analogs. For example, raspberry BAS prefers *in vitro* feruloyl-CoA (**25**) and 5-hydroxyferuloyl-CoA (**26**) over normal 4-coumaroyl-CoA (**3**), and yields 3-methoxy-4-hydroxybenzalacetone (**43**) and 3-methoxy-4,5-dihydroxybenzalacetone (**44**), respectively, without producing the corresponding chalcone (**Scheme 15a**).[41] BAS from *R. palmatum* efficiently catalyzes condensation with one molecule of malonyl-CoA (**1**) (or methylmalonyl-CoA (**24**)) by accepting a much bulkier starter unit such as *N*-methylanthraniloyl-CoA (**14**) (or anthraniloyl-CoA (**29**)) to produce 4-hydroxy-2(1*H*)-quinolones (**47–50**) (**Scheme 15b**).[46] Notably, 4-hydroxy-2(1*H*)-quinolone is a biosynthetic precursor of quinolone alkaloid occurring in greatest abundance in plants from the family of Rutaceae,[82] but not found in rhubarb. The enzyme in Rutaceae responsible for the formation of quinolone alkaloid is still missing.

1.06.2.11 Curcuminoid Synthase

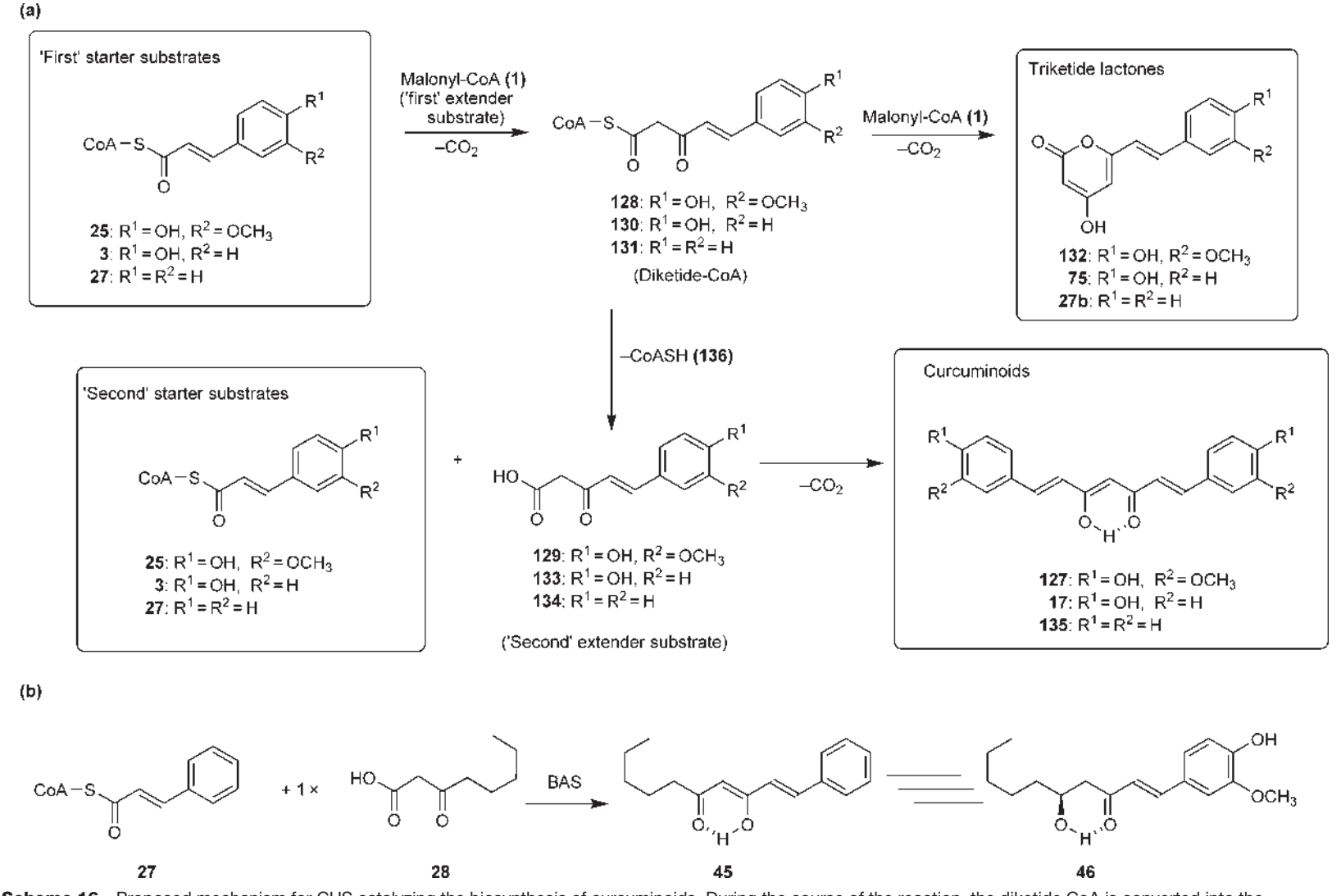
Curcuminoids, the major components of turmeric, are widely used as a traditional medicine and as food additive for their unique aromatic and coloring properties. Curcumin (**127**) has received a great deal of attention because of its anti-inflammatory, anticarcinogenic, and antitumor activities.[96–98] It has been proposed that curcuminoid scaffold is synthesized by a three-step reaction from phenylpropanoids (**Scheme 16**).[99] First, malonyl-CoA (**1**) condenses with feruloyl-CoA (**25**) to produce a diketide-CoA (**128**). The diketide-CoA is converted into β-keto acid (**129**) by hydrolysis, which finally condenses with another molecule of feruloyl-CoA (**25**) to produce curcumin (**127**). This hypothesis was not confirmed until a novel enzyme termed CUS was reported by Horinouchi and coworkers.[42] CUS from *Oryza sativa* catalyzes the condensation of two molecules of 4-coumaroyl-CoA (**3**) with one molecule of malonyl-CoA (**1**) to produce bisdemethoxycurcumin (**17**). Interestingly, the mechanism of the formation of curcuminoids is very different from the traditional head-to-tail model of polyketide assembly. During the reaction, 4-coumaroyl-CoA (**3**) also serves as a second extender to the diketide-CoA intermediate (**130**), leading to the formation of bisdemethoxycurcumin (**17**). Interestingly, CUS forms gingerol (**46**) analogs, such as cinnamoyl (hexanoyl) methane (**45**), a putative intermediate of gingerol, from cinnamoyl-CoA (**27**) and 3-oxo-octanoic acid (**28**).[100]

1.06.2.12 2-Pyrone Synthase

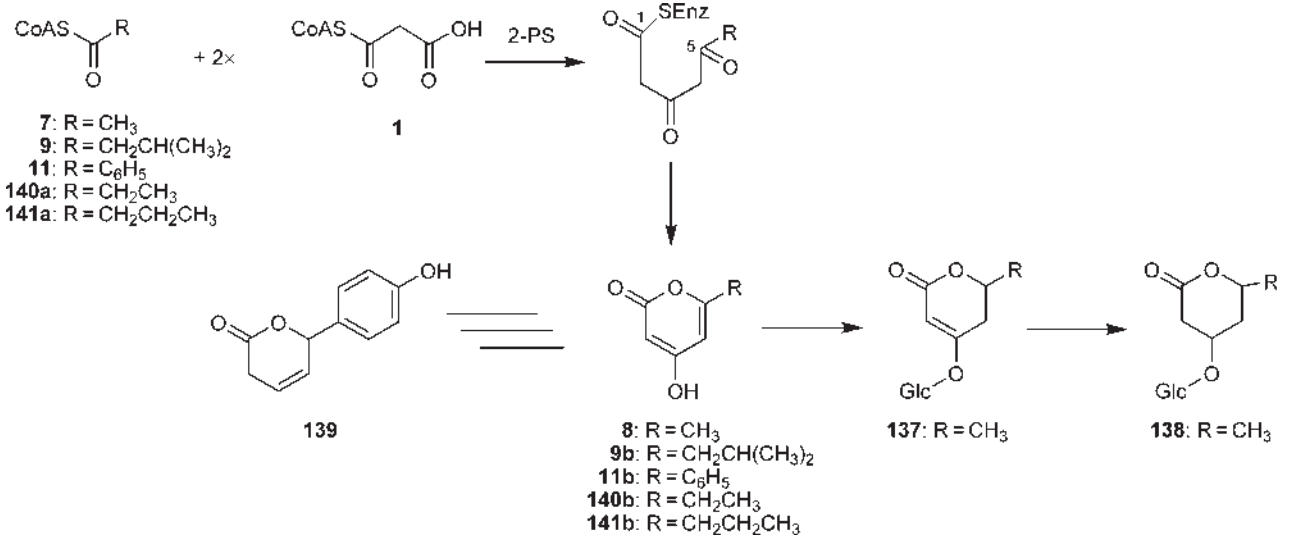
Pyrone glycosides and their derivatives showing antifungal and anti-insectant activities have been isolated from various plant species, and in *Gerbera hybrida*, 2-PS has been shown to be a related type III PKS.[13,14] Unlike CHS, 2-PS efficiently uses acetyl-CoA (**7**) as a starter substrate instead of 4-coumaroyl-CoA (**3**) and performs two decarboxylative condensation reactions with malonyl-CoA (**1**) to produce 6-methyl-4-hydroxy-2-pyrone (triacetic acid lactone (TAL) (**8**)) (**Scheme 17**), which is a biosynthetic precursor of pyrone glycoside gerberin (**137**) and parasorboside (**138**). If acetyl-CoA (**7**) is not available, the enzyme can decarboxylate malonyl-CoA (**1**) to generate its own acetyl starter. The X-ray crystal structure is now available.[18] Interestingly, 2-PS also



Scheme 15 Proposed mechanism for the enzymatic formation of (a) phenylbutanoids and relative derivatives and (b) quinolinones (**47–50**) by *R. palmatum* BAS.



Scheme 16 Proposed mechanism for CUS catalyzing the biosynthesis of curcuminoids. During the course of the reaction, the diketide CoA is converted into the corresponding β-keto acid by hydrolysis, which is then condensed with another molecule of aromatic CoA to produce curcuminoids.



Scheme 17 Formation of triketide pyrones by 2-PS and relative derivatives.

accepts benzoyl-CoA (**11**) as a starter substrate with two molecules of malonyl-CoA (**1**) to produce 6-phenyl-4-hydroxy-2-pyrone (phenylpyrone) (**11b**), corresponding to the backbone in the psilotins (**139**) of the Psilotaceae.[14] Additionally, other hydrophobic short-chain aliphatic-CoA esters such as isovaleryl-CoA (**9**), propionyl-CoA (**140a**), and butyryl-CoA (**141a**) are also utilized as a starter substrate to produce related triketide pyrones (**Scheme 17**).

However, the pathway of heterocyclic ring closure is unclear because two lactonization pathways are possible. One possibility is formation by a nucleophilic attack of C5 keto-enol oxyanion to C1 carbonyl carbon on the enzyme-bound intermediate. Another possibility is formation by a nucleophilic attack of carboxylate anion to C5 carbonyl carbon on the linear intermediate through the cleavage of the enzyme-bound intermediate by a water molecule. The latter case will occur easily in solution.

1.06.2.13 Long-chain 2-Pyrone Synthase

Two type III PKSs, PKS-A and PKS-B, that produce 4-hydroxy-2-pyrone derivatives have been characterized in *Arabidopsis thaliana*.[48] The enzymes efficiently accept long-chain fatty acyl-CoAs of C_6 (*n*-hexanoyl (**31**)) to C_{20} (*n*-eicosanoyl (**38**)) chain length as a starter substrate with almost equal efficiency and carry out sequential condensations with two or three molecules of malonyl-CoA (**1**) to produce triketide (**52–59**) and tetraketide α-pyrones (**60–67**) as the major products in addition to a trace amount of phloroglucinols (**142–149**) (**Scheme 18**). Interestingly, the enzymatic ability of the two enzymes is drastically different from that of 2-PS, but very similar to bacterial type III PKS *Mycobacterium tuberculosis* PKS18, which is thought to play the crucial role of lipid metabolite.[28] Additionally, it has been reported that several novel type III PKSs are also involved in the biosyntheses of long-chain phenolic lipids in bacteria.[16,101,102] Although the biosynthesis of α-pyrone lipids in plants has not been studied well and their presence in *A. thaliana* has not been reported so far, these observations might suggest the involvement of type III PKSs in lipid metabolism in the model plant *A. thaliana*, as in the case of Mycobacterium.

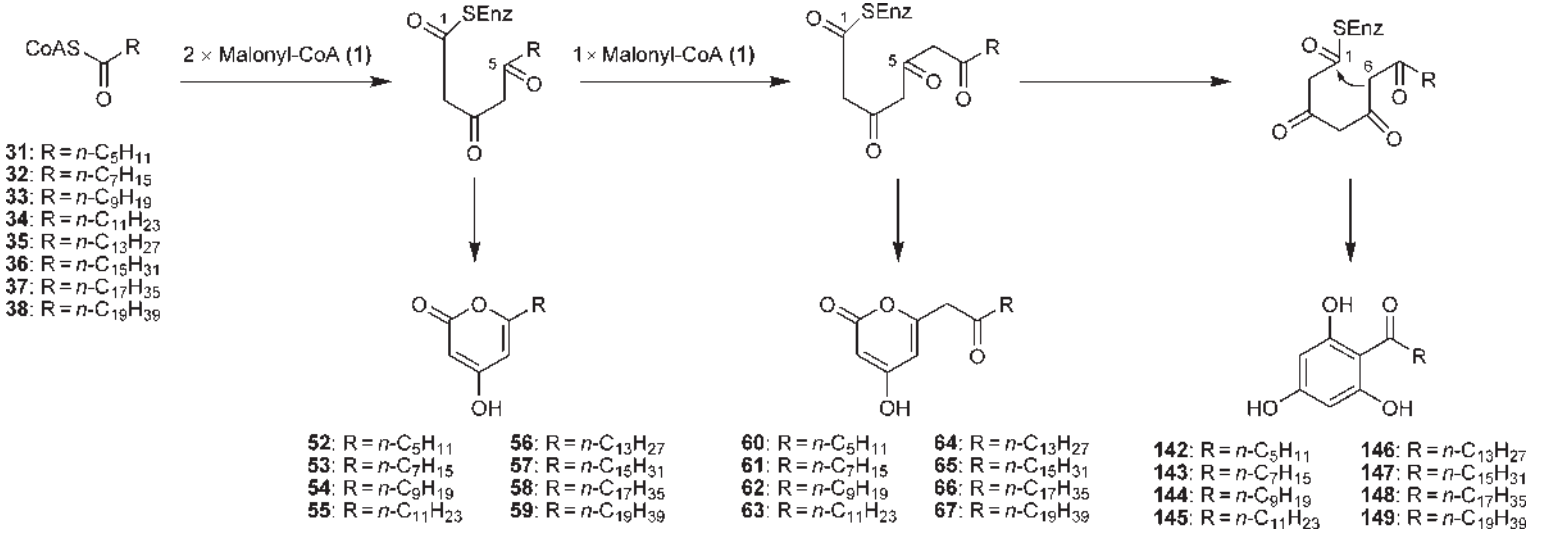
1.06.2.14 Pentaketide Chromone Synthase

PCS from Aloe (*Aloe arborescens*) catalyzes the formation of a pentaketide chromone, 5,7-dihydroxy-2-methylchrome (**18**), from five-step decarboxylative condensations of malonyl-CoA (**1**) followed by the Claisen cyclization reaction to form an aromatic ring (**Scheme 19**).[32,60] PCS also accepts acetyl-CoA (**7**), resulting from decarboxylation of malonyl-CoA (**1**), as a starter substrate, but it is a poor substrate for PCS. The pentaketide chromone (**18**) has been isolated from several plants,[92] and is known to be the biosynthetic precursor of the antiasthmatic furochromones visnagin (**150**) and kehellin (**151**) found in *Ammi visnaga*.[92] The X-ray crystal structure complexed with CoA-SH is now available.[21,103] However, it is unclear whether the heterocyclic ring closure of the pentaketide chromone is enzymatic or not, because the ring closure can take place due to spontaneous Michael-like ring closure, as in the case of flavanone formation from chalcone *in vitro*.

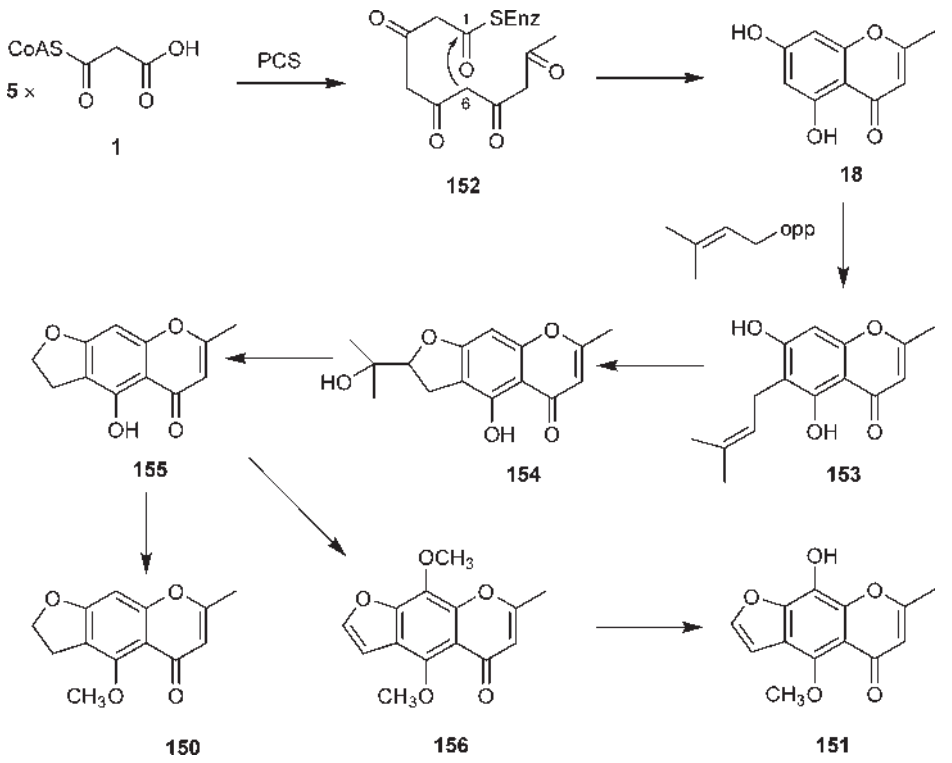
As in the case of prototype CHS, the enzyme also accepts aromatic (4-coumaroyl (**3**), cinnamoyl (**27**), and benzoyl (**11**)) and middle-chain aliphatic (*n*-hexanoyl (**31**), *n*-octanoyl (**32**), and *n*-decanoyl (**33**)) CoA esters as a starter substrate, but it yields only triketide and tetraketide α-pyrones.[32]

1.06.2.15 Aloesone Synthase

ALS is a key enzyme in the biosynthesis of heptaketide chromone aloesone derivatives such as aloesone *O*-glucoside (7-*O*-β-D-glucopyranoside) in rhubarb (*R. palmatum*)[104] and anti-inflammatory aloesone *C*-glucoside (8-*C*-β-D-glucopyranoside) (aloesin) (**157**) in Aloe (*A. arborescens*).[105] ALS from rhubarb and Aloe efficiently catalyzes the formation of a heptaketide aromatic pyrone 6-(2-(2,4-dihydroxy-6-methylphenyl)-2-oxoethyl)-4-hydroxy-2-pyronechromone (**20**) from acetyl-CoA (**7**) and six molecules of malonyl-CoA (**1**) through an aldol cyclization with a trace amount of hexaketide pyrones 6-(2′,4′-dihydroxy-6′-methylphenyl)-4-hydroxy-2-pyrone (**19**) and Tw93a (**158**) and octaketide pyrones SEK4 (**22**) and SEK4b (**23**) (**Scheme 20**).[37,96] The unstable heptaketide pyrone (**20**) (or acid form) would be then released from the active site and undergo subsequent spontaneous isomerization to the β-ketoacid chromone (**159**), which is followed



Scheme 18 Enzymatic formation of triketide and tetraketide α-pyrones and phloroglucinols by *Arabidopsis thaliana* PKS-A and PKS-B.



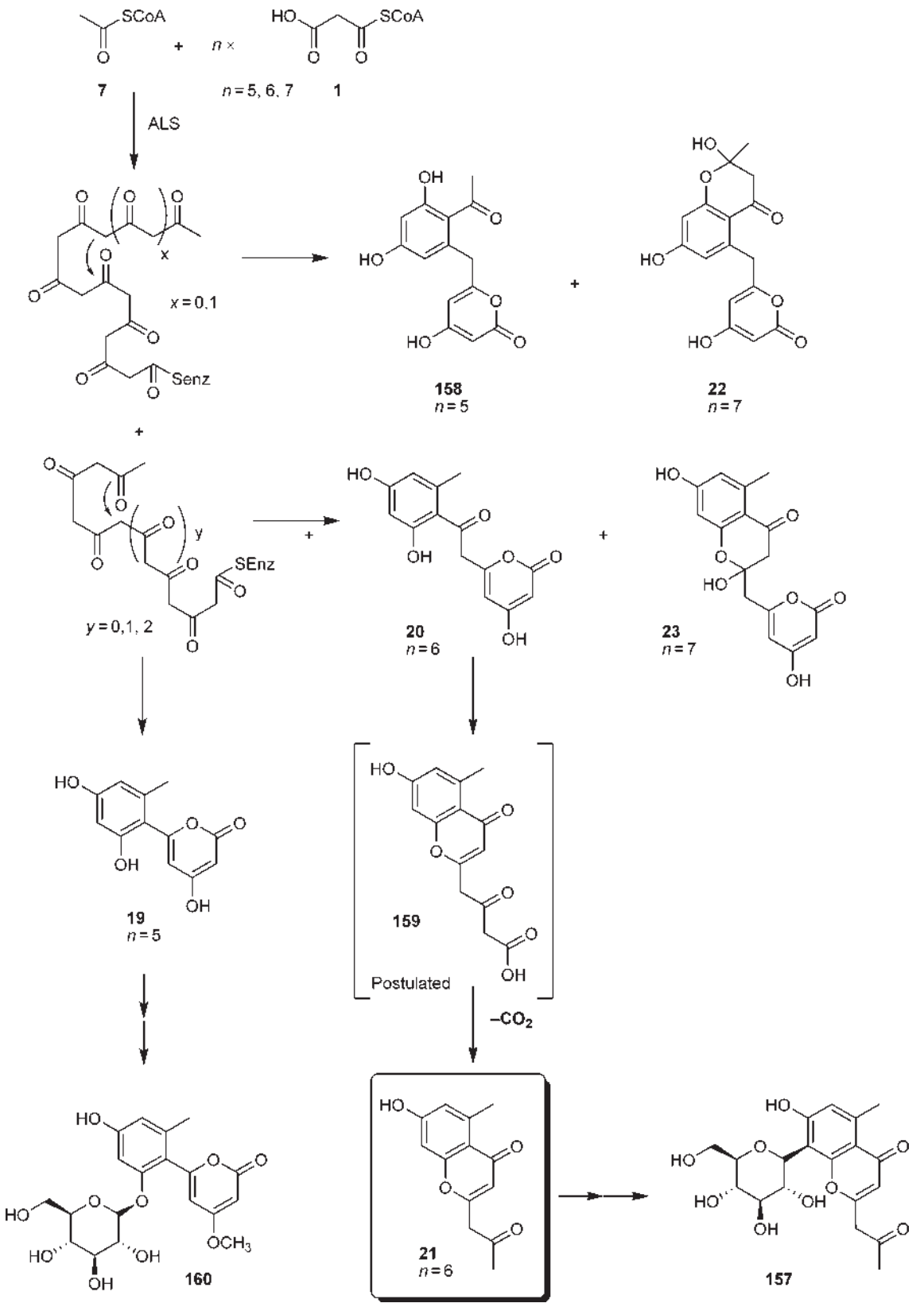
Scheme 19 Proposed mechanisms for the enzymatic formation of 5,7-dihydroxy-2-methylchromone (**18**) and its derivatives.

by decarboxylation to produce the heptaketide aloesone (**21**). In contrast, two hexaketide pyrones (**19, 158**) from iterative condensations of six acetate units have been known to be a demethylated aglycone of antihistamic *O*-glucoside aloenin (**160**) occurring only in *A. arborescens* and a product limited in *in vitro* reaction of the type II minimal PKS of *Streptomyces coelicolor*, respectively.[106,107] ALS does not accept 4-coumaroyl-CoA (**3**) or other aromatic-CoA esters, whereas aliphatic-CoA esters of medium-chain length are accepted as a starter substrate, but with less efficiency.

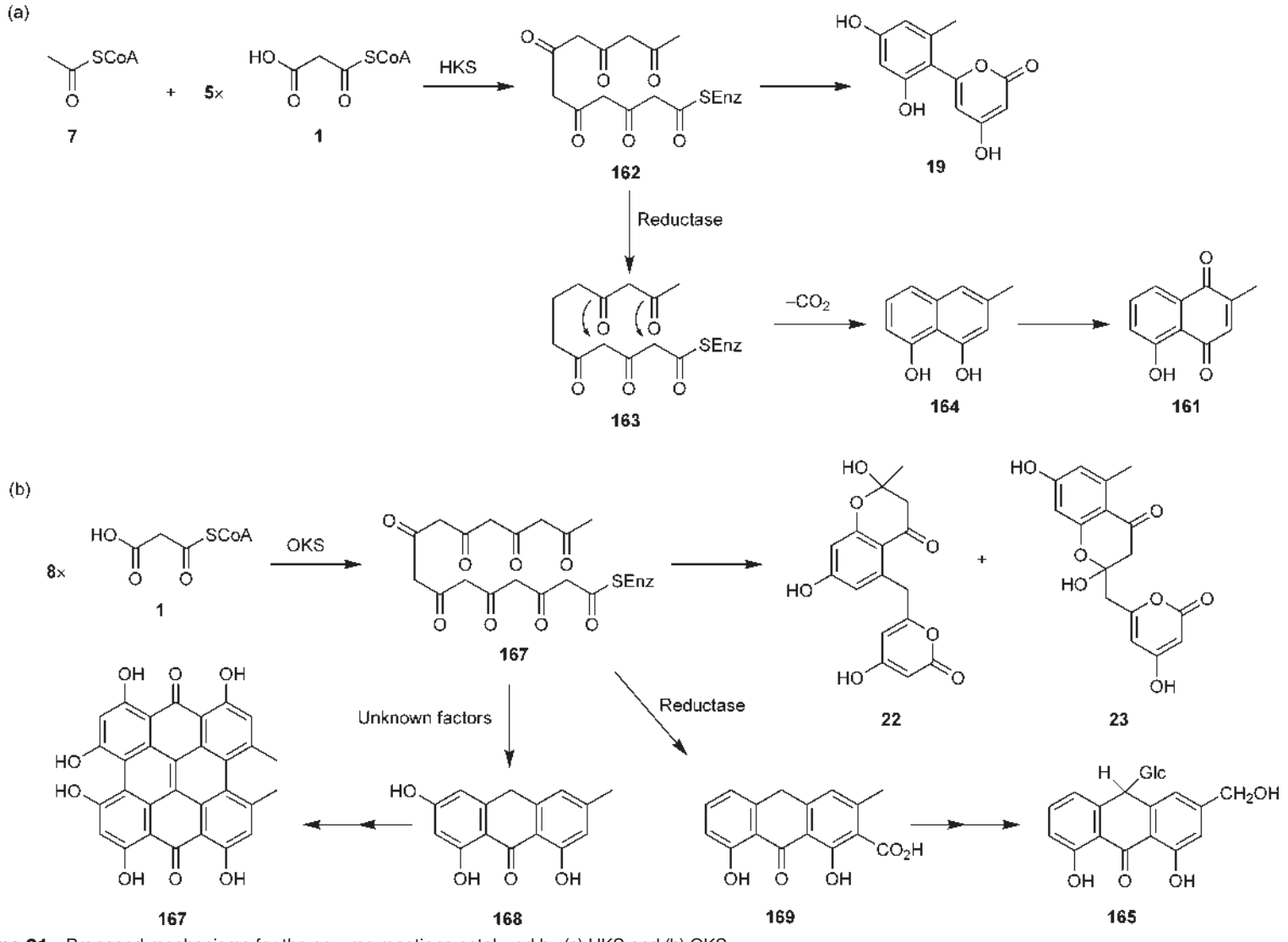
1.06.2.16 Hexaketide Synthase and Octaketide Synthase

Hexaketide synthase (HKS) efficiently produces hexaketide 6-(2′,4′-dihydroxy-6′-methylphenyl)-4-hydroxy-2-pyrone (**19**) of minor product of ALS, by catalyzing iterative condensations of five molecules of malonyl-CoA (**1**) with acetyl-CoA (**7**) as a starter substrate (**Scheme 21a**). Two HKSs have been cloned from *Plumbago indica* and *Drosophyllum lusitanicum*, although aloenin (**160**) and its derivatives have not been described as a constituent of plants.[61,62] In contrast, plants accumulate hexaketide naphthoquinones such as plumbagin (**161**), which is derived from six acetate units.[108,109] As in the case of other type III PKSs, HKSs also show broad substrate and product specificities in *in vitro* assays, but the acetogenic hexaketide naphthoquinone was not synthesized. It is therefore tempting to speculate that the enzyme is originally involved in the biosyntheses of plumbagin (**161**) and tetralones in the presence of the tailoring enzymes such as a cyclase-like cofactor and/or a reductase in plants.[61,62]

Interestingly, a similar case has also been found in the metabolite ability of OKS from Aloe and *Hypericum perforatum*.[38,39,60] The enzyme forms aromatic octaketides SEK4 (**22**) and SEK4b (**23**) constructed with the longest polyketide chain yielded by the structurally simple type III PKS, from eight-step decarboxylative condensations of malonyl-CoA (**1**) followed by aldol cyclization (**Scheme 21b**). As in the case of PCS, acetyl-CoA (**7**) is also available as a poor starter substrate. Notably, the octaketides are the products of the minimal



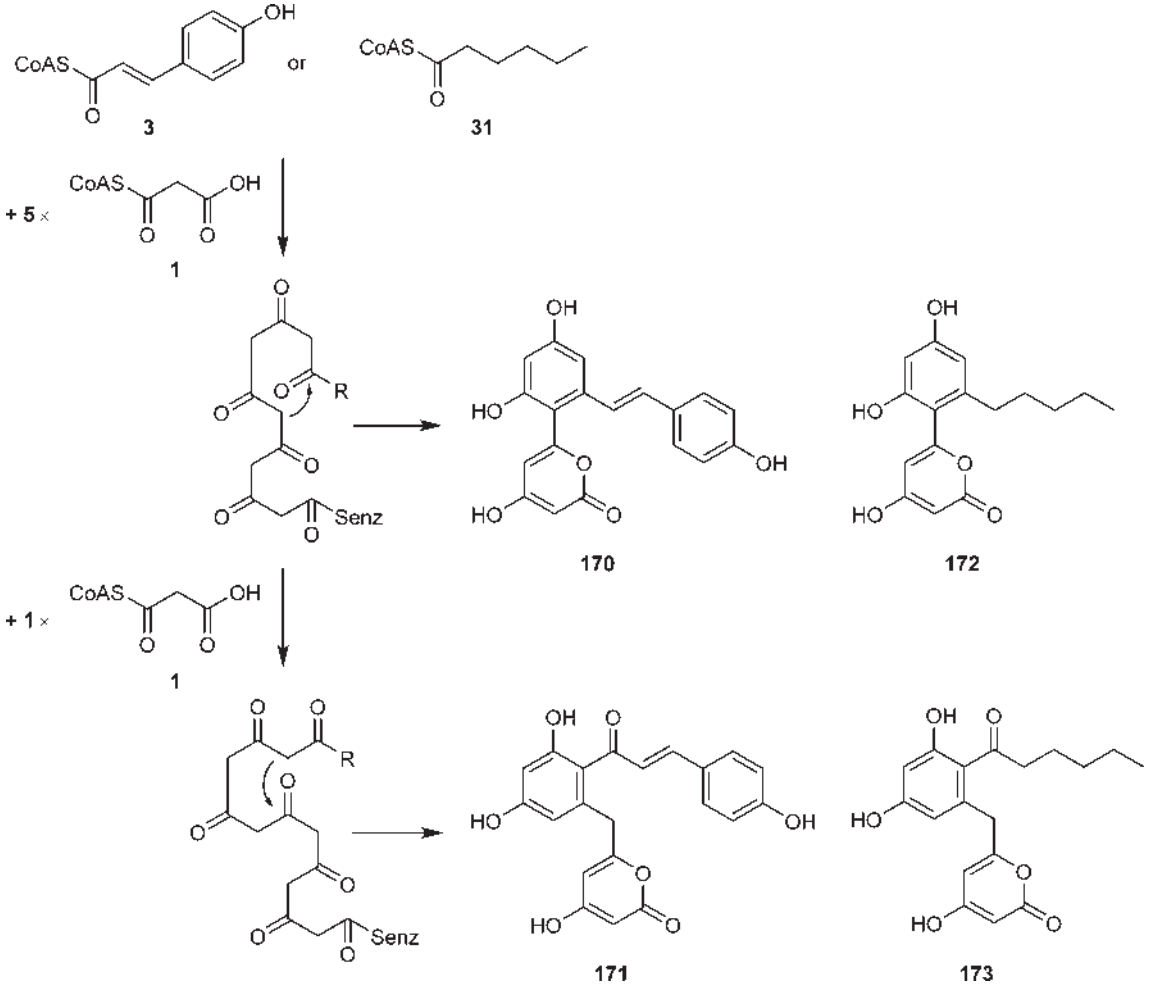
Scheme 20 Proposed mechanisms for the enzyme reactions catalyzed by ALS.



Scheme 21 Proposed mechanisms for the enzyme reactions catalyzed by (a) HKS and (b) OKS.

type II PKS for the benzoisochromanequinone actinorhodin (*act* from *S. coelicolor*),[110] and have not been isolated from any of the known plants. However, Aloe plant has been shown to be rich in pharmaceutically important anthrone glycoside, barbaroin (**165**).[60,111] The medicinal plant *H. perforatum* has been reported to accumulate pharmaceutically and physiologically important red-pigmented naphthodianthrones including hypericin (**167**),[112,113] the precursor of which has been proposed to be emodine anthrone (**168**), an anthraquinone.[39] The failure of the OKS enzymes to produce anthrone suggests the requirement of tailoring enzymes or as-yet unknown factors in the biosynthesis of octaketide anthrones/anthraquinones in the medicinal plant, as in the case of naphthoquinone biosynthesis.

OKS is a biologically quite mysterious plant type III PKS; however, it is interesting from the point of view of engineered biosynthesis of pharmaceutically important plant polyketides. In particular, when 4-coumaroyl-CoA (**3**) and malonyl-CoA (**1**) are incubated with *A. arborescens* OKS, the enzyme accepts 4-coumaroyl-CoA (**3**) less efficiently and produces unnatural hexaketide stilbene (**170**) and heptaketide chalcone (**171**) by catalyzing five- and six-round condensations of malonyl-CoA (**1**), respectively (**Scheme 22**).[114] Furthermore, the enzyme can efficiently synthesize the formation of unnatural hexaketide resorcinol (**172**) and heptaketide phloroglucinol (**173**) from *n*-hexanoyl-CoA (**31**) as a starter substrate and five and six molecules of malonyl-CoA (**1**), respectively.[114] Although these compounds have not been isolated from natural source, the hexaketide resorcinol (**172**) has been previously proposed as one of the possible reaction products of the minimal type II PKS from hexanoyl-ACP/malonyl-CoA.[115]



Scheme 22 Enzymatic synthesis of unnatural polyketides by *A. arborescens* OKS from 4-coumaroyl-CoA (**3**) and *n*-hexanoyl-CoA (**31**) as starters.

1.06.2.17 *C*-Methylchalcone Synthase

None of the polyketides isolated from plants contain methyl-extended moieties, which suggests that type III PKSs also utilize methylmalonyl-CoA (**24**) as a substrate for the C–C bond elongation reaction. Evidence favoring the selection of methylmalonyl-CoA (**24**) as a starter substrate has been obtained for CHS2 from *P. strobes*.[47] This enzyme was reported to be completely inactive with malonyl-CoA (**1**) as a substrate; however, it catalyzed a one-step condensation of (2*RS*)-methylmalonyl-CoA (**24**) with cinnamoyl diketide NAC (**30**) to produce methylated triketide styrylpyrone (**51**) (**Scheme 23**).

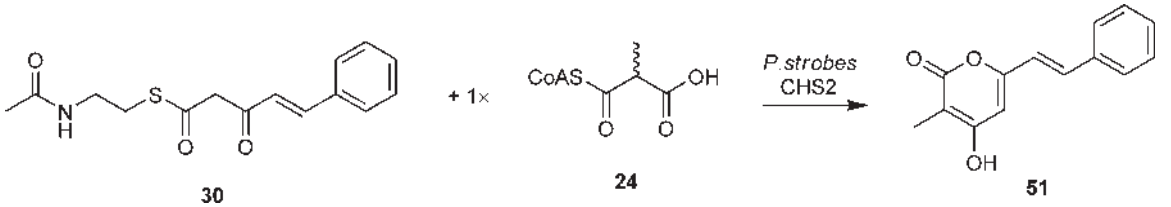
However, a few plant PKSs (*S. baicalensis* CHS, *A. hypogaea* STS, and *R. palmatum* BAS) have been known to produce an unnatural C_6-C_5 aromatic polyketide, 1-(4-hydroxyphenyl)pent-1-en-3-one (**174**), which was formed by one-step decarboxylative condensation of (2*RS*)-methylmalonyl-CoA (**24**) and 4-coumaroyl-CoA (**3**) (**Scheme 24a**).[116] Furthermore, these enzymes also produced unnatural polyketides when both the starter and extender substrates were simultaneously replaced with nonphysiological substrate analogs. Thus, *S. baicalensis* CHS afforded an unnatural novel triketide (**177**) as well as a tetraketide (**178**), when incubated with benzoyl-CoA (**11**) and (2*RS*)-methylmalonyl-CoA (**24**) as substrates. Moreover, the enzyme also accepted *n*-hexanoyl-CoA (**31**) and methylmalonyl-CoA(**24**) as substrates to produce an unnatural novel triketide (**179**) (**Scheme 24b,c**).[117]

On the contrary, it has been demonstrated that *A. arborescens* OKS efficiently accepted (2*RS*)-methylmalonyl-CoA (**24**) as a sole substrate to produce 6-ethyl-4-hydroxy-3,5-dimethyl-2-pyrone (**180**) (**Scheme 24d**).[115] Interestingly, *S. baicalensis* CHS and *R. palmatum* BAS also yielded the unnatural methylated C9 triketide lactone (**180**) as a sole product by sequential decarboxylative condensations of three molecules of (2*RS*)-methylmalonyl-CoA (**24**).[118]

1.06.3 Enzyme Structure and Site-Directed Mutagenesis

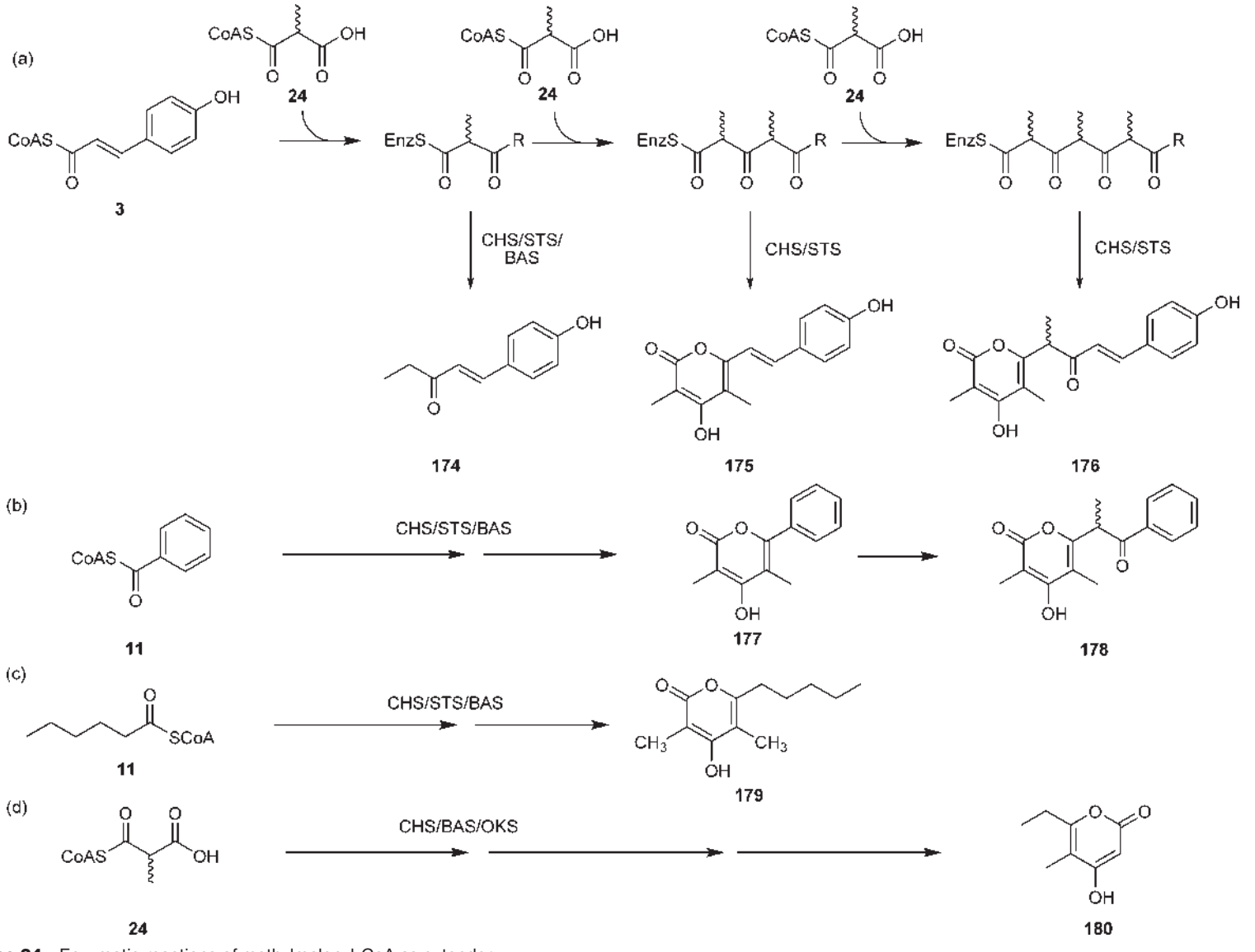
The crystal structures of functionally diverse type III PKSs from plants and bacteria have been characterized as listed in **Table 2**. There is no significant difference in the overall fold and the common catalytic machinery. Recent crystallographic and site-directed mutagenesis studies have begun to reveal that only a small modification of the active site has led to the functional diversity of type III PKSs (**Figure 3**).[4,17–25,83]

In 1999, the X-ray crystal structure of *Medicago sativa* CHS2 at 1.56 Å resolution by Noel and coworkers was solved and it revealed the common overall fold of the type III PKSs (**Figure 4a, b**).[17] Each monomer of *M. sativa* CHS2 has an $\alpha\beta\alpha\beta\alpha$ pseudosymmetric motif, originally observed in the crystal structure of homodimeric 3-ketoacyl-CoA thiolase from *Saccharomyces cerevisiae*, and forms a symmetric dimer with a two-fold axis. Furthermore, Met137 protrudes into another monomer by formation of a *cis*-peptide bond between Met137 and Pro138 on a loop, forming a partial wall of the active-site cavity. The catalytic center of type III PKSs is buried deep inside in each monomer and sits at the intersection of characteristic 16-Å-long CoA-binding tunnel and a large internal initiation/elongation/cyclization cavity, which contains the catalytic triad of Cys164, His303, and Asn336 residues (**Figure 4c**). The CoA-binding tunnel connects to the protein surface, enabling the substrates to proceed into the catalytic center. Based on the structural and site-directed mutagenesis studies of *M. sativa* CHS2, the structural machinery for the common iterative condensation reactions of type III PKSs has been proposed as follows (**Scheme 2**).[17,119,120] Under stabilization of the thiolate anion of the



30 24 51

Scheme 23 Enzymatic formation of methylstyrylpyrone (**51**) by *P. strobes* CHS2 from cinnamoyl diketide *N*-acetylcysteamine thioester (**30**) as a starter.



Scheme 24 Enzymatic reactions of methylmalonyl-CoA as extender.

Table 2 Examples of plant, bacterial, and fungal type III PKS that characterized the crystal structure

Enzyme	*Species*	*PDB Id.*	*Ligands*	*Ref.*
One cyclization reaction				
Pyrone - producing - type				
2-Pyrone synthase (2-PS)	*Gerbera hybrida*	1QLV	-	18
		1EE0	Acetoacetyl-CoA (**181**)	
Steely 1	*Dictyostelium discoideum*	2H84	Hexaethylene glycol (**182**)	23
PKS18	*Mycobacterium*	1TED	Myristic acid (**183**)	19
PKS18 C205F	*tuberculosis*	1TEE	-	
CHS - type				
Chalcone synthase2 (CHS2),		1BI5	-	17
EC 2.3.1.74		1BQ6	CoA-SH (**136**)	
	Medicago sativa	1CGK	Naringenin (**76**)	
		1CGZ	Resveratrol (**4**)	
		1CHW	*n*-Hexanoyl-CoA (**31**)	
		1CML	Malonyl-CoA (**1**)	
CHS G256L		1I89	-	121
CHS G256F		1I8B	-	
CHS F215S		1JWX	-	125
STS - type				
Stilbene synthase (STS)	*Arachis hypogaea*	1Z1E	-	83
		1Z1F	Resveratrol (**4**)	
Pinosylvin synthase	*Pinus sylvestris*	1U0U	-	20
18×CHS		1U0V	-	
		1U0W	Resveratrol (**4**)	
2′-Oxoalkylresorcylic acid	*Neurospora crassa*	3EUO	-	25
synthase (ORAS)		3EUT	Eicosanoic acid (**184**)	
		3E1H	-	24
ORAS F252G		3EUQ	PEG 2,000 (**185**)	25
More than two cyclization reactions				
Miscellaneous type				
Pentaketide chromone	*Aloe arborescens*	2D3M	CoA-SH (**136**)	21
synthase (PCS)				
PCS M207G		2D51	-	
		2D52	CoA-SH (**136**)	
Octaketide synthase (OKS)	*A. arborescens*	-	CoA-SH (**136**)	114,124
OKS N222G		-	CoA-SH (**136**)	
1,3,6,8-Tetrahydroxynaphthalene	*Streptomyces coelicolor*	1U0M	PEG (**186**)	
synthase (THNS)				20

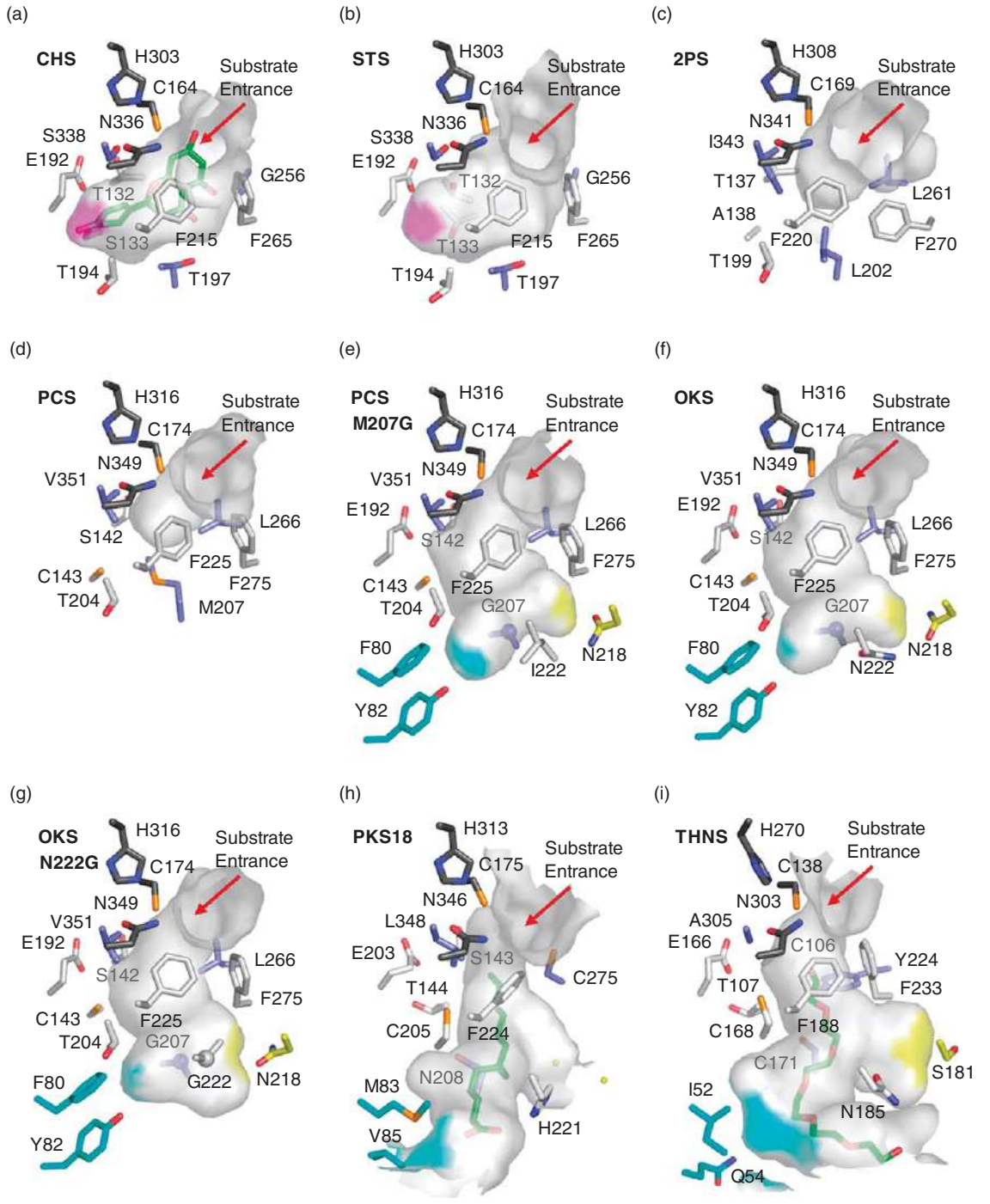


Figure 3 Comparison of the active site architectures of type III PKSs. The only residues (silver, blue, light blue, and yellow) that appear to determine the respective activities of these enzymes are shown with the catalytic triad (black). The bottom of the coumaroyl binding pocket is highlighted as purple surface. The bottoms of the polyketide elongation and another pockets of two downward expanding pockets are highlighted as light blue and yellow surfaces, respectively. The naringenin (**76**), myristic acid (**183**), and polyethylene glycol heptamer (**186**), bound to the respective active site cavities of *Medicago sativa* CHS2, *Mycobacterium tuberculosis* PKS18, and *Streptomyces coelicolor* THNS, are shown as green stick models, to indicate the coumaroyl-binding pocket and the acyl-binding tunnel of their active site cavities.

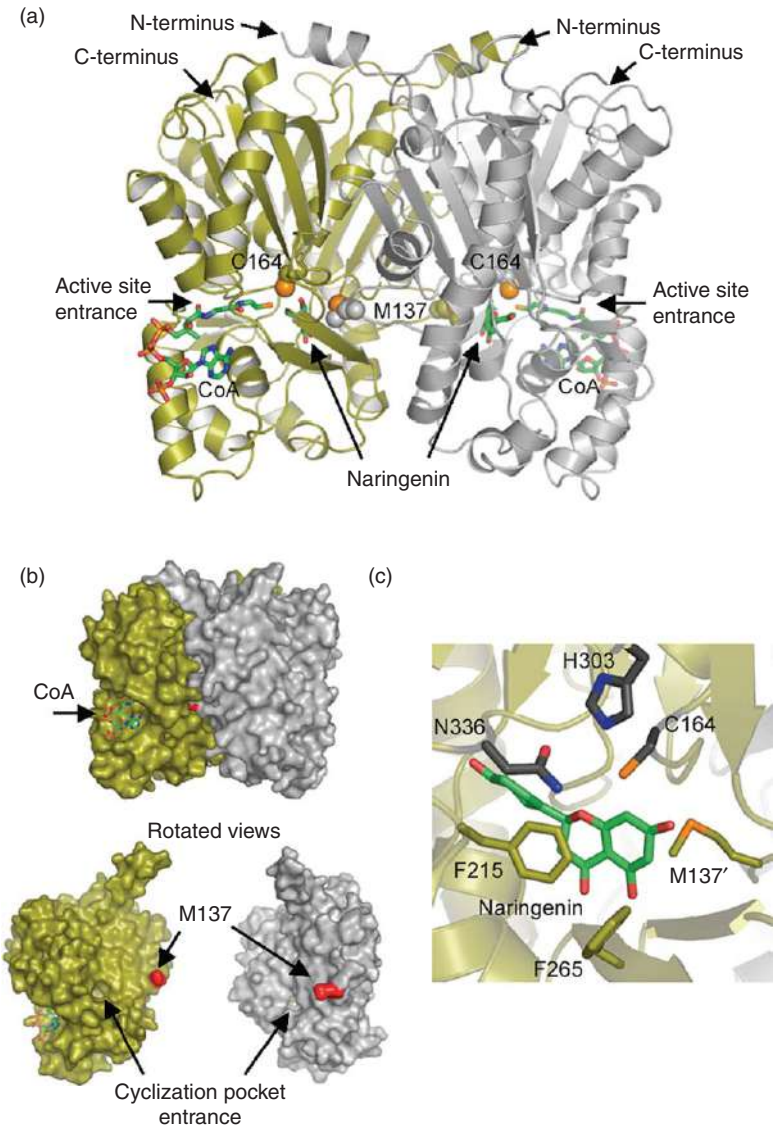


Figure 4 Structure of *Medicago sativa* chalcone synthase 2 (CHS2). (a) Ribbon representation of the CHS2 homodimer. The monomers are colored gold and silver, and the CoA-SH (**136**) and naringenin (**76**) molecules are shown as green stick models. The catalytic Cys164 and Met137, which form a partial wall of the active site cavity of another monomer, are highlighted as CPK models. (b) Molecular surface representation of CHS2. In the bottom panel, the two CHS2 monomers are separated and rotated slightly to highlight the flat dimerization interface. The Met137 is colored red, and the protruded positions of Met137 to another monomer are indicated as cyclization pocket entrance. (c) Close-up view of the CHS2 active site cavity complexed with naringenin (**76**). The Cys-His-Asn catalytic triad and the bound naringenin (**76**) are represented as black and green stick models, respectively. The two gatekeepers Phe215 and Phe265, and Met137 are also indicated.

Cys164 by an imidazolium ion on the His303, the reaction initiation starts by a nucleophilic attack of the thiolate anion to the thioester carbonyl of the starter, resulting in transfer of the acyl moiety to the cysteine side chain. Sequentially, protonated nitrogen of the side chain on His303 and Asn336 forms an 'oxyanion hole' with Phe215, where the thioester carbonyl oxygen of malonyl-CoA (1) orients, providing a nonpolar environment for the terminal carboxylate, which facilitates decarboxylation. A resonance of the enolate ion to the keto form allows for condensation of the acetyl carbanion with the enzyme-bound polyketide intermediate. Recapture of the elongated starter-acetyl-diketide-CoA by Cys164 and release of CoA set the stage for additional rounds of elongation, resulting in the formation of the final polyketide reaction intermediate, which proceeds to

intermolecular cyclization reaction. The so-called gatekeepers Phe215 and Phe265, which are situated at the active site entrance, are thought to help the folding and the internal orientation of the tetraketide intermediate during the cyclization reaction.

In *M. sativa* CHS2, the active site cavity consists of the bilobed initiation/elongation/cyclization cavity. One lobe of this cavity forms a coumaroyl-binding pocket and the other accommodates the growing polyketide chain (**Figure 3(a)**).[17] The coumaroyl-binding pocket is thought to lock the aromatic moiety derived from 4-coumaroyl-CoA (**3**) as the starter substrate on the tetraketide intermediate. It is most likely that the C6/C1 Claisen cyclization reaction with an internal proton transfer from the nucleophilic carbon at C6 to carbonyl oxygen of C6 position facilitates under the locking of aromatic moiety on the tetraketide intermediate to the coumaroyl binding pocket, thereby releasing chalcone from the enzyme active site cavity (**Figure 5(a)**).

In 2000, Noel and coworkers characterized the second crystal structure of *G. hybrida* 2-PS, which shares 74% sequence identity with *M. sativa* CHS2, but yields a different product TAL (**8**) from acetyl-CoA (**7**) and two molecules of malonyl-CoA (**1**).[14,18] The crystal structure complexed with reaction intermediate acetoacetyl-CoA (**181**) at 2.05 Å resolution revealed that *G. hybrida* 2-PS and *M. sativa* CHS2 share common three-dimensional fold, a set of conserved catalytic residues, and similar CoA-binding sites. However, the total cavity volume of the active site cavity of 2-PS is one-third that of *M. sativa* CHS2 (**Figures 3(a–b)** and **6**). Notably, this reduction in volume is achieved by the difference in the introduction of bulkier side chain at chemically inert three residues (corresponding to Leu202, Leu261, and Ile343 in 2-PS, and Thr197, G256, and Ser338 in *M. sativa* CHS2), without rearrangement of the protein main chain. As a further proof, it is shown that alternation of the CHS active site by introducing Thr197L, Gly256L, and S338I converts CHS into 2-PS. It is therefore possible that *G. hybrida* 2-PS prefers the less bulky substrate acetyl-CoA (**7**) than 4-couaroyl-CoA (**3**) as the starter substrate, and reduces chain elongation reaction up to two C_2-unit condensations, in contrast to CHS (**Figure 5(b)**). These results thus provided the structural basis that the volume and shape of the active site principally influence the functional diversity of type III PKSs.

A similar case of steric contraction by the three residues is also found in the active site architecture in *A. arborescens* PCS, which produces a pentaketide chromone (**18**) from five molecules of malonyl-CoA (**1**) (**Figure 3(d)**).[21,32] The crystal structure of *A. arborescens* PCS revealed that the active site volume is as large as that of *G. hybrida* 2-PS, and that the reduction is indeed derived from sterically bulkier residues Met207, Leu266, and Val351 than Thr197, Gly256, and Ser338 in *M. sativa* CHS2 with additional conformational differences of Cys143 and Thr204 corresponding to Ser133 and Thr194 in *M. sativa* CHS2 (**Figure 7**). In addition, electron density map showed that the side chain of Met207 of *A. arborescens* PCS is very flexible, whereas the side chains of Thr197 of *M. sativa* CHS2 and Leu261 of *G. hybrida* 2-PS are not so flexible. It is therefore most likely that the conformationally flexible Met207 locks the acetyl moiety of the pentaketide intermediate and then the Claisen-type cyclization/aromatization reaction of the enzyme-bound intermediate takes place with a common structural machinery, as in the case of CHS (**Figure 5(c)**). Interestingly, the small-to-large Gly256L substitution (numbering to *M. sativa* CHS2) is also observed in *A. arborescens* PCS, *A. arborescens* OKS, and *A. arborescens* and *R. palmatum* ALSs, each of which selects acetyl- or malonyl-CoA (**1**) as a starter substrate. In contrast, tetraketide-forming type III PKSs, including other CHS, STS, VPS, CTAS, BBS, and ACS, conserve Gly residue at this position. Site-directed mutagenesis studies suggest that Gly256 determines the starter specificity and the chain length in *M. sativa* CHS2,[121] whereas the spatial tininess of the Gly256 determines the starter selectively in *S. baicalensis* CHS and *R. palmatum* ALS Gly256.[122,123]

In contrast, further crystal structure of the octaketide-producing PCS Met207G mutant revealed that the residue 207 (corresponding to Thr197 in *M. sativa* CHS2) lining the active site cavity occupies a crucial position for the polyketide chain elongation reactions (**Figures 3(e)** and **8**). Met207 of the wild-type PCS blocks the entrances of two novel, hidden pockets that extend into the 'floor' of the active site cavity.[6,21,60] The large-to-small substitution thus drastically increases the active site volume 2.6 times, compared with that of wild-type PCS, by opening the gate to the downward pockets, which leads to the formation of longer octaketides (**22, 23**) instead of the pentaketide (**18**). Importantly, the hidden pockets differ from the coumaroyl-binding pocket observed in the active site in *M. sativa* CHS2, but the location and orientation of one of the pockets accommodate the long substrate-binding tunnel in the similar active site architecture of the bacterial type III PKSs including *M. tuberculosis* PKS18,[19] *S. coelicolor* THNS,[20] *Dictyostelium discoideum* C-terminal Steely 1,[23] and *N. crassa* ORAS (**Figure 3**).[24,25] In addition, almost the same active site architecture with the

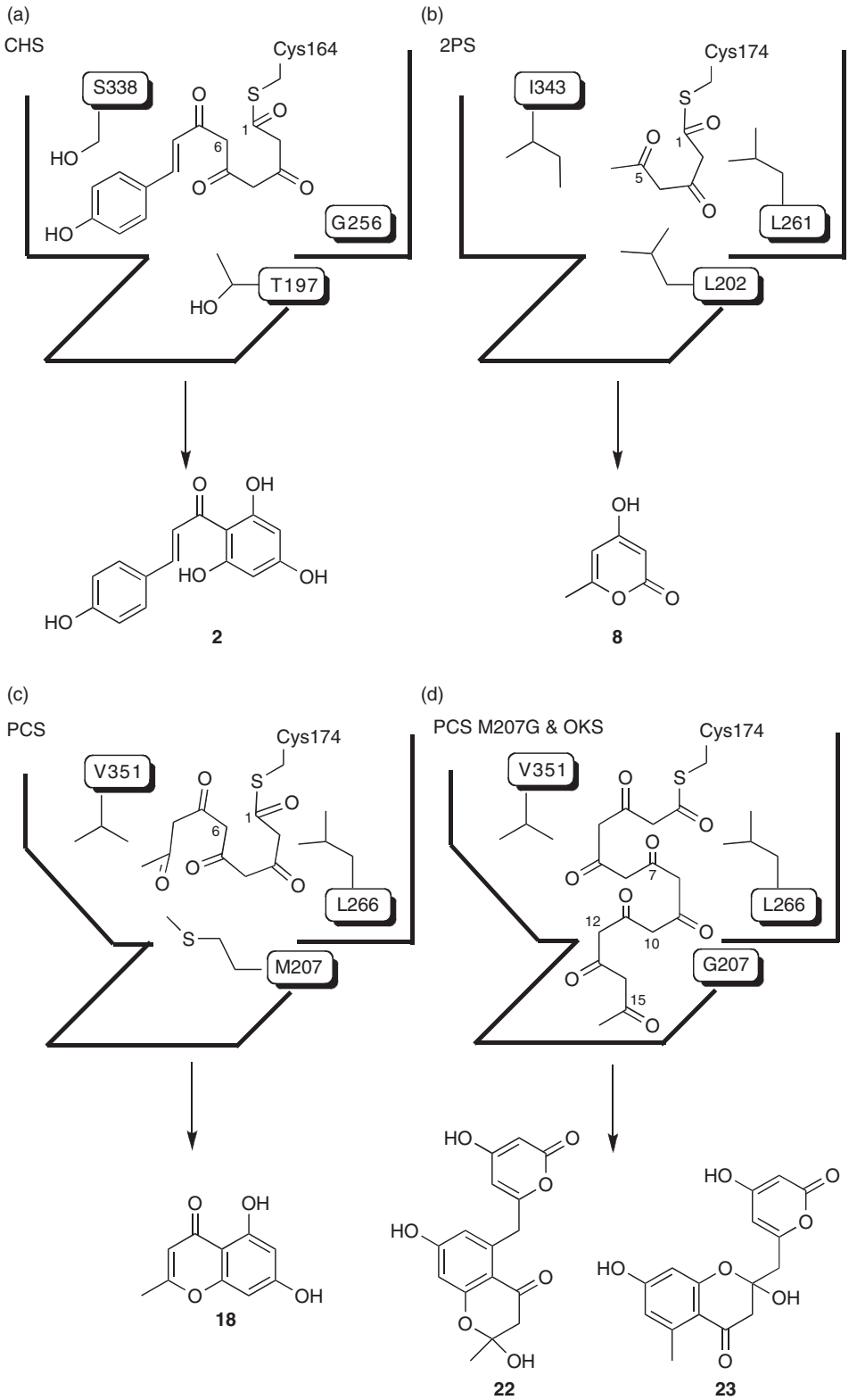


Figure 5 Schematic representation of the active site architecture of (a) *Medicago sativa* CHS2, (b) *Gerbera hybrida* 2-PS, (c) *Aloe arborescens* PCS, and (d) both PCS M207G mutant and *A. arborescens* OKS, and their products.

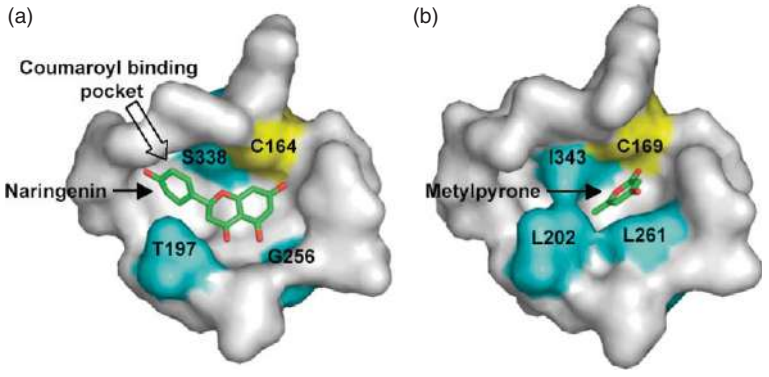


Figure 6 Comparison of the active site cavity of (a) *Medicago sativa* CHS2 and (b) *Gerbera hybrida* 2-PS. Phe215 and Phe265 in CHS2 and Phe220 and Phe270 in 2-PS were removed from the respective surface representation. Methylpyrone in the 2-PS active site cavity is a model. The catalytic residue Cys164 is colored yellow, whereas the chemically inert three residues that play an important role in the functional diversity of both enzymes are highlighted in light blue.

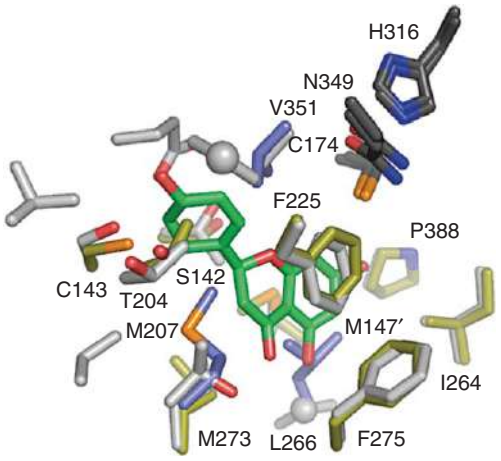


Figure 7 Comparison of the active site cavity of *Medicago sativa* CHS2 (silver) and *Aloe arborescens* PCS (gold). The chemically inert three residues in *A. arborescens* PCS is colored blue. The catalytic triads in both enzymes are indicated in black.

downward expanding polyketide elongation tunnel of the PCS M207G mutant was observed in the crystal structure of octaketide-producing *A. arborescens* OKS in natural resource (**Figure 3(f)**).[114,124] Furthermore, a chain length control depending on the steric bulk of single side chain at residue 197 has been shown in octaketide-producing *A. arborescens* OKS and heptaketide-producing *R. palmatum* and *A. arborescens* ALSs.[38,63,123] For example, small-to-large substitutions (Gly207A, Gly207T, Gly207M, Gly207L, and Gly207W) in *A. arborescens* OKS resulted in the loss of octaketide-forming activity and the concomitant formation of shorter chain length polyketides. It is thus strongly suggested that the related malonyl-primed long-chain polyketide-producing plant type III PKSs, such as heptaketide-producing *A. arborescens* and *R. palmatum* ALSs, share a common active site architecture and reaction machinery, as in the case of the PCS M207G mutant.

The CHS active site cavity can be extended to allow for the synthesis of a trace amount of longer octaketides (**22,23**) by a single Ser338V substitution.[122] The crystal structure also exhibits the presence of an additional buried pocket that extends into the 'floor' of the traditional CHS active site, as in the case of the PCS M207G mutant and *A. arborescens* OKS enzymes. Based on the crystal structures of the plant type III PKSs, Ser338 is located in proximity of the catalytic Cys164, which is the covalent attachment site for the growing polyketide intermediates. These observations suggest that the residue 338 plays an important role in steric guidance of the

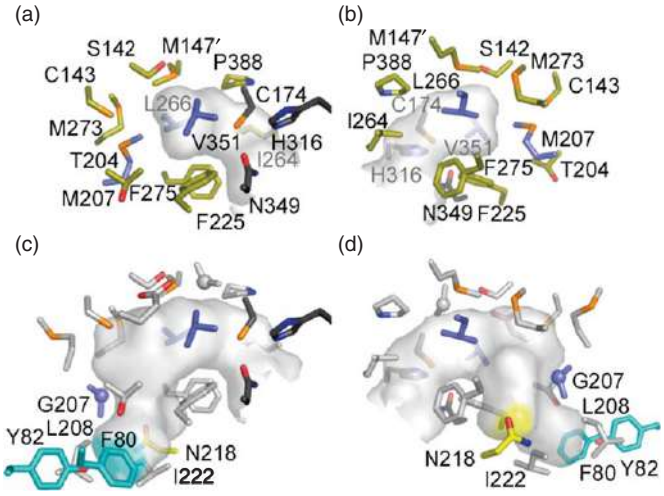
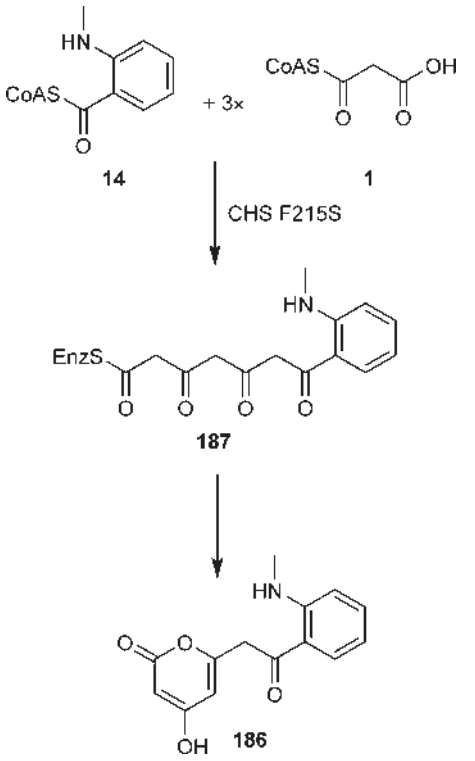


Figure 8 Comparison of the active site cavity of *Aloe arborescens* PCS and the PCS M207G mutant. (a) Front and (b) rear views of the active site cavity surface of wild-type PCS. (c) Front and (d) rear views of the active-sity cavity surface of the PCS M207G mutant. The active site residues and two pockets are colored as in **Figure 3**.

growing intermediate so that the linear intermediate extends into the additional buried pocket, thereby leading to longer octaketides (**22**,**23**). The improvement of the steric guidance at residue 338 would be provided from loss of the heptaketide-forming activity in *R. palmatum* ALS by Thr338S substitution and from twofold increase in the benzalacetone-forming activity in *R. palmatum* BAS by Ser338V substitution (numbering to *M. sativa* CHS2).[44,123] On the contrary, Noel and coworkers provided the only known CHS mutant that can accept *N*-methylanthraniloyl-CoA (**14**) as a starter substrate to produce the unnatural alkaloid *N*-methylanthraniloyltriacetic acid lactone (**186**) by a Phe215S substitution in *M. sativa* CHS2 (**Scheme 25**).[125] For the structural machinery, crystal structure analysis of the mutant enzyme suggests that the serine hydroxyl group forms a hydrogen bond with the backbone carbonyl oxygen of Gly211, which widens the entrance to the active site of the wild-type CHS2 by 4–5 Å. It is thus most likely that the Phe215S substitution opens a space at the cavity entrance to accommodate the methylmalonylamine moiety of *N*-methylanthraniloyl-CoA (**14**) and allows for positioning of the thioester carbonyl moiety next to Cys164, His303, and Asn336.[125] It should be noted that diketide-forming activity of *R. palmatum* BAS is attributed to the characteristic substitution of the conserved active site Phe215 with leucine (numbering to *M. sativa* CHS2).[126] In the case of BAS, Leu215F substitution restores the chalcone-forming activity. Additionally, substitutions of Ser132, Ala133, and Val265 with Thr, Ser, and Phe also transform acridone-producing *R. graveolens* ACS into CHS (numbering to *M. sativa* CHS2).[127,128] Although the intimate structural machinery has remained unclear, the functional diversity of *R. palmatum* BAS and *R. graveolens* ACS has been suggested to be controlled by a steric factor.[44,128]

Besides the simple steric contraction, structure analyses of *P. sylvestris* STS and *A. hypogaea* STS have provided new insights into the molecular diversity of the STS enzymes (**Figure 3(b)** and **Figure 9**).[20,83] It is interesting to note that unlike the simple steric contraction, an electronic effect rather than the steric factor determines the product specificity of STS. Importantly, a significant backbone change occurs on a loop between residues 132 and 136 in the active site of STS, thereby resulting in a subtle displacement of Thr132, and bringing its side chain hydroxyl moiety within the hydrogen-bonding distance of a water molecule. As a result, the so-called 'aldol-switch' thioesterase-like electronic hydrogen-bond network involving Ser338–H_2O–Thr132–Glu192 is rearranged in proximity of the catalytic Cys164 in the active site cavity in STS. Additionally, there is no reorientation of resveratrol (**4**) and naringenin (**76**) binding to the active site cavity between the STS and CHS enzymes. Furthermore interestingly, in 1966, Harris and others have indicated that when C1 position of the linear polyketide is part of an ester (or thioester) bond, C6/C1 Claisen cyclization predominates, whereas C7/C2 aldol cyclization is favored when C1 position is free acid.[129] Based on these findings, the reaction machinery in the active site of STS is most likely that the nucleophilic water by the aldol



Scheme 25 Enzymatic synthesis of unnatural *N*-methylanthraniloyltriacetic acid lactone by *M sativa* CHS F215S mutant enzyme from *N*-methylanthraniloyl-CoA (**14**) as a starter.

switch initially cleaves the thioester bond of the cysteine-bound linear intermediate.[20] Thus, the C7/C2 aldol cyclization predominantly proceeds within the resultant acidic intermediate and then the decarboxylation from the cyclized carboxylic intermediate finally occurs to produce resveratrol (**4**) (**Scheme 11a**).

1.06.4 Protein Engineering

Plant type III PKSs show promiscuous broad substrate specificity, and can produce unnatural polyketides. In fact, as described above, unnatural polyketides have been synthesized by *in vitro* reaction with various nonphysiological substrates such as the aromatic ring-replaced (furan or thiophene) coumaroyl-CoA analogs (**81a, 82a**) and methylmalonyl-CoA (**24**).[6] It is therefore theoretically possible that if we can control the starter selectivity, polyketide chain length, and/or the cyclization reaction, as well as prepare more substrate analogs and use combinational modification enzymes, billions of polyketide analogs can be synthesized under mild conditions. Structure-based engineering of type III PKSs has further increased the possibility of enzymatic synthesis of various polyketides.

For example, on the basis of the crystal structures of wild-type and M207G mutant of PCS, F80A/Y82A/M207G triple mutant has succeeded in the release of an unnatural novel nonaketide naphthopyrone (**188**) through condensation with nine molecules of malonyl-CoA (**1**) (**Figure 10**).[60,130] A homology model predicted that the active site cavity volume of the triple mutant increased 4 times that of the wild-type PCS. On the contrary, large-to-small Asn222G substitution at the bottom of the polyketide chain elongation tunnel let to decaketide-producing enzyme from octaketide-producing OKS, resulting in the formation of unnatural decaketide benzophenone, SEK15 (**189**), by 10-round condensations (**Figures 3(f) 3(g)** and **Scheme 26**).[114] SEK15 has been previously reported as a product of genetically engineered type II PKSs,[110] and now the longest polyketide produced by the structurally simple type III

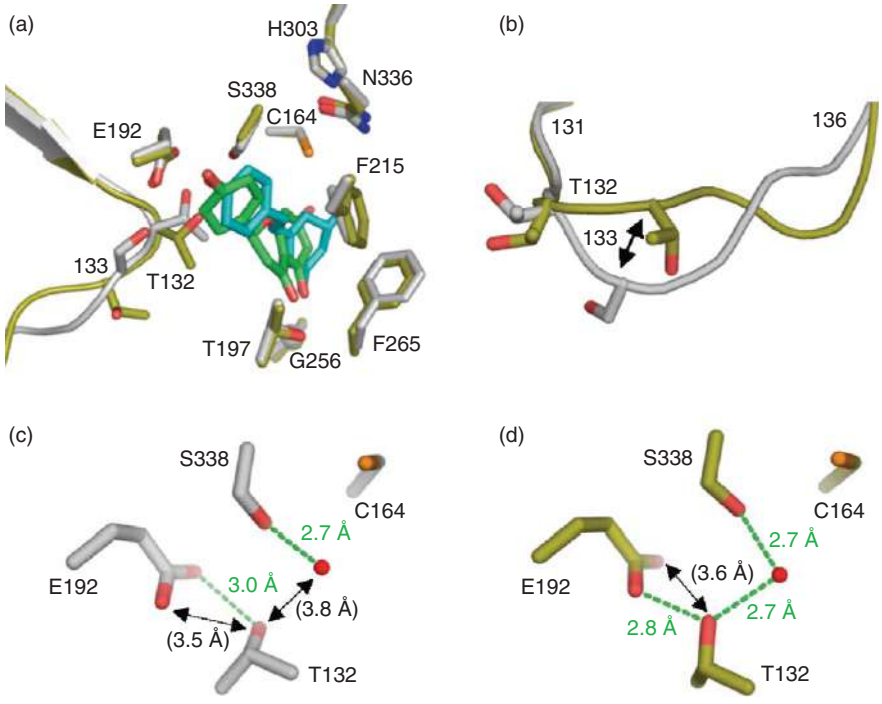


Figure 9 (a) Comparison of the active site of *Medicago sativa* CHS2 (silver) and *Pinus sylvestris* STS (gold). Naringenin (**76**) complexed with *M. sativa* CHS2 is shown in a green stick model. Resveratrol (**4**) in structure of 18 × CHS, which is a mutant enzyme having the STS activity instead of CHS activity, is shown in light blue stick model, and is superimposed in (a), to indicate that intermediate of both enzymes are locked in an almost identical position. (b) Close-up view of the loops between 131 and 136 in STS and CHS. (c and d) Close-up view of the hydrogen bond networks in (c) CHS and (d) STS. The hydrogen bond formations and their distances are indicated with green parentheses. The putative nucleophilic water molecule in STS and the relative water molecule in CHS2 are shown as red spheres.

PKS. Furthermore, in contrast to wild-type OKS, the OKS N222G mutant efficiently accepts 4-coumaroyl-CoA as a starter and yields the unnatural hexaketide stilbene (**170**) and heptaketide chalcone (**171**) as the major products (**Scheme 22**).[114] A similar profile was also obtained when *n*-hexanoyl-CoA (**31**) was used as a starter, and the heptaketide phloroglucinol (**173**) was efficiently produced. Structure-based engineering of type III PKSs would thus lead to further production of chemically and structurally disparate unnatural novel polyketides.

1.06.5 Combinational Biosynthesis

Combinatorial biosynthesis is an attractive method to synthesize potential seeds for clinical drugs, or to synthesize a series of candidate compounds for high-throughput screening.[129–133] Successful examples are the production of flavonoids by recombinant *E. coli* and yeast cells.[134–140] Coincubation of yeast cells carrying isoflavone synthase (IFS) and recombinant *E. coli* cells carrying phenylalanine ammonia-lyase (PAL), cinnamate/coumarate:CoA-ligase (CCL), CHS, and CHI cDNAs with tyrosine (**190**) expectedly led to the production of genistein (**191**), an isoflavone that is famous for significant estrogen-like activities (**Scheme 27**).[136] Using a similar methodology, Horinouchi and coworkers have successfully constructed a system to synthesize nearly 100 polyketides, which includes 36 unnatural flavonoids and stilbenes (**Scheme 28**).[137] In addition, same researchers have also provided the high-yield curcuminoids production system using *E. coli*.[139] This outstanding achievement has provided a promising methodology to construct a chemical library of polyketides.

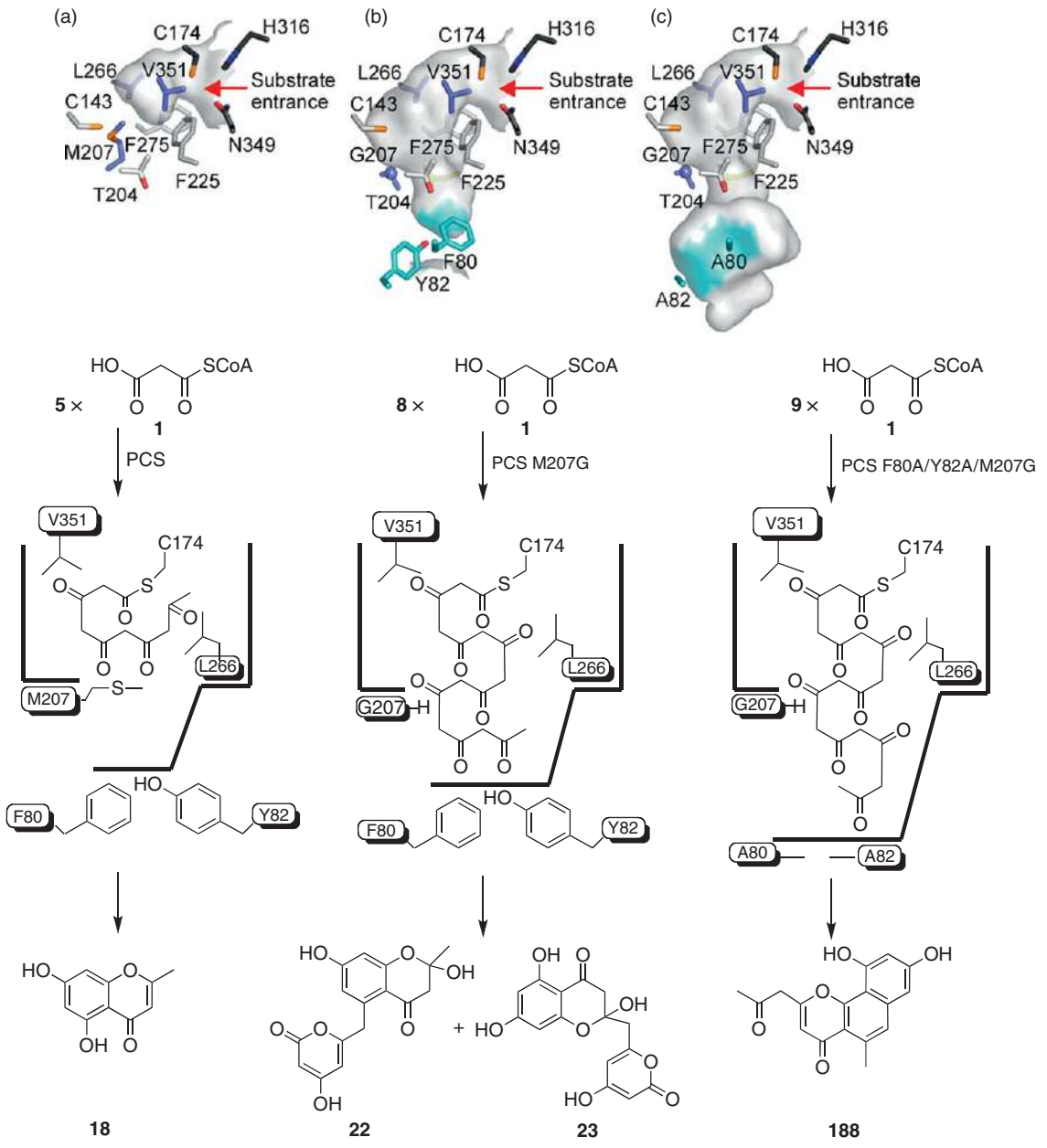
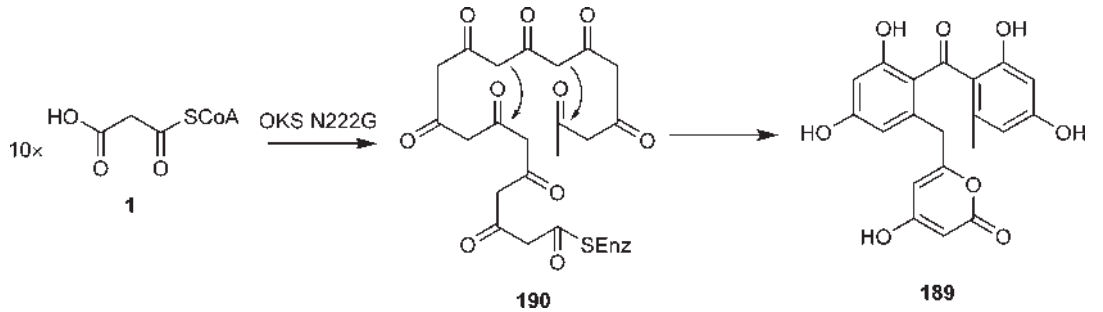
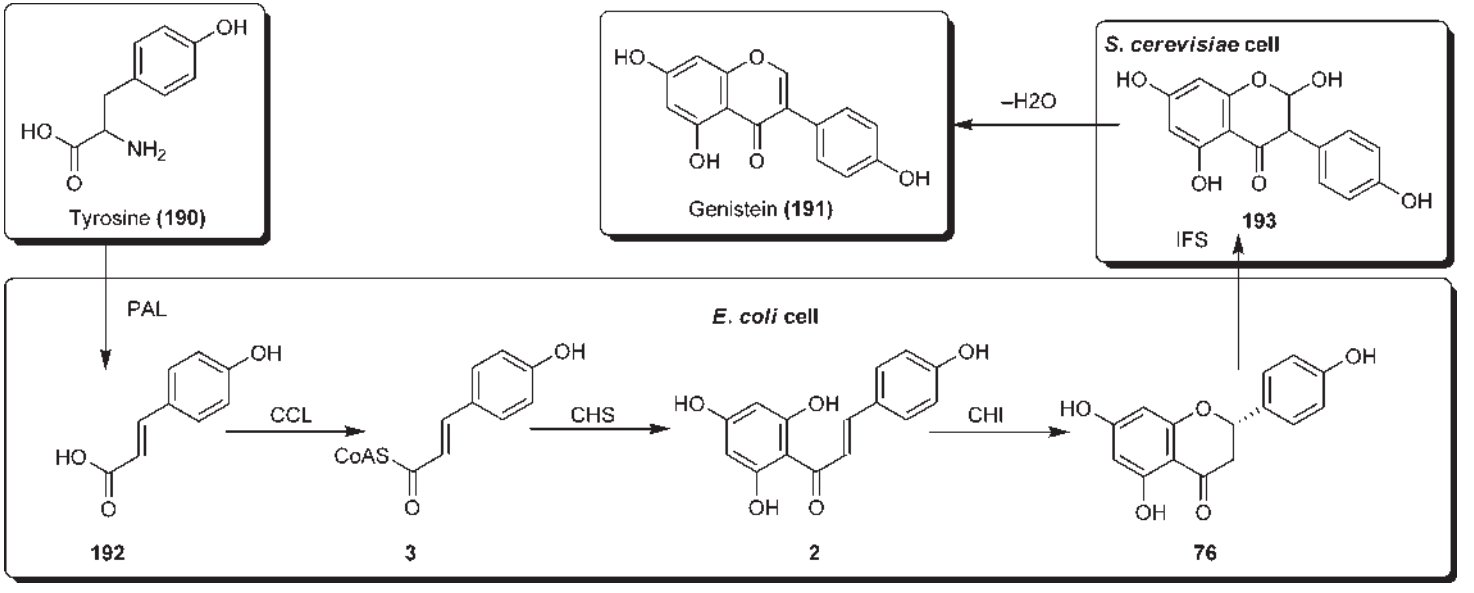


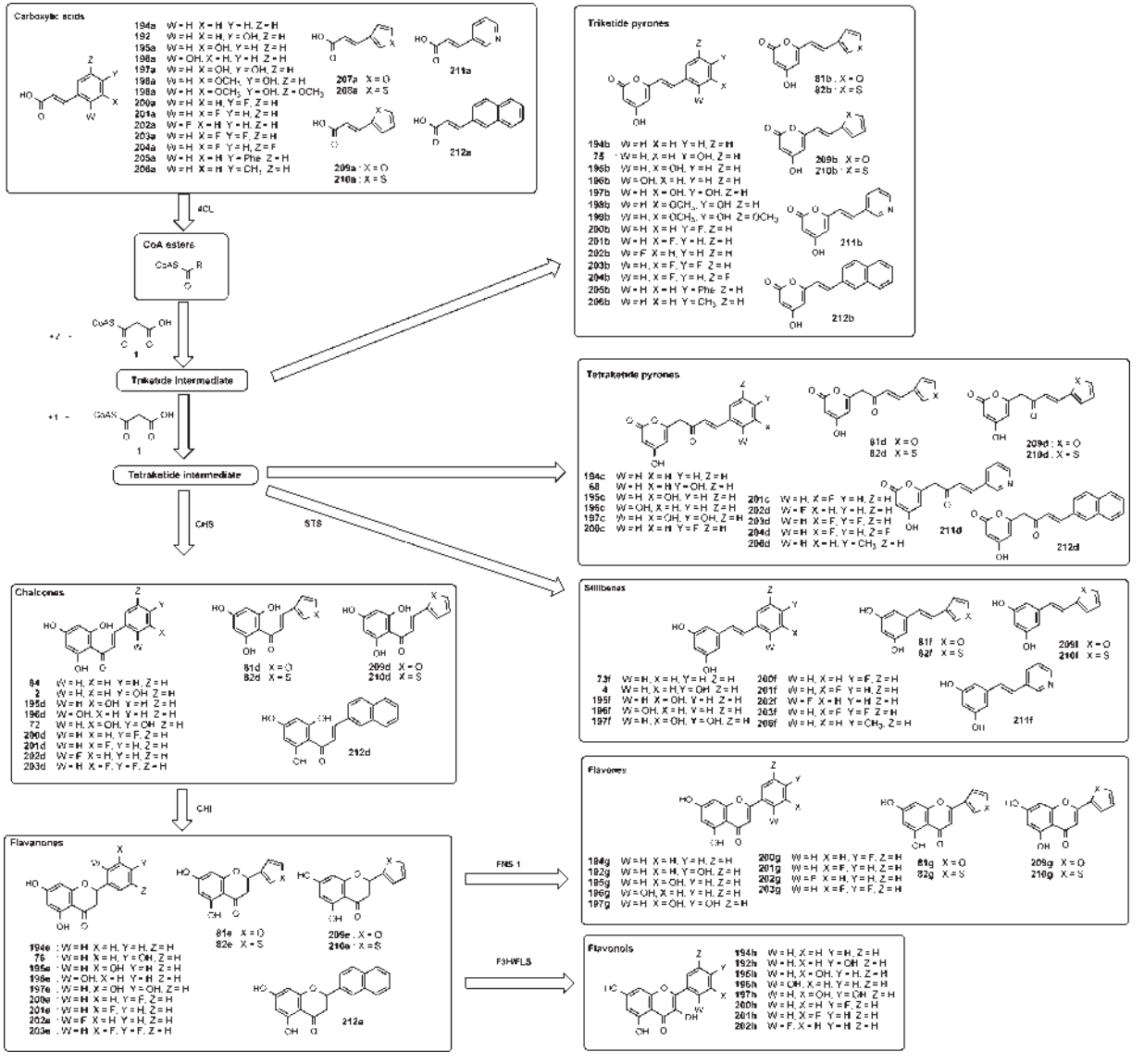
Figure 10 Structure-based engineering of *Aloe arborescens* PCS and enzymatic formation of an unnatural nonaketide naphthopyrone (**188**). (a) wild-type PCS, (b) PCS M207G mutant, and (c) PCS F80A/Y82A/M207G mutant.



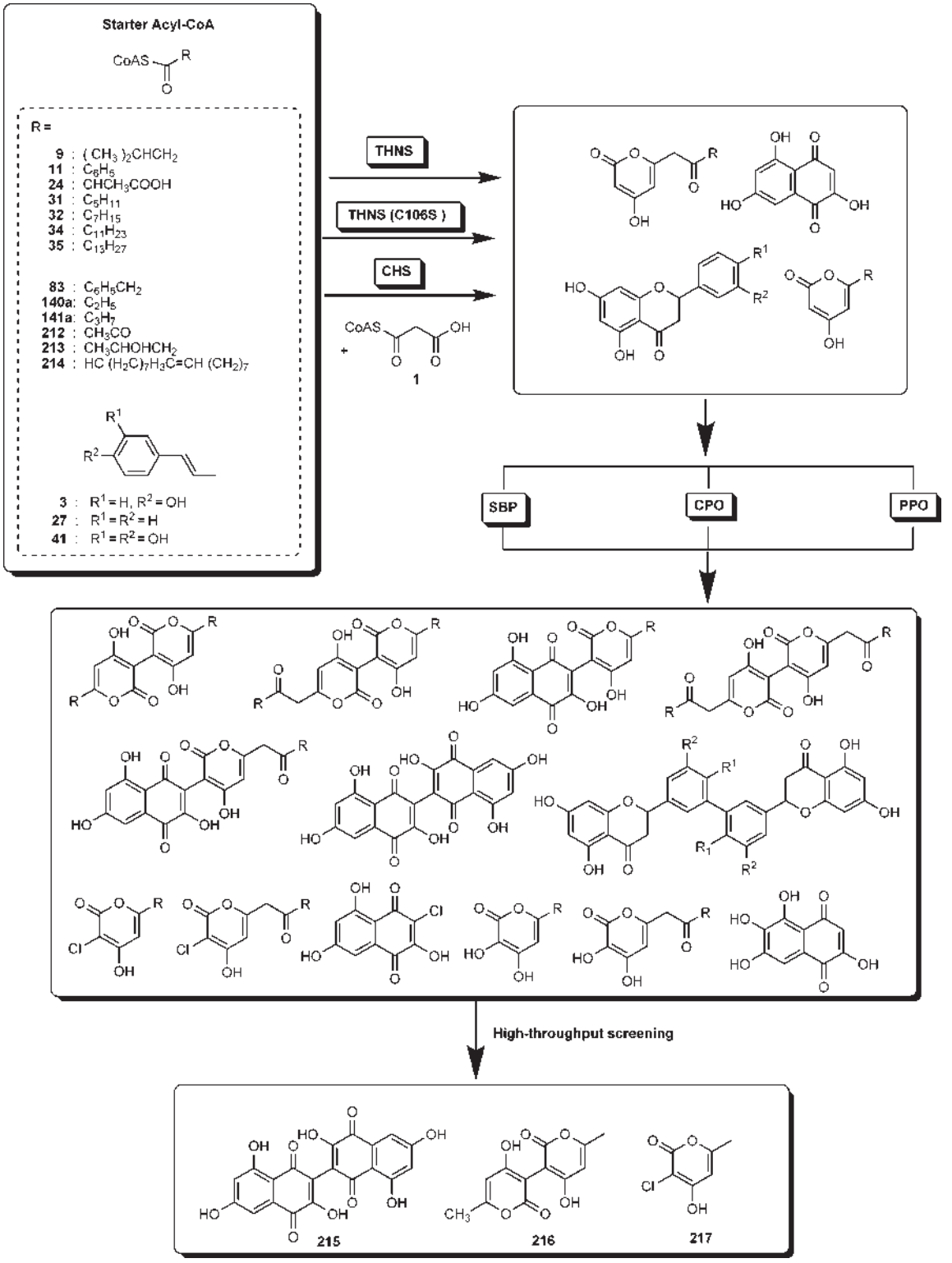
Scheme 26 Enzymatic synthesis of SEK15 (**189**) by OKS N222G mutant from 10 molecules of malonyl-CoA (**1**).



Scheme 27 Production of genistein (191) by coincubation of yeast carrying IFS and *E. coli* cells carrying PAL, CCL, CHS, and CHI cDNAs. Abbreviations: IFS, isoflavone synthase; PAL, phenylalanine ammonia-lyase; CCL, cinnamate/coumarate, CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase.



Scheme 28 Strategy for generating polyketides by type III PKSs. Abbreviations: 4CL, 4-coumarate: CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; FNS 1, flavone synthase I; F3H, flavanone 3β-hydroxylase; FLS, flavonol synthase; STS, stilbene synthase.



Scheme 29 Strategy for generating polyketides by type III PKSs and tailoring enzymes. Abbreviations: CPO, chloroperoxidase; PPO, soybean peroxidase; SBP, polyphenol oxidase; THNS, 1,3,6,8-tetrahydroxynaphthalene synthase.

On the contrary, Dordick and coworkers have developed a high-throughput microarray-based method to synthesize polyketide derivatives *in vitro*. Using several type III PKSs together with oxidative post-PKS tailoring enzymes, nearly 200 individual or multienzymatic reactions were successfully performed on a single glass microarray (**Scheme 29**). Subsequent array-based screening with a human tyrosine kinase led to the identification of three compounds that acted as modest inhibitors in the low-micromolar range.[141] The design of *in vitro* biosynthesis coupled with high-throughput technology thus serves as a new platform for the synthesis and identification of novel biologically active compounds.[141]

1.06.6 Conclusions

Polyketides with diverse structural scaffolds and biological activities are precious treasures for mankind. In the last several years, a large number of type III PKSs have been discovered in plants, and intimate catalytic mechanism has also been elucidated. However, many problems are still remaining. For example, cyclization mechanism including bi- and tricyclization and their timing are still unclear. On the contrary modern technologies now provide us with powerful tools to exploit the mysterious mechanisms of the biosynthetic pathways, which make it possible to generate important pharmaceutical seed compounds by enzymatic synthesis, and to screen out target compounds with promising biological activities. Further analyses of the catalytic potential and plasticity of functionally divergent type III PKS enzymes promise to reveal intimate structural details of the enzyme-catalyzed processes, and provide strategies for the structure-based rational engineering of enzymes and for manipulating substrate and product specificities of the polyketide formation reactions.

Acknowledgments

The authors express their appreciation to their coworkers, in particular Professor Yutaka Ebizuka, Dr. Toshiyuki Kohno, and Dr. Shigetoshi Sugio. Financial support at Shizuoka has been provided by the PRESTO program of Japan Science and Technology Agency, and Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

Abbreviations

ACS	acridone synthase
ALS	aloesone synthase
BAS	benzalacetone synthase
BBS	bibenzyl synthase
Bgh	*Blumeria graminis* f.sp *hordei*.
BIS	biphenyl synthase
BNY	bis-noryangonin
BPS	benzophenone synthase
BUS	isobutyrophenone synthase
4CL	4-coumarate:CoA ligase
CCL	cinnamate/coumarate:CoA ligase
CHI	chalcone isomerase
CHS	chalcone synthase
CPO	chloroperoxidase
CTAL	4-coumaroyltriacetic acid lactone
CTAS	coumaroyl triacetic acid lactone synthase
CUS	curcuminoid synthase
FAS	fatty acid synthase

F3H	flavanone 3β-hydroxylase
FLS	flavonol synthase
FNS 1	flavone synthase I
HEDS or HvCHS2	homoeriodictyol/eriodictyol synthase
HKS	hexaketide synthase
IFS	isoflavone synthase
KAS	ketoacyl synthase
NAC	*N*-acetylcysteamine
OKS	octaketide synthase
OMT	*O*-methyltransferase
ORAS	2′-oxoalkylresorcylic acid synthase
PAL	phenylalanine ammonia-lyase
PCS	pentaketide chromone synthase
PKS	polyketide synthase
PPO	soybean peroxidase
2-PS	2-pyrone synthase
SBP	polyphenol oxidase
STCS	stilbenecarboxylate synthase
STS	stilbene synthase
THNS	1,3,6,8-tetrahydroxynaphthalene synthase
VPS	phlorisovalerophenone synthase

References

1. R. A. Dixon, *Nature* **2001**, *411*, 843–847.
2. M. Jang; L. Cai; G. O. Udeani; K. V. Slowing; C. F. Thomas; C. W. Beecher; H. H. Fong; N. R. Farnsworth; A. D. Kinghorn; R. G. Mehta; R. C. Moon; J. M. Pezzuto, *Science* **1997**, *275*, 218–220.
3. R. Li; G. L. Kenyon; F. E. Cohen; X. Chen; B. Gong; J. N. Dominguez; E. Davidson; G. Kurzban; R. E. Miller; E. O. Nuzum; P. J. Rosenthal, *J. Med. Chem.* **1995**, *38*, 5031–5037.
4. M. B. Austin; J. P. Noel, *Nat. Prod. Rep.* **2003**, *20*, 79–110.
5. J. Schröder, The Chalcone/Stilbene-Synthase Type Family of Condensing Enzymes. In *Comprehensive Natural Products Chemistry*; U. Sankawa, Ed.; Elsevier: Oxford, 1999; Vol. 2, 749–771.
6. S. P. Shi; H. Morita; K. Wanibuchi; Y. Mizuuchi; H. Noguchi; I. Abe, *Curr. Org. Synth.* **2008**, *5*, 250–266.
7. G. Schröder; J. W. Brown; J. Schröder, *Eur. J. Biochem.* **1988**, *172*, 161–169.
8. J. Fliegmann; G. Schröder; S. Schanz; L. Britsch; J. Schröder, *Plant Mol. Biol.* **1992**, *18*, 489–503.
9. F. Melchior; H. Kindl, *FEBS Lett.* **1990**, *268*, 17–20.
10. W. Wang; S. B. Wan; P. Zhang; H. L. Wang; J. C. Zhan; W. D. Huang, *Plant Physiol. Biochem.* **2008**, *46*, 1085–1092.
11. K. T. Junghanns; R. E. Kneusel; A. Baumert; W. Maier; D. Groger; U. Matern, *Plant Mol. Biol.* **1995**, *27*, 681–692.
12. R. Lukacin; K. Springob; C. Urbanke; C. Ernwein; G. Schroder; J. Schröder; U. Matern, *FEBS Lett.* **1999**, *448*, 135–140.
13. Y. Helariutta; P. Elomaa; M. Kotilainen; R. J. Griesbach; J. Schröder; T. H. Teeri, *Plant Mol. Biol.* **1995**, *28*, 47–60.
14. C. Eckermann; G. Schröder; J. Schmidt; D. Strack; R. A. Edrada; Y. Helariutta; P. Elomaa; M. Kotilainen; I. Kilpeläinen; P. Proksch; T. H. Teeri; J. Schröder, *Nature* **1998**, *396*, 387–390.
15. N. Funa; Y. Ohnishi; I. Fujii; M. Shibuya; Y. Ebizuka; S. Horinouchi, *Nature* **1999**, *400*, 897–899.
16. N. Funa; T. Awakawa; S. J. Horinouchi, *Biol. Chem.* **2007**, *282*, 14476–14481.
17. J. L. Ferrer; J. M. Jez; M. E. Bowman; R. A. Dixon; J. P. Noel, *Nat. Struct. Biol.* **1999**, *6*, 775–784.
18. J. M. Jez; M. B. Austin; J. Ferrer; M. E. Bowman; J. Schröder; J. P. Noel, *Chem. Biol.* **2000**, *7*, 919–930.
19. R. Sankaranarayanan; P. Saxena; U. B. Marathe; R. S. Gokhale; V. M. Shanmugam; R. Rukmini, *Nat. Struct. Mol. Biol.* **2004**, *11*, 894–900.
20. M. B. Austin; M. E. Bowman; J. L. Ferrer; J. Schroder; J. P. Noel, *Chem. Biol.* **2004**, *11*, 1179–1194.
21. H. Morita; S. Kondo; S. Oguro; H. Noguchi; S. Sugio; I. Abe; T. Kohno, *Chem. Biol.* **2007**, *14*, 359–369.
22. M. B. Austin; M. Izumikawa; M. E. Bowman; D. W. Udwary; J. L. Ferrer; B. S. Moore; J. P. Noel, *J. Biol. Chem.* **2004**, *279*, 45162–45174.
23. M. B. Austin; T. Saito; M. E. Bowman; S. Haydock; A. Kato; B. S. Moore; R. R. Kay; J. P. Noel, *Nat. Chem. Biol.* **2006**, *2*, 494–502.
24. A. Goyal; P. Saxena; A. Rahman; P. K. Singh; D. P. Kasbekar; R. S. Gokhale; R. Sankaranarayanan, *J. Struct. Biol.* **2008**, *162*, 411–421.
25. S. B. Rubin-Pitel; H. Zhang; T. Vu; J. S. Brunzelle; H. Zhao; S. K. Nair, *Chem. Biol.* **2008**, *15*, 1079–1090.
26. U. Reimold; M. Kroger; F. Kreuzaler; K. Hahlbrock, *EMBO J.* **1983**, *2*, 1801–1805.

27. B. S. Moore; J. N. Hopke, *Chembiochem.* **2001**, *2*, 35–38.
28. P. Saxena; G. Yadav; D. Mohanty; R. S. Gokhale, *J. Biol. Chem.* **2003**, *278*, 44780–44790.
29. S. Tropf; T. Lanz; S. A. Rensing; J. Schröder; G. Schröder, *J. Mol. Evol.* **1994**, *38*, 610–618.
30. N. B. Paniego; K. W. Zuurbier; S. Y. Fung; R. van der Heijden; J. J. Scheffer; R. Verpoorte, *Eur. J. Biochem.* **1999**, *262*, 612–616.
31. Y. Okada; K. Ito, *Biosci. Biotechnol. Biochem.* **2001**, *65*, 150–155.
32. I. Abe; Y. Utsumi; S. Oguro; H. Morita; Y. Sano; H. Noguchi, *J. Am. Chem. Soc.* **2005**, *127*, 1362–1363.
33. T. Reinecke; H. Kindl, *Phytochemistry* **1994**, *35*, 63–66.
34. R. Preisig-Muller; P. Gnau; H. Kindl, *Arch. Biochem. Biophys.* **1995**, *317*, 201–207.
35. B. Liu; T. Raeth; T. Beuerle; L. Beerhues, *Planta* **2007**, *225*, 1495–1503.
36. C. Eckermann; G. Schröder; S. Eckermann; D. Strack; J. Schmidt; B. Schneider; J. Schröder, *Phytochemistry* **2003**, *62*, 271–286.
37. I. Abe; Y. Utsumi; S. Oguro; H. Noguchi, *FEBS Lett.* **2004**, *562*, 171–176.
38. I. Abe; S. Oguro; Y. Utsumi; Y. Sano; H. Noguchi, *J. Am. Chem. Soc.* **2005**, *127*, 12709–12716.
39. K. Karppinen; J. Hokkanen; S. Mattila; P. Neubauer; A. Hohtola, *FEBS J.* **2008**, *275*, 4329–4342.
40. I. Abe; Y. Takahashi; H. Morita; H. Noguchi, *Eur. J. Biochem.* **2001**, *268*, 3354–3359.
41. D. Zheng; G. Hrazdina, *Arch. Biochem. Biophys.* **2008**, *470*, 139–145.
42. Y. Katsuyama; M. Matsuzawa; N. Funa; S. Horinouchi, *J. Biol. Chem.* **2007**, *282*, 37702–37709.
43. W. Borejsza-Wysocki; G. Hrazdina, *Plant Physiol.* **1996**, *110*, 791–799.
44. T. Abe; H. Morita; H. Noma; T. Kohno; H. Noguchi; I. Abe, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 3161–3166.
45. Y. Katsuyama; M. Matsuzawa; N. Funa; S. Horinouchi, *Microbiology* **2008**, *154*, 2620–2628.
46. I. Abe; T. Abe; K. Wanibuchi; H. Noguchi, *Org. Lett.* **2006**, *8*, 6063–6065.
47. J. Schröder; S. Raiber; T. Berger; A. Schmidt; J. Schmidt; A. M. Soares-Sello; E. Bardshiri; D. Strack; T. J. Simpson; M. Veit; G. Schröder, *Biochemistry* **1998**, *37*, 8417–8425.
48. Y. Mizuuchi; Y. Shimokawa; K. Wanibuchi; H. Noguchi; I. Abe, *Biol. Pharm. Bull.* **2008**, *31*, 2205–2210.
49. T. Akiyama; M. Shibuya; H. M. Liu; Y. Ebizuka, *Eur. J. Biochem.* **1999**, *263*, 834–839.
50. P. Klingauf; T. Beuerle; A. Mellenthin; S. A. El-Moghazy; Z. Boubakir; L. Beerhues, *Phytochemistry* **2005**, *66*, 139–145.
51. L. Beerhues, *FEBS Lett.* **1996**, *383*, 264–266.
52. B. Liu; H. Falkenstein-Paul; W. Schmidt; L. Beerhues, *Plant J.* **2003**, *34*, 847–855.
53. K. T. Junghanns; R. E. Kneusel; D. Groger; U. Matern, *Phytochemistry* **1998**, *49*, 403–411.
54. K. Springob; R. Lukacin; C. Ernwein; I. Groning; U. Matern, *Eur. J. Biochem.* **2000**, *267*, 6552–6559.
55. K. Wanibuchi; P. Zhang; T. Abe; H. Morita; T. Kohno; G. Chen; H. Noguchi; I. Abe, *FEBS J.* **2007**, *274*, 1073–1082.
56. A. B. Christensen; P. L. Gregersen; J. Schröder; D. B. Collinge, *Plant Mol. Biol.* **1998**, *37*, 849–857.
57. A. Schoppner; H. Kindl, *J. Biol. Chem.* **1984**, *259*, 6806–6811.
58. S. Raiber; G. Schröder; J. Schröder, *FEBS Lett.* **1995**, *361*, 299–302.
59. S. Schanz; G. Schröder; J. Schröder, *FEBS Lett.* **1992**, *313*, 71–74.
60. I. Abe, *Chem. Pharm. Bull. (Tokyo)* **2008**, *56*, 1505–1514.
61. K. Springob; S. Samappito; A. Jindaprasert; J. Schmidt; J. E. Page; W. De-Eknamkul; T. M. Kutchan, *FEBS J.* **2007**, *274*, 406–417.
62. A. Jindaprasert; K. Springob; J. Schmidt; W. De-Eknamkul; T. M. Kutchan, *Phytochemistry* **2008**, *69*, 3043–3053.
63. Y. Mizuuchi; S. P. Shi; K. Wanibuchi; A. Kojima; H. Morita; H. Noguchi; I. Abe, *FEBS J.* **2009**, *276*, 2391–2401.
64. F. Kreuzaler; K. Hahlbrock, *Arch. Biochem. Biophys.* **1975**, *169*, 84–90.
65. F. Kreuzaler; K. Hahlbrock, *FEBS Lett.* **1972**, *28*, 69–72.
66. W. Heller; K. Hahlbrock, *Arch. Biochem. Biophys.* **1980**, *200*, 617–619.
67. R. Sutfeld; R. Wiermann, *Arch. Biochem. Biophys.* **1980**, *201*, 64–72.
68. F. Kreuzaler; K. Hahlbrock, *Eur. J. Biochem.* **1975**, *56*, 205–213.
69. T. Lanz; S. Tropf; F. J. Marner; J. Schröder; G. Schröder, *J. Biol. Chem.* **1991**, *266*, 9971–9976.
70. F. Kreuzaler; H. Ragg; W. Heller; R. Tesch; I. Witt; D. Hammer; K. Hahlbrock, *Eur. J. Biochem.* **1979**, *99*, 89–96.
71. I. Abe; H. Morita; A. Nomura; H. Noguchi, *J. Am. Chem. Soc.* **2000**, *122*, 11242.
72. H. Morita; Y. Takahashi; H. Noguchi; I. Abe, *Biochem. Biophys. Res. Commun.* **2000**, *279*, 190–195.
73. K. W. Zuurbier; J. Leser; T. Berger; A. J. Hofte; G. Schröder; R. Verpoorte; J. Schröder, *Phytochemistry* **1998**, *49*, 1945–1951.
74. H. Oku; S. Ouchi; T. Shiraishi; Y. Komoto; K. Oki, *Ann. Phytopathol. Soc. Jpn.* **1975**, *41*, 185–191.
75. L. M. Vieira; A. Kijjoa, *Curr. Med. Chem.* **2005**, *12*, 2413–2446.
76. R. W. Fuller; J. W. Blunt; J. L. Boswell; J. H. Cardellina, 2nd; M. R. Boyd; *J. Nat. Prod.* **1999**, *62*, 130–132.
77. L. H. Hu; K. Y. Sim, *Tetrahedron* **2000**, *56*, 1379–1386.
78. W. Schmidt; L. Beerhues, *FEBS Lett.* **1997**, *420*, 143–146.
79. M. Iinuma; H. Tosa; T. Tanaka; F. Asai; Y. Kobayashi; R. Shimano; K. Miyauchi, *J. Pharm. Pharmacol.* **1996**, *48*, 861–865.
80. Y. Kwok; L. H. Hurley, *J. Biol. Chem.* **1998**, *273*, 33020–33026.
81. A. Urbain; A. Marston; E. F. Queiroz; K. Ndjoko; K. Hostettmann, *Planta Med.* **2004**, *70*, 1011–1014.
82. J. P. Michael, *Nat. Prod. Rep.* **2008**, *25*, 166–187.
83. Y. Shomura; I. Torayama; D. Y. Suh; T. Xiang; A. Kita; U. Sankawa; K. Miki, *Proteins* **2005**, *60*, 803–806.
84. R. Hain; H. J. Reif; E. Krause; R. Langebartels; H. Kindl; B. Vornam; W. Wiese; E. Schmelzer; P. H. Schreier; R. H. Stocker, *et al.*, *Nature* **1993**, *361*, 153–156.
85. B. D. Gehm; J. M. McAndrews; P. Y. Chien; J. L. Jameson, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 14138–14143.
86. H. Morita; H. Noguchi; J. Schroder; I. Abe, *Eur. J. Biochem.* **2001**, *268*, 3759–3766.
87. R. Gehlert; H. Kindl, *Phytochemistry* **1996**, *30*, 457–460.
88. B. Liu; T. Beuerle; T. Klundt; L. Beerhues, *Planta* **2004**, *218*, 492–496.
89. R. Mechoulam; S. Ben-Shabat, *Nat. Prod. Rep.* **1999**, *16*, 131–143.
90. M. Fellermeier; W. Eisenreich; A. Bacher; M. H. Zenk, *Eur. J. Biochem.* **2001**, *268*, 1596–1604.

91. C. Taguchi; F. Taura; T. Tamada; Y. Shoyama; H. Tanaka; R. Kuroki; S. Morimoto; *Acta Crystallogr. Sect F Struct. Biol. Cryst. Commun.* **2008**, *64*, 217–220.
92. P. M. Dewick, *Medicinal Natural Products*, second ed.; Wiley: West Sussex, 2002.
93. S. Shibutani; T. Nagasawa; H. Oura; G. Nonaka; I. Nishioka, *Chem. Pharm. Bull. (Tokyo)* **1981**, *29*, 874–878.
94. M. Nishizawa; T. Yamagishi; G. Nonaka; I. Nishioka; T. Nagasawa; H. Oura, *Chem. Pharm. Bull. (Tokyo)* **1983**, *31*, 2593–2600.
95. W. Borejsza-Wysocki; G. Hrazdina, *Phytochemistry* **1994**, *35*, 623–628.
96. R. K. Maheshwari; A. K. Singh; J. Gaddipati; R. C. Srimal, *Life Sci.* **2006**, *78*, 2081–2087.
97. P. Anand; A. B. Kunnumakkara; R. A. Newman; B. B. Aggarwal, *Mol. Pharm.* **2007**, *4*, 807–818.
98. S. Shishodia; M. M. Chaturvedi; B. B. Aggarwal, *Curr. Prob. Cancer* **2007**, *31*, 243–305.
99. P. J. Roughley; D. A. Whiting, *J. Chem. Soc. Perkin. Trans. I* **1973**, 2379–2388.
100. P. Denniff; I. Macleod; D. A. Whiting, *J. Chem. Soc. Perkin. Trans. I* **1980**, 2637–2644.
101. N. Funa; H. Ozawa; A. Hirata; S. Horinouchi, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 6356–6361.
102. M. Funabashi; N. Funa; S. J. Horinouchi, *Biol. Chem.* **2008**, *283*, 13983–13991.
103. H. Morita; S. Kondo; T. Abe; H. Noguchi; S. Sugio; I. Abe; T. Kohno, *Acta Crystallogr. F Biol. Crystallogr.* **2006**, *62*, 899–901.
104. Y. Kashiwada; G. Nonaka; I. Nishioka, *Phytochemistry* **1990**, *29*, 1007–1009.
105. D. K. Holdsworth, *Planta Med.* **1972**, *22*, 54–58.
106. T. Suga; T. Hirata, *Bull. Chem. Soc. Jpn.* **1978**, *51*, 872–877.
107. T.-W. Yu; Y. Shen; R. McDaniel; H. Floss; C. Khosla; D. Hopwood; B. Moore, *J. Am. Chem. Soc.* **1998**, *120*, 7749–7759.
108. R. Durand; M. H. Zenk, *Tetrahedron Lett.* **1971**, *12*, 3009–3012.
109. R. Durand; M. H. Zenk, *FEBS Lett.* **1974**, *39*, 218–220.
110. H. Fu; D. A. Hopwood; C. Khosla, *Chem. Biol.* **1994**, *1*, 205–210.
111. T. Hirata; T. Suga, *Z. Naturforsch. [C]* **1977**, *32*, 731–734.
112. A. Kirakosyan; H. Hayashi; K. Inoue; A. Charchoglyan; H. Vardapetyan, *Phytochemistry* **2000**, *53*, 345–348.
113. J. Barnes; L. A. Anderson; J. D. Phillipson, *J. Pharm. Pharmacol.* **2001**, *53*, 583–600.
114. S. P. Shi; K. Wanibuchi; H. Morita; K. Endo; H. Noguchi; I. Abe, *Org. Lett.* **2009**, *11*, 551–554.
115. T. P. Nicholson; C. Winfield; J. Westcott; J. Crosby; T. J. Simpson; R. Cox, *J. Chem. Commun. (Camb)* **2003**, *21*, 686–687.
116. I. Abe; Y. Takahashi; H. Noguchi, *Org. Lett.* **2002**, *4*, 3623–3626.
117. I. Abe; Y. Takahashi; W. Lou; H. Noguchi, *Org. Lett.* **2003**, *5*, 1277–1280.
118. T. Abe; H. Noma; H. Noguchi; I. Abe, *Tetrahedron Lett.* **2006**, *47*, 8727–8730.
119. J. M. Jez; J. P. Noel, *J. Biol. Chem.* **2000**, *275*, 39640–39646.
120. J. M. Jez; J. L. Ferrer; M. E. Bowman; R. A. Dixon; J. P. Noel, *Biochemistry* **2000**, *39*, 890–902.
121. J. M. Jez; M. E. Bowman; J. P. Noel, *Biochemistry* **2001**, *40*, 14829–14838.
122. I. Abe; T. Watanabe; H. Morita; T. Kohno; H. Noguchi, *Org. Lett.* **2006**, *8*, 499–502.
123. I. Abe; T. Watanabe; W. Lou; H. Noguchi, *FEBS J.* **2006**, *273*, 208–218.
124. H. Morita; S. Kondo; R. Kato; K. Wanibuchi; H. Noguchi; S. Sugio; I. Abe; T. Kohno, *Acta Crystallogr. Sect F Struct. Biol. Cryst. Commun.* **2007**, *63*, 947–949.
125. J. M. Jez; M. E. Bowman; J. P. Noel, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 5319–5324.
126. I. Abe; Y. Sano; Y. Takahashi; H. Noguchi, *J. Biol. Chem.* **2003**, *278*, 25218–25226.
127. R. Lukacin; S. Schreiner; U. Matern, *FEBS Lett.* **2001**, *508*, 413–417.
128. R. Lukacin; S. Schreiner; K. Silber; U. Matern, *Phytochemistry* **2005**, *66*, 277–284.
129. T. M. Harris; R. L. Carney, *J. Am. Chem. Soc.* **1966**, *88*, 2053–2054.
130. I. Abe; H. Morita; S. Oguro; H. Noma; K. Wanibuchi; N. Kawahara; Y. Goda; H. Noguchi; T. Kohno, *J. Am. Chem. Soc.* **2007**, *129*, 5976–5980.
131. R. S. Gokhale; S. Y. Tsuji; D. E. Cane; C. Khosla, *Science* **1999**, *284*, 482–485.
132. C. Khosla; P. B. Harbury, *Nature* **2001**, *409*, 247–252.
133. B. A. Pfeifer; S. J. Admiraal; H. Gramajo; D. E. Cane; C. Khosla, *Science* **2001**, *291*, 1790–1792.
134. E. I. Hwang; M. Kaneko; Y. Ohnishi; S. Horinouchi, *Appl. Environ. Microbiol.* **2003**, *69*, 2699–2706.
135. M. Kaneko; E. I. Hwang; Y. Ohnishi; S. Horinouchi, *J. Ind. Microbiol. Biotechnol.* **2003**, *30*, 456–461.
136. I. Miyahisa; N. Funa; Y. Ohnishi; S. Martens; T. Moriguchi; S. Horinouchi, *Appl. Microbiol. Biotechnol.* **2006**, *71*, 53–58.
137. Y. Katsuyama; N. Funa; I. Miyahisa; S. Horinouchi, *Chem. Biol.* **2007**, *14*, 613–621.
138. Y. Katsuyama; I. Miyahisa; N. Funa; S. Horinouchi, *Appl. Microbiol. Biotechnol.* **2007**, *73*, 1143–1149.
139. J. A. Chemeler; Y. J. Yanm; E. Leonard; M. A. G. Koffas, *Org. Lett.* **2007**, *9*, 1855–1858.
140. Y. Zhang; S. Z. Li; J. Li; X. Pan; R. E. Cahoon; J. G. Jaworski; X. Wang; J. M. Jez; F. Chen; O. Yu, *J. Am. Chem. Soc.* **2006**, *128*, 13030–13031.
141. S. J. Kwon; M. Y. Lee; B. Ku; D. H. Sherman; J. S. Dordick, *ACS Chem. Biol.* **2007**, *2*, 419–425.

Biographical Sketches

Hiroyuki Morita graduated from Tokyo University of Pharmacy and Life Sciences in 1996, where he studied the organic synthesis of lignan analogs. He then obtained his Ph.D. in 2001 from the University of Shizuoka, School of Pharmaceutical Sciences under the direction of Professor Hiroshi Noguchi, where he studied chemistry and biochemistry of natural products biosynthesis. After 1 year postdoctoral research with Professor John C. Vederas at University of Alberta (2001–2002), he investigated the structural enzymology of protein kinases and type III polyketide synthases as a postdoctoral fellow at Mitsubishi Chemical Corporation (2002–2004) and Mitsubishi Kagaku Institute of Life Sciences (2004–2008), and as an assistant professor in University of Shizuoka, School of Pharmaceutical Sciences (2008–2009). Now he is an assistant professor in the University of Tokyo. His research interests include structure-based engineering of a plant type III polyketide synthase.

Ikuro Abe graduated from the University of Tokyo in 1984, and obtained his Ph.D. degree in 1989 from the same university under the direction of Professor Yutaka Ebizuka, where he studied chemistry and biochemistry of natural products biosynthesis. After 2 years postdoctral research with Professor Guy Ourisson at the CNRS Institut de Chimie des Substances Naturelles, and mostly with Professor Michel Rohmer at the Ecole Nationale Supérieure de Chimie de Mulhouse (1989–91), he moved to the United States to work with Professor Glenn D. Prestwich at the State University of New York at Stony Brook (1991–96) and The University of Utah (1996–98) as a Research Assistant Professor. In 1998, he returned to Japan at University of Shizuoka, School of Pharmaceutical Sciences, and was an investigator of PRESTO, Japan Science and Technology Agency (2005-2009). In 2009, he moved back to The University of Tokyo as Professor of Natural Products Chemistry. His research interests involve exploring and engineering the biosynthesis of natural products.

Hiroshi Noguchi graduated from the University of Tokyo in 1975, and obtained his Ph.D. degree in 1980 from the same university under the direction of Professor Ushio Sankawa, where he studied pharmacognosy and biosynthesis of natural products. He became research associate at the University of Tokyo in 1979 and worked as a postdoctoral fellow at Alberta University with Professor John Vederas in 1984–85, and at Brown University with Professor David Cane in 1985–86. He was promoted to lecturer in 1994, and then moved to the University of Shizuoka in 1995 as Professor of Pharmacognosy. His research interests are the chemical biology of bioactive natural products and the biosynthesis of natural products.

1.07 Type II PKS

Jürgen Rohr, University of Kentucky, Lexington, KY, USA

Christian Hertweck, Leibniz Institute for Natural Product Research and Infection Biology, HKI, Jena, Germany

© 2010 Elsevier Ltd. All rights reserved.

1.07.1	**Introduction**	228
1.07.2	**Enzymes of Type II PKS-Derived Natural Products**	228
1.07.2.1	Priming, Polyketide Chain Assembly, and Enzymes Responsible for the Formation of the Primary Framework	228
1.07.2.1.1	PKS priming	228
1.07.2.1.2	Chain assembly and control of chain length	232
1.07.2.1.3	Cyclizations and ring topologies	235
1.07.2.2	Polyketide Frame Modifying and Decorating Enzymes (Tailoring Enzymes)	236
1.07.2.2.1	Group transferases	236
1.07.2.2.2	Oxidoreductases	236
1.07.2.2.3	Halogenases	238
1.07.3	**Natural Structural Diversity**	239
1.07.3.1	Natural Products Derived from Short Polyketides	239
1.07.3.1.1	Benzoisochromanequinones	239
1.07.3.1.2	Other octaketide derivatives	239
1.07.3.2	Natural Products Derived from Decaketides that Undergo Initial 7,12-Cyclization	239
1.07.3.2.1	Anthracycline folding	242
1.07.3.2.2	Tetracycline folding	245
1.07.3.2.3	Angucycline folding	248
1.07.3.3	Natural Products Derived from Decaketides that Undergo Initial 9,14-Cyclization	262
1.07.3.3.1	Tetracenomycin folding	262
1.07.3.3.2	Discoid folding	262
1.07.3.4	Natural Products Derived from Larger Polyketides	266
1.07.3.4.1	Angucylines from angular heterocyclic folding	269
1.07.4	**Artificial Structural Diversity through Combinatorial Biosynthesis**	270
1.07.4.1	'Unnatural Natural Products' Generated by Manipulation of Type II Polyketide Synthases	272
1.07.4.1.1	New aromatic polyketides by manipulation of the PKS and PKS-associated enzymes	272
1.07.4.1.2	Artificial aromatic polyketides through utilization of enzymes involved in the priming process	273
1.07.4.1.3	Outlook	280
1.07.4.2	Derivatives of Natural Products Generated by Recombination with Genes Encoding Post-PKS Tailoring Enzymes	280
1.07.4.2.1	Exploiting deoxysugar pathways and glycosyltransferases	280
1.07.4.2.2	Exploiting group transferases other than glycosyltransferases	287
1.07.4.2.3	Exploiting oxidoreductases	287
1.07.5	**Conclusions**	292
References		293

1.07.1 Introduction

Type II polyketide synthases (type II PKSs) generate the core of many natural products, those produced mainly by bacteria. The versatility of these enzyme complexes, along with a broad array of often uniquely operating post-PKS tailoring enzymes, makes the group of type II PKS-derived secondary metabolites one of the most structurally diverse groups found in nature. Usually, the structures contain a multicyclic, aromatic core, and are also referred to as multicyclic or polycyclic aromatic polyketides, or sometimes simply as polyphenols. Almost all of these compounds are biologically active, the activity spectrum reaching from antibacterial, anticancer, antifungal, and antiviral to special activities, such as inhibition of platelet aggregation, and is typically achieved through enzyme inhibition, DNA binding, DNA damaging, or receptor antagonism. Thus, it is not surprising that many of these drugs and their derivatives are clinically used, well-known examples being the tetracyclines as antibacterial drugs (antibiotics), or doxorubicin and mithramycin as anticancer drugs.[1–3]

1.07.2 Enzymes of Type II PKS-Derived Natural Products

During the last two decades, an impressive number of almost 40 biosynthesis gene clusters coding for type II PKS pathways have been localized, cloned, and in most cases fully sequenced. The identity of the gene clusters has been verified almost exclusively through gene inactivation and/or heterologous expression. From the targeted inactivation, cross-complementation, and gene swapping experiments, *in vitro* biochemical and structural analysis, and chemical structural elucidation, a large body of knowledge has been gained on how aromatic polyketide assembly is controlled and how the multienzyme complexes function in generating structural diversity (**Table 1**).

1.07.2.1 Priming, Polyketide Chain Assembly, and Enzymes Responsible for the Formation of the Primary Framework

A hallmark of type II PKS is the so-called 'minimal PKS' consisting of two ketosynthase units (KS_α and KS_β, or chain length factor (CLF)), and an acyl-carrier protein (ACP), which serves as an anchor for the growing polyketide chain.[54] This minimal biosynthetic machinery is usually sufficient for the iterative condensation of malonyl extender units to an acyl starter unit to yield a polyketide chain of a defined length (**Scheme 1**). In type II PKS systems, only malonyl CoA extender units are employed, which is in marked contrast to modular polyketide synthases that are known to accept alternative building blocks such as methylmalonyl or methoxymalonyl.[3,55,56] Transfer of the malonyl extender units may involve a potential fourth PKS component, a malonyl-CoA:ACP transferase (MCAT). Due to the absence of MCAT genes in most type II PKS gene clusters, it was proposed that malonylation activity is recruited from fatty acid biosynthesis.[57–60] An alternative model was proposed by Simpson and coworkers, who demonstrated that the ACP is capable of self-malonylation *in vitro*, albeit only at elevated concentrations of malonyl-CoA.[61–63]

Apart from the minimal PKS, further accessory components are required for loading alternative starter units, getting the nascent chain into shape, and tailoring the basic carbon skeleton.

1.07.2.1.1 PKS priming

A large majority of type II PKSs generate acetate-primed polyketide chains. Leadlay and coworkers have demonstrated for actinorhodin and tetracenomycin biosynthesis that acetyl-*S*-KS actually results from decarboxylation of a malonyl unit. The decarboxylase activity resides within the KS_β, which is mechanistically reminiscent of the KS_Q loading domain of modular PKS.[64] Nonetheless, various alternative primers can be employed by type II PKS. In most cases, these starters are linear or branched short-chain fatty acids, and only rarely are malonate or benzoate starters seen.[65] Two principal alternative priming strategies have been postulated to introduce such nonacetate units: priming similar to fatty acid synthase (FAS), priming using a FabH homologue, or a nonribosomal peptide synthetase (NRPS)-like mechanism.

Table 1 Analyzed type II PKS gene loci

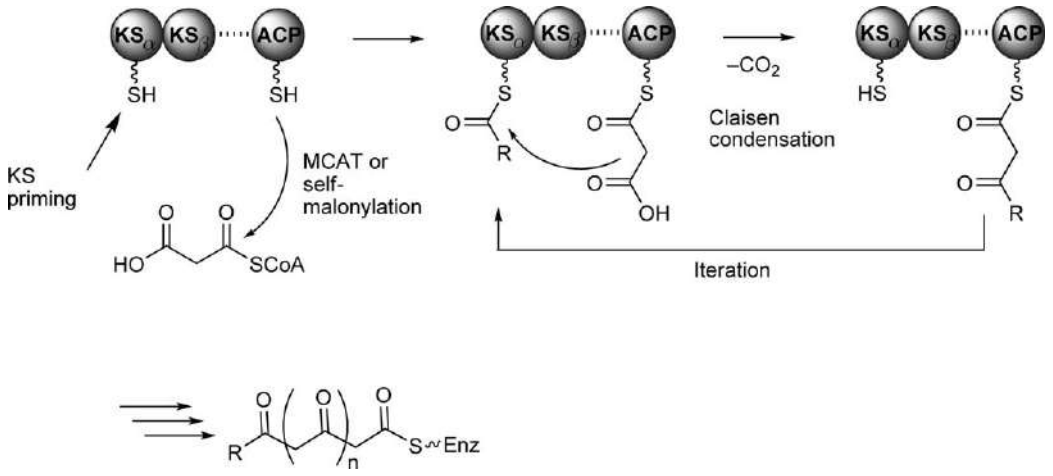
Metabolite	*Abbreviation*	*Type*	*Strain/source*	*Year*	*References*
Actinorhodin	Act	BIQ	Streptomyces coelicolor	1986/1992	4,5
Tetracenomycin	Tcm	TCM (DEC)	Streptomyces glaucescens	1987	6
Granaticin	Gra	BIQ	Streptomyces violaceoruber Tü22	1989	7
Oxytetracycline	Otc or Oxy	TET (DEC)	Streptomyces rimosus	1989/2006	8,9
Jadomycin	Jad	ANG (DEC)	Streptomyces venezuelae	1994	10
Griseusin	Gris	BIQ	Streptomyces griseus	1994	11
Daunorubicin	Dnr	ANT (DEC)	Streptomyces sp. (C5)	1994	12, partial
Doxorubicin	Dox	ANT (DEC)	Streptomyces peucetius	1994	13, partial
Frenolicin	Fren	BIQ	Streptomyces roseofulvus	1994	14
Urdamycin	Urd	ANG (DEC)	Streptomyces fradiae (Tü2717)	1995	15, partial
Elloramycin	Ell	TCM (DEC)	Streptomyces olivaceus (Tü2353)	1995	16, partial
WhiE spore pigment (unknown structure, known shunt products)	WhiE	SPP	Streptomyces coelicolor	1995	17
Mithramycin	Mtm	TET (AUR)	Streptomyces argillaceus	1996	18
Pradimicin	Prm/Pdm	PEN	Actinomadura hibisca (P157-2)	1997 (partial), 2007	19–21
Kinamycin	Kin	ANG (DEC)	Streptomyces murayamaensis	1998	22
Landomycin	Lan	ANG (DEC)	Streptomyces cyanogenus (S136)	1999	23
Naphthocyclinone	Ncn	BIQ	Streptomyces arenae	1999	24
Enterocin	Enc	OCT	Streptomyces maritimus	2000	25
R1128	Zhu	BIQ	Streptomyces sp. (R1128)	2000	26
Nogalamycin	Sno	ANT (DEC)	Streptomyces nogalater	2001	27
Rubromycin	Rub	PEN	Streptomyces collinus	2001 (only db)	28
Aklacinomycin	Akn	ANT (DEC)	Streptomyces galilaeus	2002	29
Griseorhodin	Grh	PEN	Streptomyces sp. (JP95)	2002	30
Auricin (unknown structure)	Aur	ANG?	Streptomyces aureofaciens	2002	31
Medermycin	Med	BIQ	Streptomyces sp. (AM-7161)	2003	32
Gilvocarcin	Gil	ANG (DEC)	Streptomyces griseoflavus	2003	33
Hedamycin	Hed	PLU	Streptomyces griseoruber	2004	34
Chromomycin	Cmm	TET (AUR)	Streptomyces griseus	2004	35
Resistomycin	Rem	DIS (DEC)	Streptomyces resistomycificus	2004	36
Spore pigment (unknown structure)	WhiESa	SPP	Streptomyces aureofaciens (CCM 3239)	2004	37

(*Continued*)

Table 1 (Continued)

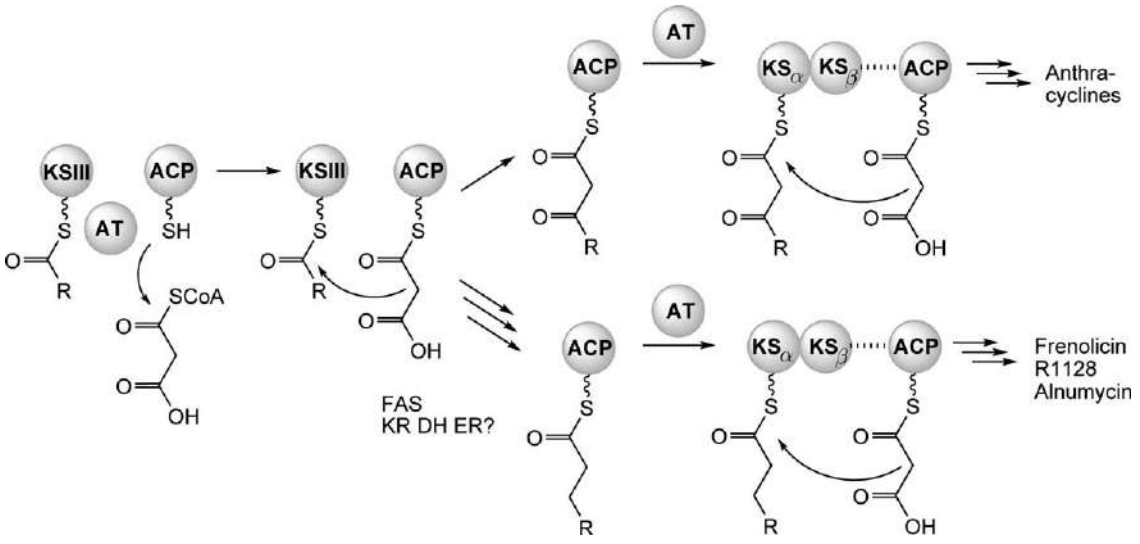
Metabolite	*Abbreviation*	*Type*	*Strain/source*	*Year*	*References*
Unknown (unknown structure)	Aur2	?	Streptomyces aureofaciens	2004	38
Cervimycin	Cer	TET (DEC)	Streptomyces tendae	2004 (partial)	39
Gaudimycin (cryptic angucycline)	Pga	ANG (DEC)	Streptomyces sp. PGA64	2004, 2007	40,41
Gaudimycin (cryptic angucycline)	Cab	ANG (DEC)	Streptomyces sp. HO21	2004, 2007	40,41
Chartreusin	Cha	ANT (DEC)	Streptomyces chartreusis	2005	42
Oviedomycin	Ovi	ANG (DEC)	Streptomyces antibioticus	2005	43
Fredericamycin	Fdm	PEN	Streptomyces griseus	2005	44
Alpomycin (unknown structure)	Alp	?	Streptomyces ambofaciens	2005 (partial)	45
Steffimycin	Stf	ANT (DEC)	Streptomyces steffisburgensis	2006	46
Unknown (unknown structure)	?	?	Streptomyces toxytricini NRRL 15,443	2006	47
Sch 47554, 47555 (unknown structure)	Sch	SPP	Streptomyces sp. (SCC-2136)	2006	48
Benastatin	Ben	PEN	Streptomyces sp.	2007	49
Aurachin	Aua	X	Stigmatella aurantiaca	2007	50
Anthraquinones (AQ-256, etc.)	Ant	X	Photorhabdus luminescens	2007	51
Lactanomycin (unknown structure)	Lct/Lcz	?	Streptomyces rishiriensis/Streptomyces sanglieri	2008	52
Alnumycin	Aln	BIQ	Streptomyces sp. (CM020)	2008	53

ANG, angucyclic; ANT, anthracyclic; AUR, aureolic acid; BIQ, benzoisochromanequinone; DEC, decaketide; DIS, discoid; OCT, misc. octaketide; PEN, pentangular polyketides; PLU, pluramycin-type; SPP, spore pigment; TCM, tetracenomycin; TET, tetracycline; X, other.

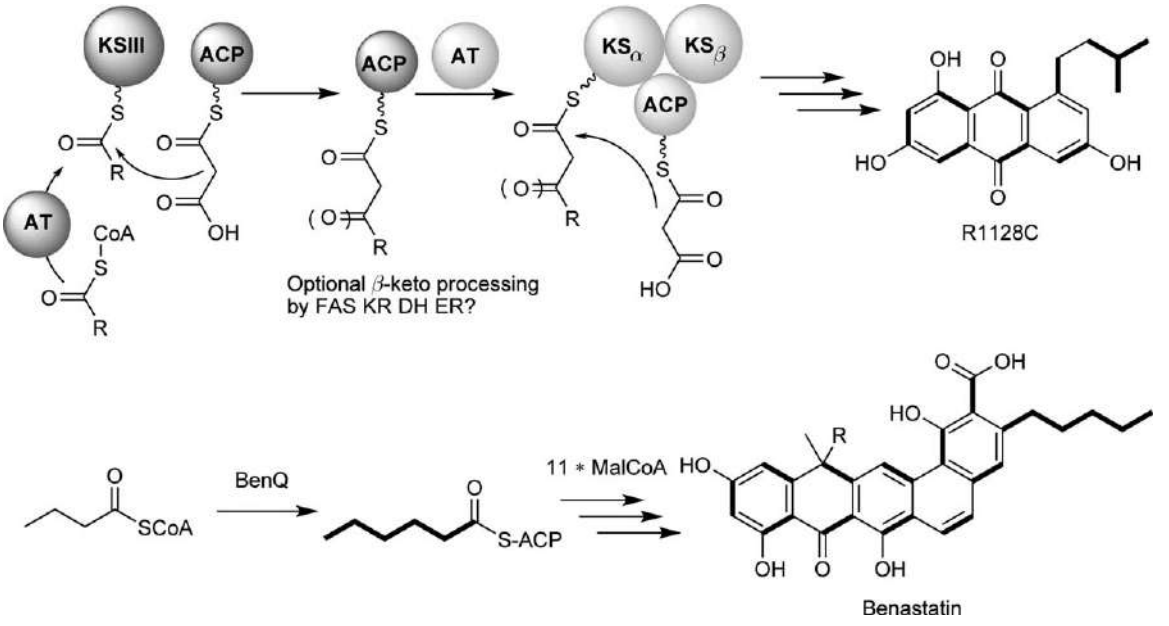


Scheme 1 Basic mechanisms in polyketide formation by a (minimal) type II PKS.

Two additional PKS components encoded in the daunorubicin (*dnr*) and doxorubicin (*dox*) gene clusters were implicated in starter unit selection and chain initiation; gave a first clue on how propionate is loaded onto the PKS.[12,66] Functional analyses revealed that a FabH-like KSIII (DpsC)[13] and an AT (DpsD) are essential in selecting, processing, and loading the correct primer. Common to all KSIII is a catalytic triad composed of Cys112, His244, and Asn274 (ecFabH numbering) in the active center that performs the transacylation, decarboxylation, and condensation reactions.[67] Accordingly, DpsC functions as a KSIII, catalyzing the first condensation of malonyl-CoA with propionyl-CoA. The resulting β-keto thioester is transferred onto the minimal PKS by DpsD.[12,13] In the absence of DpsC, the *dps* PKS exhibits a relaxed starter unit specificity and also employs acetyl-CoA. However, propionate-derived polyketides may also be formed.[68,69] *In vitro* DpsC is indeed required for the choice of the correct starter unit for daunorubicin biosynthesis.[70] Genes coding for KSIII components have been identified in various other type II gene clusters that are involved in the biosynthesis of fatty acid-derived polyphenols, such as frenolicin,[14] R1128,[26] benastatin,[49] and alnumycin.[53] The role of KSIII in priming has also been implicated in the hedamycin[34] and fredericamycin[44] pathways. In some cases, accessory proteins for priming have been identified. In frenolicin (*frn*) biosynthesis, PKS priming involves not only the FabH orthologue (FrnI) but also an additional ACP (FrnJ) for tethering the starter.[14] The ACP-bound β-keto thioester is likely processed to the fully saturated butyryl moiety by FAS processing enzymes (**Scheme 2**).



Scheme 2 PKS priming involving KSIII components.



Scheme 3 PKS priming in the R1128 and benastatin pathways.

In the R1128 pathway, the situation is even more complex (**Scheme 3**). The *zhu* PKS accepts butyryl, valeryl, and 4-methylvaleryl starters. In addition to the KSIII (ZhuH), a malonyl AT (ZhuC) and an ACP (ZhuG) are required for the priming process.[26,71] It is important to note that the *zhu* ketosynthases involved in the initiation (ZhuH) and elongation (ZhuA) have orthogonal ACP specificity.[72] By recombination of the R1128 'loading module' consisting of ZhuC, ZhuH, and ZhuG with other minimal PKS components, novel aromatic polyketides were engineered.[73–75] The crystal structure of the priming KSIII (ZhuH) revealed critical gatekeeper residues that are relevant for primer selection.[76]

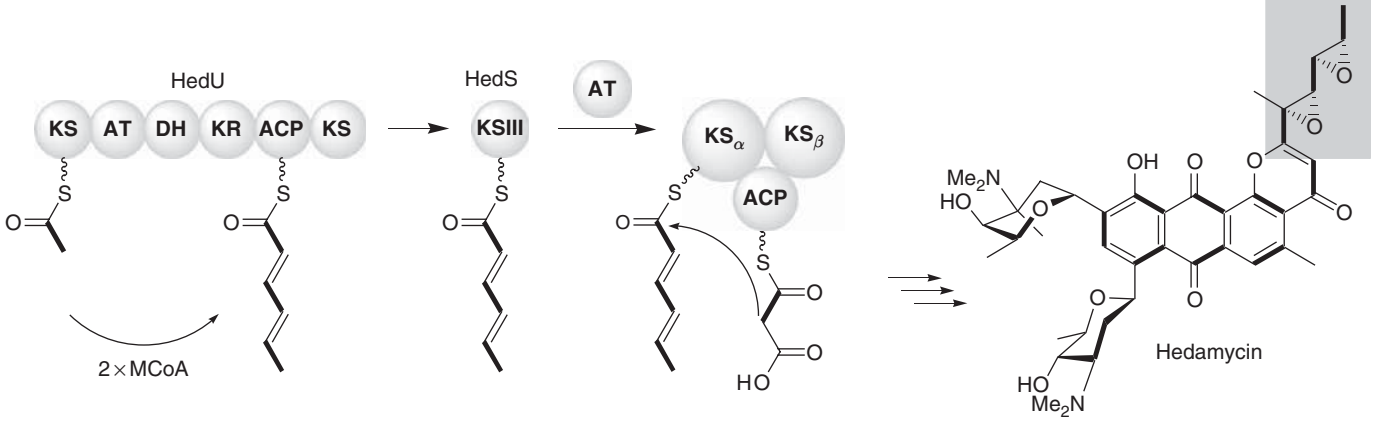
In contrast to the R1128 pathway, no designated priming ACP or AT genes were detected in the benastatin biosynthesis gene cluster (**Scheme 3**). Nonetheless, the KSIII (BenQ) alone controls hexanoate primer unit selection, as shown by mutational studies. In the absence of BenQ, various analogues with modified side chains are produced. (If a shorter fatty acid (butyrate) starter is incorporated, the length of the polyketide backbone is increased, resulting in the formation of an extended ring system that is reminiscent of the proposed intermediates in the griseorhodin and fredericamycin pathways.[49,77] The titer of the butyryl-derived compounds was increased 10-fold after exchanging BenQ with the alnumycin priming cassette.[77]

In the biosynthesis of hedamycin, an iterative type I PKS (HedT) is utilized for the formation of a short-chain starter unit (**Scheme 4**). A KSIII (HedS) and an AT (HedF) likely shuttle the hexadienoate starter from the type I (HedT) to the type II PKS (HedCDE).[34]

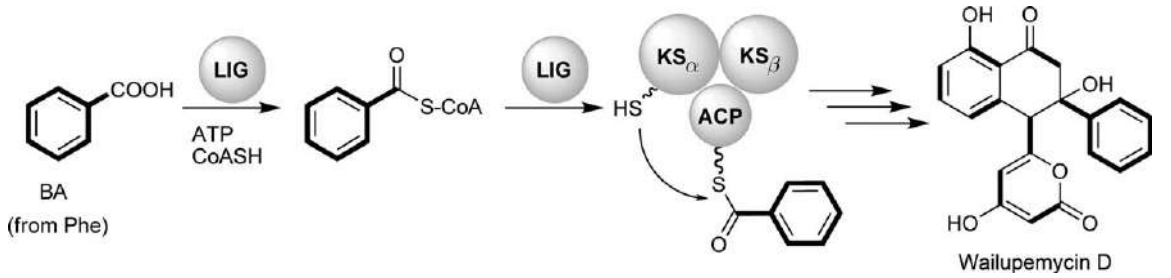
The aryl side chains of the enterocin and wailupemycins from the marine bacterium *Streptomyces maritimus* result from the incorporation of a rare phenylalanine-derived benzoyl starter unit (**Scheme 5**).[25,78,79] *In vivo* and *in vitro* work by Moore and coworkers demonstrated that a single ligase (EncN) is capable of adenylating benzoate,[80,81] activating it as CoA thioester, and transferring benzoyl-CoA onto the ACP (EncC).[82] Thus, this priming mechanism is reminiscent of the activation of NRPS building blocks. Blockage of primer biosynthesis and attachment allowed for the mutasynthesis of enterocins and wailupemycins.[80]

1.07.2.1.2 Chain assembly and control of chain length

The heart of any given type II PKS is the KS_α/KS_β heterodimer. Although both KS monomers show high amino acid sequence similarities, their functions diverge. In contrast to KS_α, which catalyzes the Claisen condensations, the KS_β component is not capable of performing this reaction because a crucial active site cysteine residue is missing. Apart from the abovementioned decarboxylation activity to yield the acetyl starter



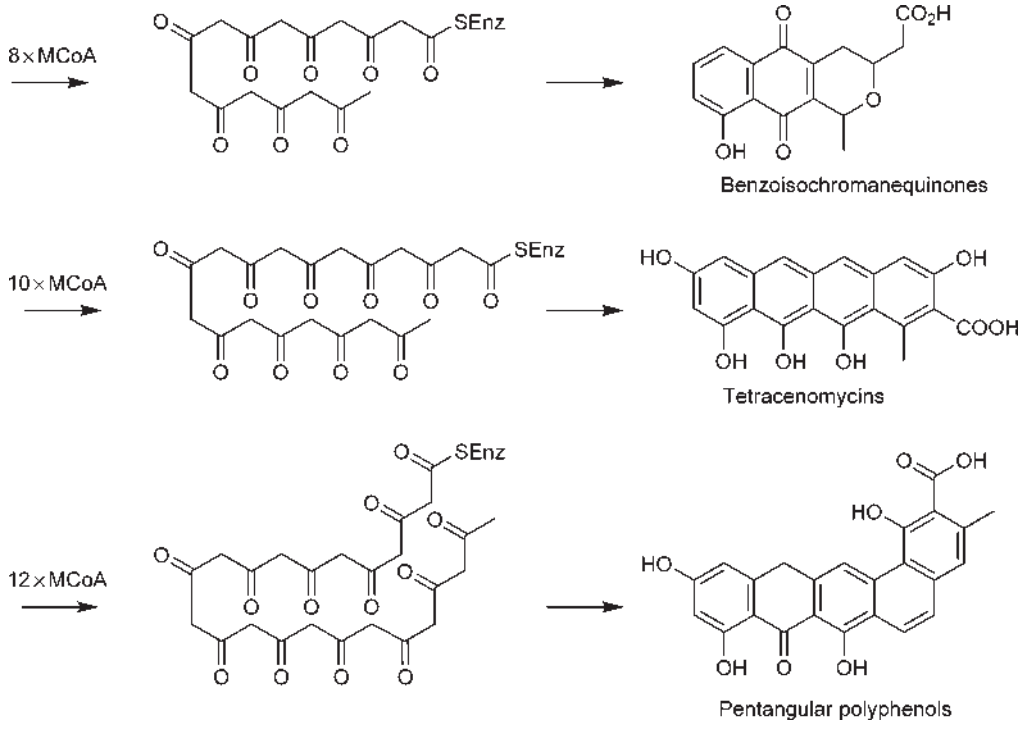
Scheme 4 Model for PKS priming in the hedamycin pathway.



Scheme 5 PKS priming with benzoate.

molecule from malonyl-CoA, the KS_β is also the primary determinant of carbon chain length.[83] The direct proof for a KS heterodimer was given by a crystal structure reported by Stroud, Khosla, and coworkers[84] showing that KS and CLF (or KS_β) have evolved highly complementary contacts. *In vivo* and *in vitro* experiments demonstrated the importance of the CLF in controlling chain length,[85] which is determined by measuring, not by counting.[86] The chain length of polyketides synthesized by type II PKS usually ranges between 16 (octaketides, such as actinorhodin), 20 (decaketides, such as tetracenomycin), and 24 (dodecaketides, such as pradimicin). The longest chains are assembled in the griseorhodin (tridecaketide), benastatin (tetradecaketide), and fredericamycin (pentadecaketide) pathways (**Scheme 6**).

Bioinformatic analyses and modeling studies revealed gatekeeper amino acid residues located at the interface of the dimer that define the chain length.[85,87,88] Experimental evidence for the role of the CLF was provided by swapping experiments and generation of chimeric genes that resulted in products with altered chain lengths.[85,88,89] Nonetheless, other type II PKS components can have an influence on the number of chain elongations.[17,90] While the polyketide chain is growing, the KS–CLF protein cleft keeps the nascent polyketide chain extended, thus separating the reactive ketide groups. It is likely that the highly reactive intermediates contact the KS–CLF heterodimer in the enol form.[84]

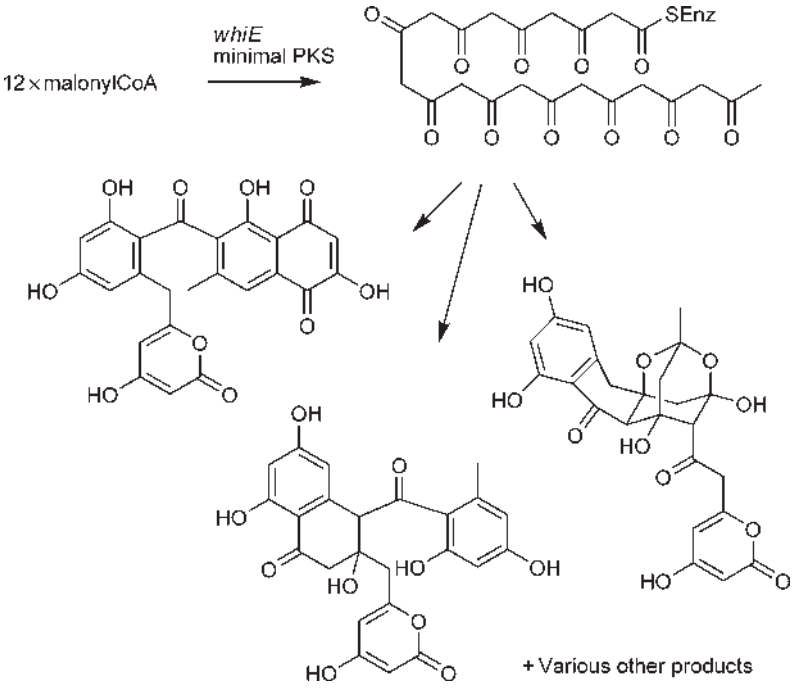


Scheme 6 Examples of octa-, deca-, and dodecaketides.

1.07.2.1.3 Cyclizations and ring topologies

After the final polyketide chain length has been achieved, the thioester is cleaved by a yet unknown enzymatic activity and the chain is released. In the absence of any further enzymes, the polyketide chain undergoes random aldol cyclocondensations and pyrone formation (**Scheme 7**).[17]

A controlled cyclization of the poly-β-keto chains into defined polyphenol structures is achieved only in the presence of cyclases, which exhibit a 'chaperone-like' function and catalyze specific aldol condensations. Ketoreductases aid in preforming the polyketide chain, and the so-called aromatases support the cyclodehydratation process. Typical primary products of type II PKS resulting from the concerted type II PKS action are polyphenols, which can be classified as linear tetracyclines, anthracyclines, benzoisochromanequinones (BIQs), tetracenomycins, aureolic acids, and the angular angucyclines as well as a group of pentangular polyphenols. Through systematic gene inactivation, recombination, and *in vitro* enzymatic studies, a large number of cyclases involved in the formation of polyphenolic ring topologies have been identified.[1] The first ring cyclization largely depends on the minimal PKS and whether or not a ketoreduction has taken place. However, in the biosynthesis of mithramycin and anthracyclines, folding of the polyketide chain is not dictated by the minimal polyketide synthase, but by ketoreduction and cyclases.[91] Also, the cyclase TcmN seems to play a crucial role in tetracenomycin biosynthesis in first ring formation.[92] Only few *in vitro* biochemical studies were conducted, including TcmF2 cyclase[93] and TcmN[94] from the tetracenomycin pathway, and the ester cyclases SnoaL[95] and AknH,[96] involved in nogalamycin and aclacinomycin biosynthesis, respectively. More recently, the crystal structure of the cyclase/aromatase region of TcmN (Tcm CYC/ARO) revealed a substrate-binding pocket with two critical residues (R69 and Y35) that are essential for promoting first- and second-ring cyclization specificity.[97] Previous gene inactivation and recombination experiments[2,54] elucidated the functions of the downstream cyclases TcmN, TcmI, and TcmJ, which were identified as second, third, and fourth ring cyclases in tetracenomycin biosynthesis.[54,93,94] The family of cyclases is extremely heterogeneous; nonetheless, they can be grouped into clades with particular functions on the basis of their amino acid sequences. Considering the plethora of aromatic polyketide structures known, it is quite remarkable that all of the basic polyketide structures are formed by a U-shaped folding of the poly-β-keto intermediates. Consequently, only a limited number or modes of cyclizations is realized, and virtually all polyphenols have a linear or angular



Scheme 7 Examples of polyketides resulting from spontaneous cyclization of a poly-β-keto intermediate.

architecture.[36,98] Clear exceptions from this biosynthetic scheme are the pentacyclic polyphenols resistomycin and resistoflavin. The multiple perifused rings result from an unparalleled S-shaped folding and cyclization of a decaketide.[36] Mutational analyses support a model according to which the 'discoid' ring system is shaped by a cage-like multienzyme complex and not by sequentially acting cyclases (**Scheme 8**).[99]

1.07.2.2 Polyketide Frame Modifying and Decorating Enzymes (Tailoring Enzymes)

1.07.2.2.1 Group transferases

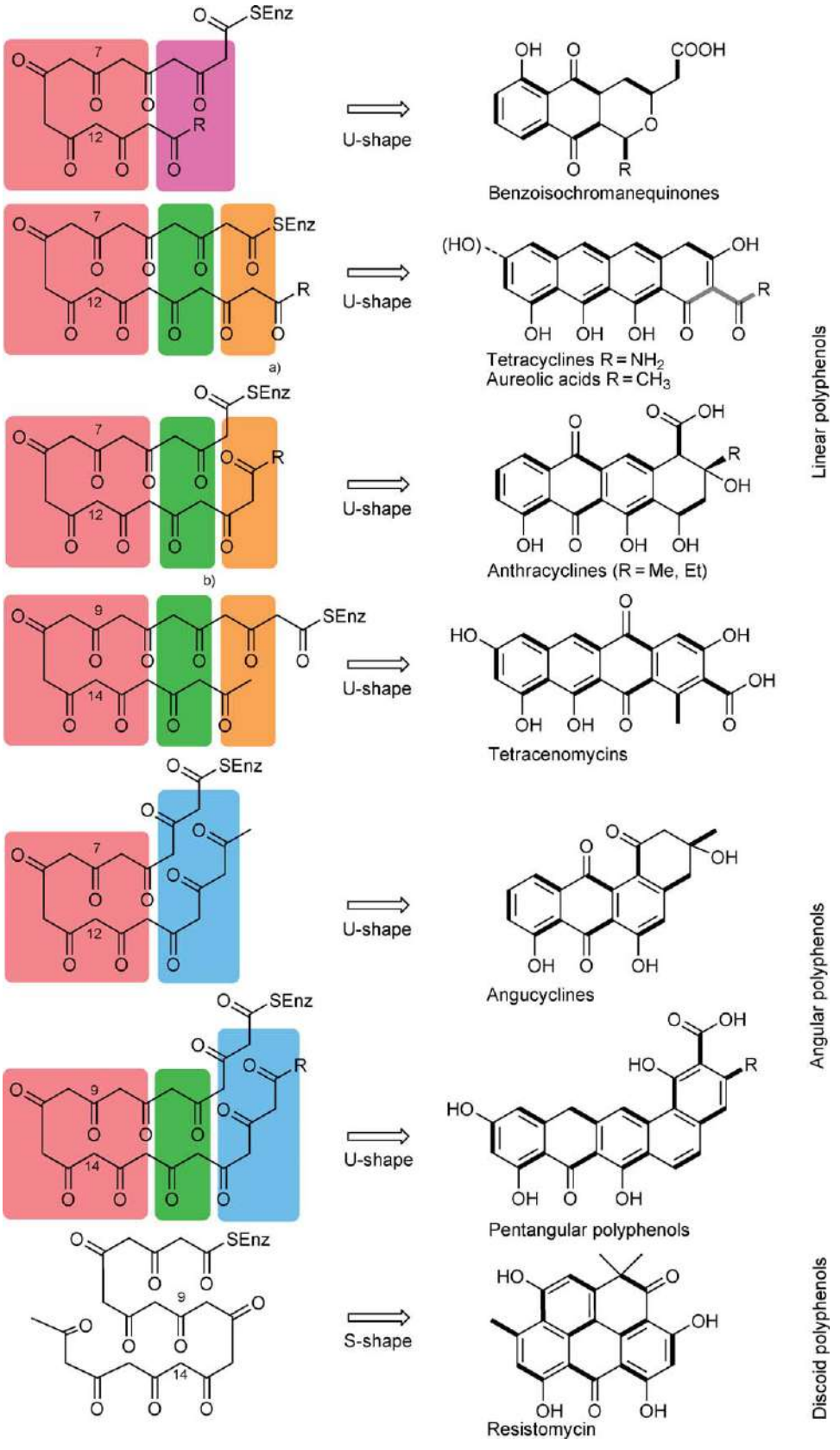
The term 'group transferase' refers to enzymes that catalyze the transfer of certain chemical groups, thereby introducing a new functional group. The functional groups introduced this way have both new reactivity and stereoelectronic impact on the substrate. This large group of enzymes contains important subgroups such as amino- and amidotransferases, alkyl (usually methyl) transferases, acyl (usually acetyl) transferases, glycosyltransferases (GTs), and kinases. Methyltransferases (MTs) and GTs are most important for tailoring type II PKS-derived scaffolds. Group transferases require cofactors. For MTs, this is usually *S*-adenosylmethionine (SAM), and most aminotransferases use alanine or glutamic acid as amino donors, and special pathways exist for their generation (including transferases that can utilize the inorganic nitrogen source ammonia).

MTs can decorate oxygen, nitrogen, sulfur, or carbon atoms with methyl groups. Examples of interesting and well-studied MTs in the context of type II PKS-derived natural products are the bifunctional cyclase/*O*-MT TcmN of the tetracenomycin pathway,[94] the C-bis-methylating *C*-MT BenF of the benastatin pathway,[100] the early acting *C*-MT OxyF and the N,N-dimethylating *N*-MT OxyT of the oxytetracycline pathway,[101,102] or the relatively late acting *C*-MT MtmMII of the mithramycin pathway,[103,104] which is essential for the biological activity and DNA–drug interaction. The pathways are discussed below in more detail.

GTs use nucleoside diphosphate (NDP)-activated sugars as donor substrates, and thymidine diphosphate (TDP) is by far the most common in the context of post-PKS modifications of type II PKS-derived compounds.[105–108] The NDP-activated sugars are sugar donor substrates or sugar cosubstrates. Sugar transfer leads to O-, N-, S-, or C-glycosidic modifications, O-glycosylations being by far the most abundant among type II PKS-derived natural products. However, there is a growing number of *C*-glycosidic compounds. These are mostly aryl-*C*-glycosides including classical angucyclines, pluramycins, gilvocarcins, anthracyclines (particularly of the nogalamycin type), and isochromane quinones.[109] For many sugar donor substrates, especially NDP-deoxysugars, special biosynthetic pathways exist, and the genes encoding these pathways are often clustered with the other genes of the metabolites, to whose biosynthesis they contribute.[107,110] GTs that are responsible for the attachment of sugar moieties add important features to the shape and the stereoelectronic properties of a molecule, which often essential for the biological activity of the natural product drugs. Many GTs possess unexpected inherent substrate flexibility, either toward their acceptor substrate or regarding their NDP–sugar donor substrates, and sometimes both; thus, GTs belong to the most versatile group of enzymes used for combinatorial biosynthesis.[105,106,111–113] GTs are classified on the basis of their substrate specificity and the type of reaction catalyzed or similarity to the amino acid sequence. Depending on the configuration of the anomeric functional group of the NDP-glycosyl donor substrate and the configuration of the resulting glycoside, all known GTs can be divided into (1) retaining GTs and (2) inverting GTs.[114–116] So far, all investigated GTs from the *Streptomyces* genus that are associated with secondary metabolism belong to the GT-1 family of inverting enzymes[117] and are the reason for Klyne's rule, which generally assigns α-configuration to L-sugars and β-configuration to D-sugars.[118] The advantage of a classification system based on structural features is its ability to predict the three-dimensional (3D) structure of GTs within the same family once the X-ray structure of one member is determined.[115,119–121] To date, various crystal structures of GTs involved in the secondary metabolism of microorganisms have been solved. Their 3D structures all fall into the GT-B fold.[122–127]

1.07.2.2.2 Oxidoreductases

The most frequently found post-PKS modifications are catalyzed by oxidoreductases, a very broad group of enzymes consisting of oxygenases, oxidases, peroxidases, reductases (e.g., ketoreductases), and dehydrogenases. In general, these enzymes can introduce oxygen-containing functionalities, such as hydroxy groups (hydroxylases), aldehyde or keto groups, and epoxides (epoxidases), or modify such functionalities by



Scheme 8 Overview of the polyketide cyclization patterns of bacterial type II PKS products.

adding or removing hydrogen atoms by transforming a ketone into a secondary alcohol or an aldehyde into a carboxylic acid, for example. All of these reactions have an impact on the stereoelectronic and physicochemical properties of the substrates by creating or removing chiral centers, introducing highly reactive functional groups, such as aldehydes or epoxides, changing the solubility of a molecule, or converting hydrogen bond donors (e.g., hydroxy groups) into hydrogen bond acceptors (e.g., keto groups) and vice versa. Furthermore, they can provide a handle for additional post-PKS modifications, for instance methyl- or glycosyltransfer. Thus, oxidoreductases can have an impact on the binding properties of a molecule with respect to a biological ligand molecule (receptor protein, enzyme, DNA, etc.), although they provide or modify usually only relatively small functional groups.[106] Enzyme nomenclature distinguishes between oxygenases (mono- and dioxygenases), oxidases, peroxidases, and dehydrogenases when an oxidation reaction is catalyzed.[128] The term 'oxygenase' refers to how a substrate molecule is modified by oxygen atoms. Oxygenases reductively activate oxygen for insertion into a substrate. Whereas a monooxygenase inserts one oxygen atom of molecular oxygen into a C–H bond of an organic substrate molecule, a dioxygenase incorporates both oxygen atoms of the dioxygen molecule into the substrate. The former results typically in a hydroxy group or an epoxide, and the latter results in a dioxetane or a peroxide, often highly reactive species that can rearrange or further react, for example, into a 1,2-diol or even undergo C–C bond cleavage into two carbonyl functions.[128,129] The terms 'oxidase' and 'peroxidase' only refer to what electron acceptor is used, that is, molecular oxygen in the case of oxidases and hydrogen peroxide in the case of peroxidases. Thus, both enzyme types can catalyze either the insertion of oxygen or a dehydrogenation reaction. Dehydrogenases remove two protons and two electrons from a substrate or act as reductases via the addition of two protons and two electrons to a substrate, as both directions of the reaction are catalyzed. All these enzymes are contained in a more or less firmly bound state, either a redox-active transition metal (e.g., iron and copper) or a nonmetal cofactor, such as FAD or FMN, and are often distinguished in this respect. Oxygenases are sometimes clustered in enzyme complexes and can catalyze cascades of reactions, which can lead to an enormous structural diversity when acting on PKS products. Various oxygenases that are involved in post-PKS tailoring steps catalyzing a wide array of chemical reactions including hydroxylations, epoxidations, anthrone oxidations, and oxidative Favorskii and Baeyer–Villiger rearrangements are known (for examples, see below).[106] Most common among type II PKS tailoring enzymes are P-450 and flavin-dependent oxygenases and cofactor-free anthrone oxygenases. The crystal structures of some oxygenases involved in tailoring steps of aromatic polyketides have been elucidated in the past several years and will provide further insight into this class of enzymes.[130–134]

1.07.2.2.3 Halogenases

Enzymes catalyzing the regio- and/or stereospecific introduction of halogen atoms become more and more important because the attached halogen atoms influence the biological activity of many bioactive natural products. The best known example of a halogenated natural product among the type II PKS-derived natural products is the antibiotic chlortetracycline. The 7-chlortetracycline halogenase Chl from *Streptomyces aureofaciens* was characterized by mutant analysis and incorporation experiments by McCormick *et al.* at the end of the 1950s and the early 1960s.[135–137] More recently, characterization[138] and cloning of the *chl* gene revealed the chlorination step to occur as the last step of the chlortetracycline biosynthesis.[139] Halogen atoms are most frequently found in amino acid–derived natural products, such as the antibiotics chloramphenicol and vancomycin or the anticancer drug rebeccamycin. Various halogen-introducing enzymes, including haloperoxidases and halogenases, were explored in the context of biosyntheses of secondary metabolites.[106,140–145] The most important class of halogenases are flavin-dependent halogenases, which are responsible for the chlorination of tetracycline and all other abovementioned natural product drugs. Recently, protein crystal structures of some of these flavin-dependent halogenases were clarified.[146,147] However, only a few examples exist where such halogenases were exploited for combinatorial biosynthesis – for example, on rebeccamycin/staurosporin analogues or on aminocoumarin antibiotic analogues.[148–151] No halogenase gene has yet been used to generate new derivatives of type II PKS-derived multicyclic aromatic compounds.

1.07.3 Natural Structural Diversity

Nature uses type II PKS systems to generate hundreds of natural products with highly diverse structures. Well-known groups of type II PKS-derived natural products are the BIQs, tetracyclines, anthracyclines, angucyclines, aureolic acids, and tetracenomycins. Although some of these groups contain only a few members (e.g., tetracenomycins and aureolic acids), other groups contain more than 100 members (e.g., anthracyclines and angucyclines). Tetracyclines (anti-infective antibiotics), anthracyclines (anticancer antibiotics), and the aureolic acids (anticancer antibiotics) are used clinically. At least some members of these groups are usually decorated with sugar moieties, which are essential for or add to the biological activities. This plethora of compounds reflects nature's impressive diversity, which we are only beginning to understand. This review is naturally restricted to examples whose biosynthesis is well studied and whose biosynthetic gene clusters were cloned and explored. Biosynthetically related groups are discussed together.

1.07.3.1 Natural Products Derived from Short Polyketides

1.07.3.1.1 Benzoisochromanequinones

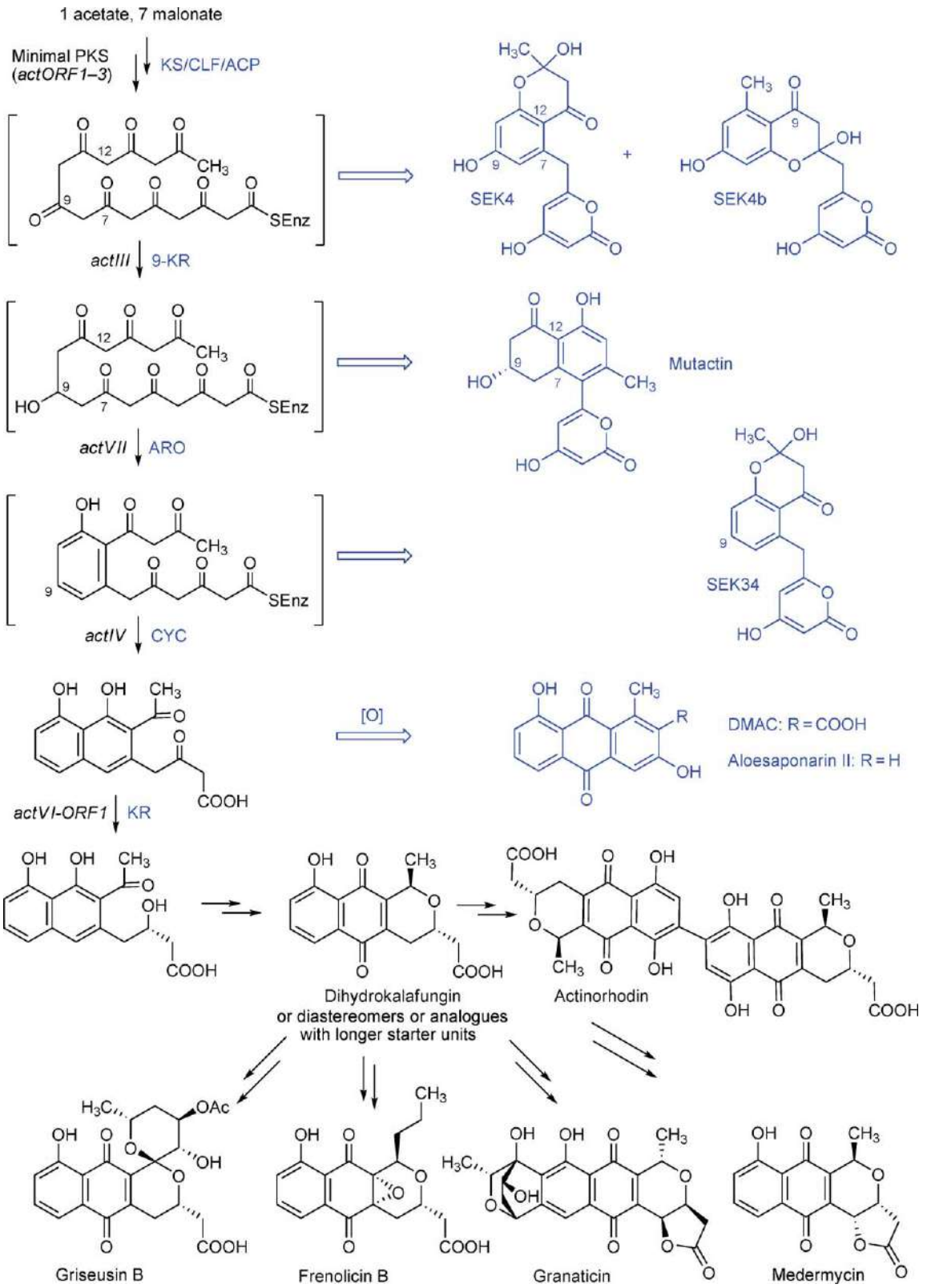
The BIQ, or isochromanequinone, class of antibiotics contains many well-studied members, for example, actinorhodin, griseusin A, frenolicin, medermycin, and granaticin. Besides antibacterial activities, cytotoxic activities were also found,[152] but none of these compounds has advanced to clinical applications. Actinorhodin, produced by *S. coelicolor*, is nevertheless the most famous and most studied aromatic polyketide and became the prototype and model for all biosyntheses of aromatic polyketides. The early steps are particularly exemplary for the aromatic polyketides. The basic biosynthetic studies were carried out mainly by Hopwood *et al.* and Khosla *et al.*, who gained an understanding of the early steps of the pathway by genetic step-by-step recombination of the *act* PKS genes.[5,11,153–167] Their work also paved the way for the generation of many artificial polyketides through combinatorial biosynthesis and the recent generation of new octaketides by rational mutation of the *act* PKS (see Section 1.07.4). Various other pathways to isochromanequinones were also studied, for example, the granaticin and medermycin pathways by Floss *et al.* and Ichinose *et al.*, respectively (**Scheme 9**).[2,7,32,155,168–172]

1.07.3.1.2 Other octaketide derivatives

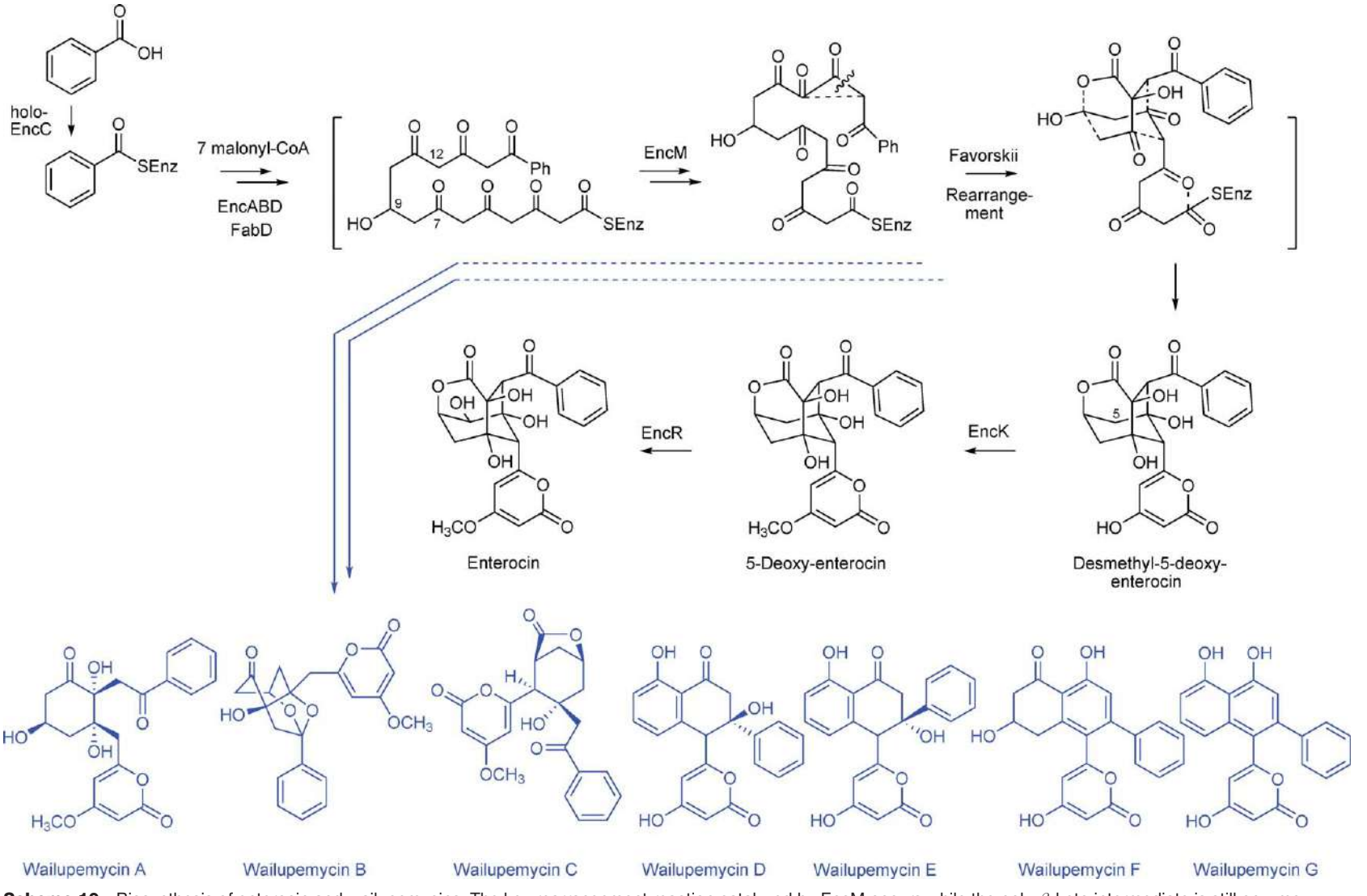
Enterocin, a bacteriostatic agent,[173,174] is unique among type II PKS-derived natural products because with its unique caged tricyclic nonaromatic core, it does not look like one due to its lack of directly fused aromatic rings that are otherwise typical structural elements of the multicyclic or polycyclic aromatic natural products, often used as synonym for type II PKS-derived compounds. In fact, the only aromatic ring of enterocin is phenylalanine derived. The biosynthesis of enterocin was thoroughly studied by the Moore group, which focused on various aspects of the pathway.[25,78,80,81,175–181] Most significant is the unusual starter unit, which can be gained either from phenylalanine or through direct activation from benzoate and the early rearrangement through oxygenase (favorskiiase) EncM. EncM catalyzes the oxidative Favorskii rearrangement, two aldol condensation, and two lactone-forming reactions in a mechanism that somewhat resembles DXP (1-deoxy-D-xylulose 5-phosphate) reductioisomerase.[182] The research culminated in the total enzymatic synthesis of the natural product (**Scheme 10**).[175] The wailupemycin shunt products, obtained through (part) inactivation of EncM or heterologous construction omitting EncM, possess fundamentally different interesting structures. The knowledge gained from the studies was exploited to design various analogues (see Section 1.07.4).

1.07.3.2 Natural Products Derived from Decaketides that Undergo Initial 7,12-Cyclization

The most typical and best investigated compounds derived from type II PKSs are tetracyclic decaketides, particularly those that undergo initial 7,12-cyclization to establish their tetracyclic frameworks.



Scheme 9 Biosynthesis of actinorhodin (adapted from[2]) and other isochromanequinones (black). Shunt products formed through spontaneous cyclizations of early reactive biosynthetic intermediates (blue) were generated by engineering cassettes containing the indicated genes, which encode the enzyme activity shown in blue (KS = ketosynthase-β, CLF = chain length factor or KS-β, ACP = acyl-carrier protein, KR = ketoreductase, ARO = aromatase, CYC = cyclase).



Scheme 10 Biosynthesis of enterocin and wailupemycins. The key rearrangement reaction catalyzed by EncM occurs while the poly-β-keto intermediate is still enzyme tethered. Shunt products (wailupemycins) from spontaneous cyclization/dehydration reactions or incomplete rearrangements are shown in blue.

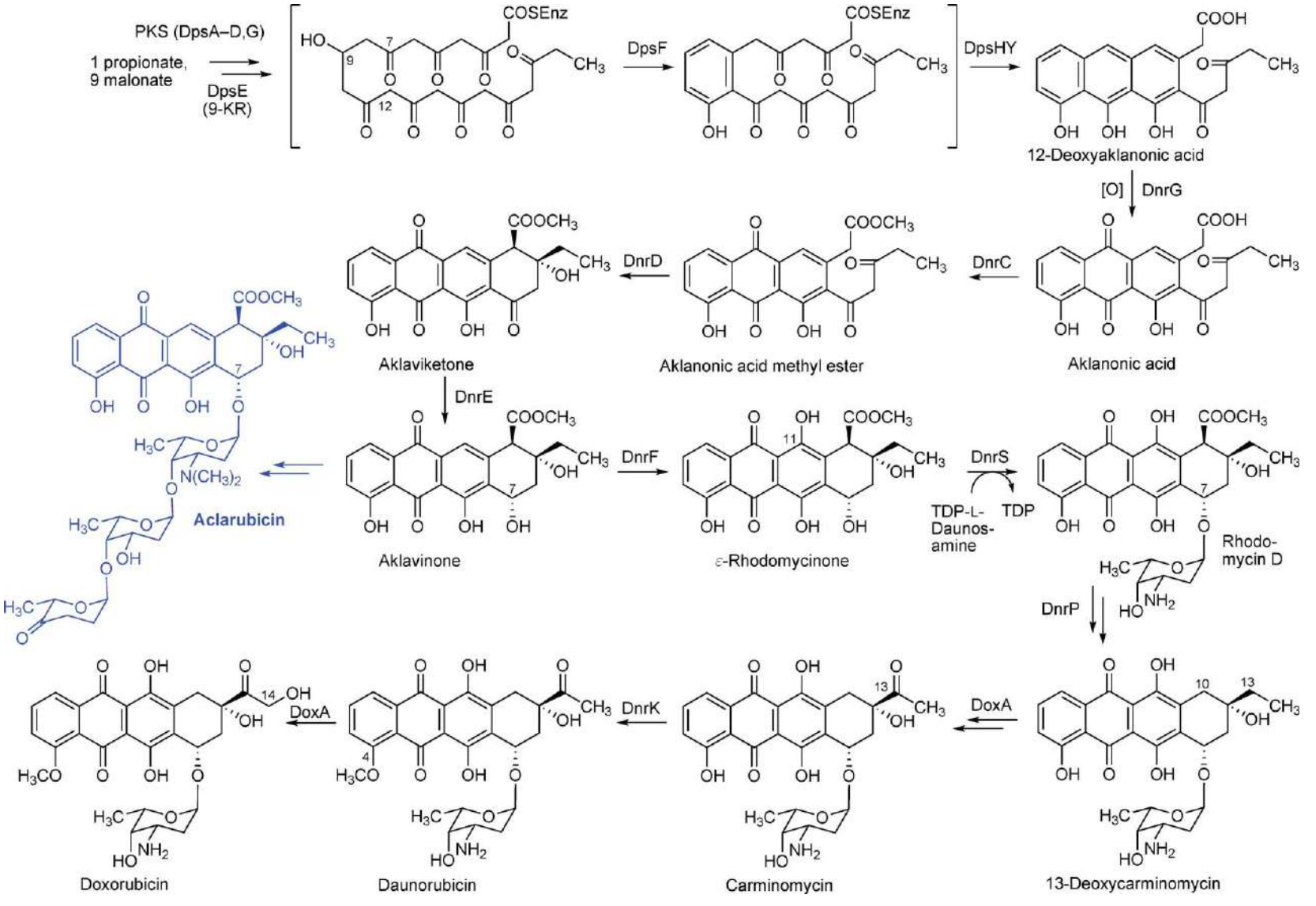
1.07.3.2.1 Anthracycline folding

1.07.3.2.1(i) Anthracyclines Anthracyclines, particularly daunorubicin and doxorubicin and their derivatives, are important clinically used anticancer drugs.[183,184] In addition, aclacinomycin A (aclarubicin), pirarubicin, and nogalamycin are also approved as clinical anticancer drugs. The activity of anthracyclines seems to be based on various mechanisms of action, including DNA intercalation, topoisomerase II inhibition, and radical generation – their biggest side effect being cumulative cardiomyopathy.[185–187] Several hundred natural, semisynthetic, and synthetic analogues exist, and the biosynthetic pathway, studied mainly by the research groups of K. Eckardt, C. R. Hutchinson, P. Mäntsälä, and W. R. Strohl, is one of the best explored and is exemplary not just for the anthracyclines but for many aromatic polyketides. Many of the key enzymes involved were studied, and several protein crystal structures are available.[2,13,66,70,184,188–199] For instance, the release from the PKS (KS/ACP) occurs at the stage of a tricyclic intermediate, as also found later for the tetracenomycins and the aureolic acids. Furthermore, the (first) glycosylation occurs after finishing the tetracyclic ring scaffold, as also found later for the urdamycins and aureolic acids. Finally, a complex oxygenase reaction, here triple hydroxylation by DoxA, finishes the biosynthesis (**Scheme 11**). Similarly, TcmG finishes the tetracenomycin C biosynthesis, and MtmOIV catalyzes the second last step of the mithramycin biosynthesis. The knowledge gained from the biosynthetic studies could be exploited to engineer various aklanonic acid derivatives and high producers of the commercially important doxorubicin as well as epirubicin, which was formerly available only by a more complicated chemical derivatization process.[73,190,200–208] Various pathways to other anthracyclines were also studied in detail, particularly the nogalamycin pathway and the aclacinomycin A (aclarubicin) pathway, which deviate from the doxorubicin pathway at different stages. For instance, the aclacinomycins (aclarubicins) derive from aklavinone, although nogalamycins and steffimycins/aranciamycins use acetate starters and fourth-ring cyclases (DnrD analogues) that provide an opposite stereochemistry at C-9, the steffimycin pathway lacks a DpsE analogue, that is, the 9-ketoreduction step within the PKS, and so on.[27,46,95,96,130–132,209–222]

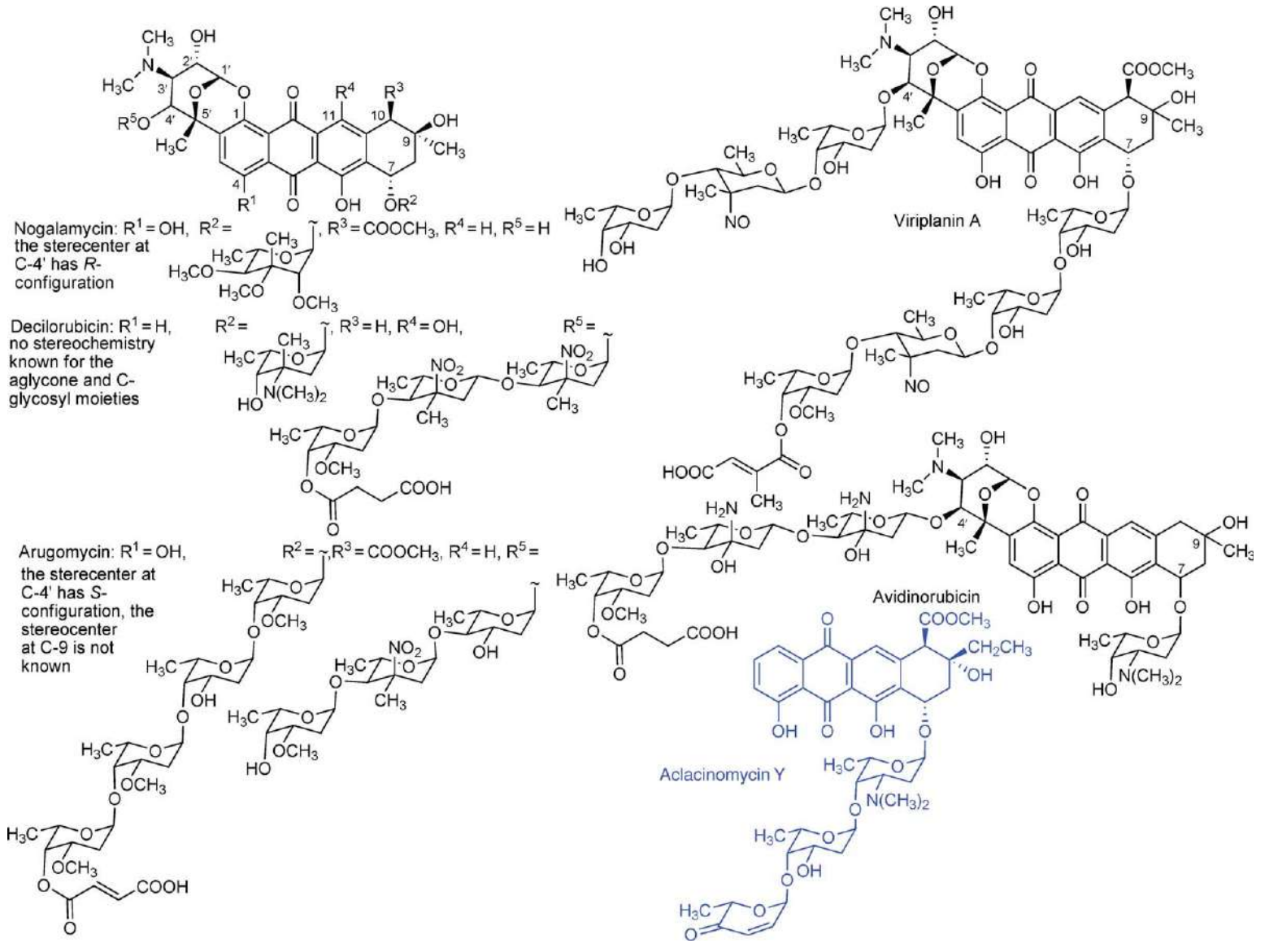
Some anthracyclines possess far more complex deoxysugar chains, which contain unique sugars, for example, the ketosugars L-cinerulose and L-aculose in the aclacinomycins (aclarubicins) or the unusually partly C-linked aminosugar as well as unique nitro- or nitrososugars found in nogalamycin and its more complex derivatives decilorubicin, arugomycin, avidinorubicin, and viriplanin (**Scheme 12**). The latter group also shows various uncommonly β-glycosidically linked L-sugars (anti-Klyne's rule). However, in the case of viriplanin, a careful reanalysis of the structure showed that the nitrososugars, originally thought to be β-L-configurated, were indeed D-configurated.

1.07.3.2.1(ii) Chartreusins Chartreusin (**Scheme 13**) is an aromatic polyketide glycoside isolated from *S. chartreusis* consisting of fucose, digitalose, and an unusual bislactone aglycone (chartarin).[223,224] Although chartreusin was first investigated because of its antibacterial activity,[223] further studies revealed a significant chemotherapeutic activity against various tumor cell lines, such as murine P388 and L1210 leukemia and B16 melanoma cells.[225] Unfortunately, chartreusin development has been hampered by its unfavorable pharmacokinetics due to rapid biliary excretion and slow gastrointestinal absorption.[225] The natural derivative elsamicin A (or elsamitrucin), produced by an unidentified actinomycete strain, has an improved water solubility due to amino sugar moiety.[226] Both elsamicin A and a semisynthetic derivative of chartreusin, IST-622, which appears to be a prodrug with more suitable pharmacokinetic properties, have reached phase II clinical trials.[227] Pharmacological studies revealed that chartreusin and its derivatives exert their antitumor activities through binding to DNA,[228] radical-mediated single-strand breaks, and inhibition of topoisomerase II.[229] In all cases, the unusual polycyclic aromatic aglycone – chartarin – represents a key element for bioactivity.[230,231] Early ^{13}C-labeling experiments indicated that the coumarin-related aglycone of chartreusin and elsamicin is derived entirely from acetate and is thus supposedly of polyketide origin.[232–234] However, various acetate units are disrupted, and it was suggested that either two individual polyketide chains are fused during the biosynthesis[235] or that an undecaketide-derived benzopyrene intermediate would undergo a series of oxidative C–C bond fissions and rearrangements.[232,233]

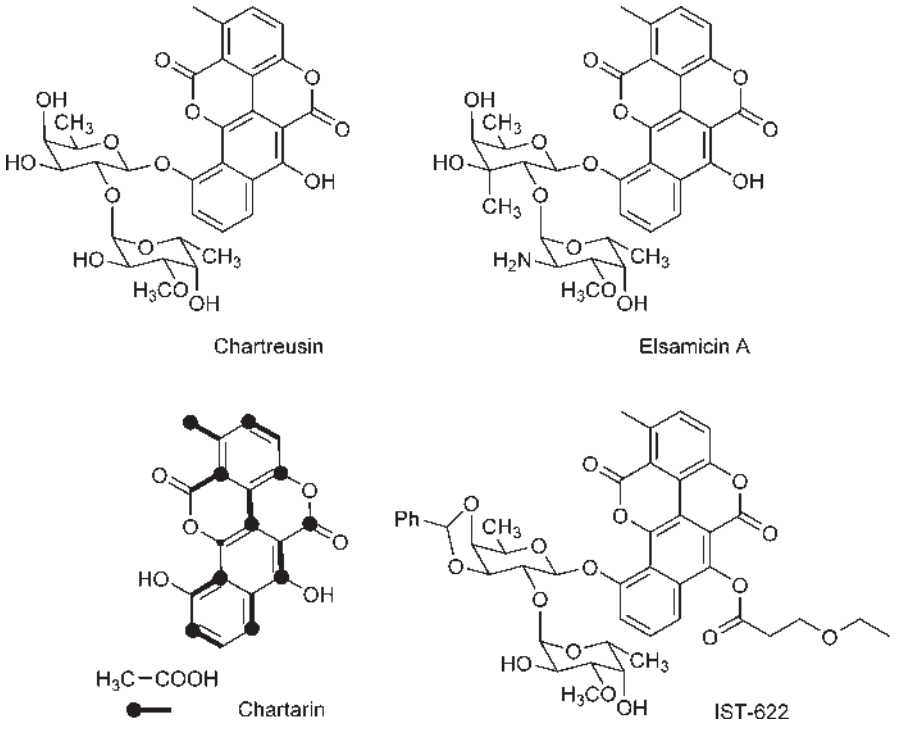
The analysis of the molecular basis of chartreusin biosynthesis shed light on its unusual pathway. The chartreusin biosynthesis gene cluster of *S. chartreusis* HKI-249 was cloned and sequenced, and its identity was confirmed by heterologous expression.[42] Surprisingly, the polyketide biosynthesis genes share highest



Scheme 11 Biosynthesis of daunorubicin and doxorubicin. This pathway is exemplary not only for many anthracyclines but also for aromatic polyketides in general (see text). Branches to other anthracyclines exist at different stages of the pathway, for example, at aklavinone for the aclarubicins (blue).



Scheme 12 Structures of complex anthracyclines of the nogalamycin (black) and aclacinomycin (blue) type.



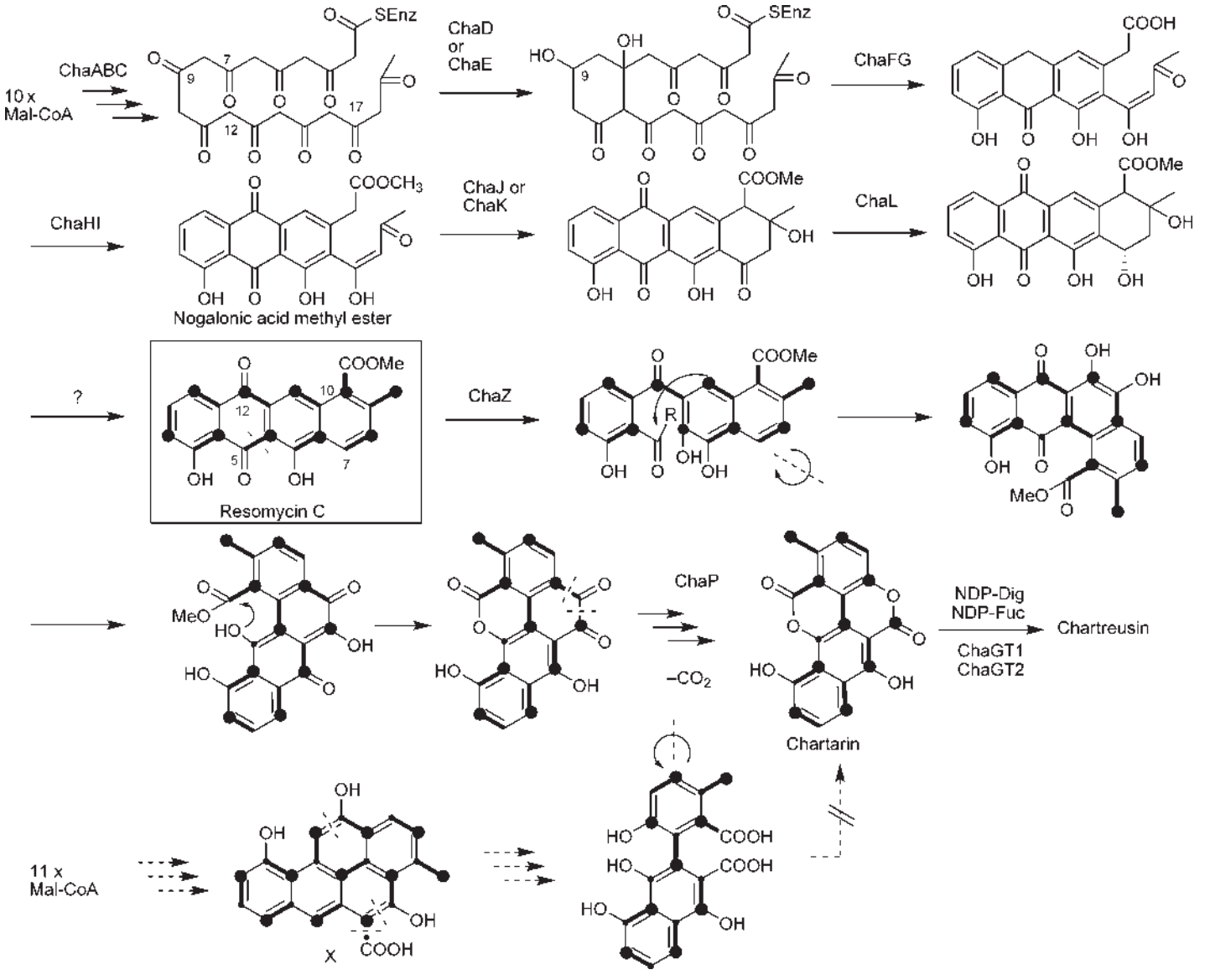
Scheme 13 Structures of the potent antitumor agents chartreusin, elsamicin, and IST-622, and acetate labeling pattern of the aglycone chartarin.

similarity with the previously studied anthracycline biosynthesis genes. The hypothesis that chartreusin is derived from an anthracyclic precursor was supported by a knockout of the oxygenase gene *chaZ*, which led to the accumulation of resomycin,[236] an anthracyclic precursor or shunt product of the chartreusin pathway.[42] On the basis of this information and on earlier isotope labeling results, a revised biosynthetic model involving an unprecedented rearrangement of an anthracyclic ring system was proposed (**Scheme 14**).

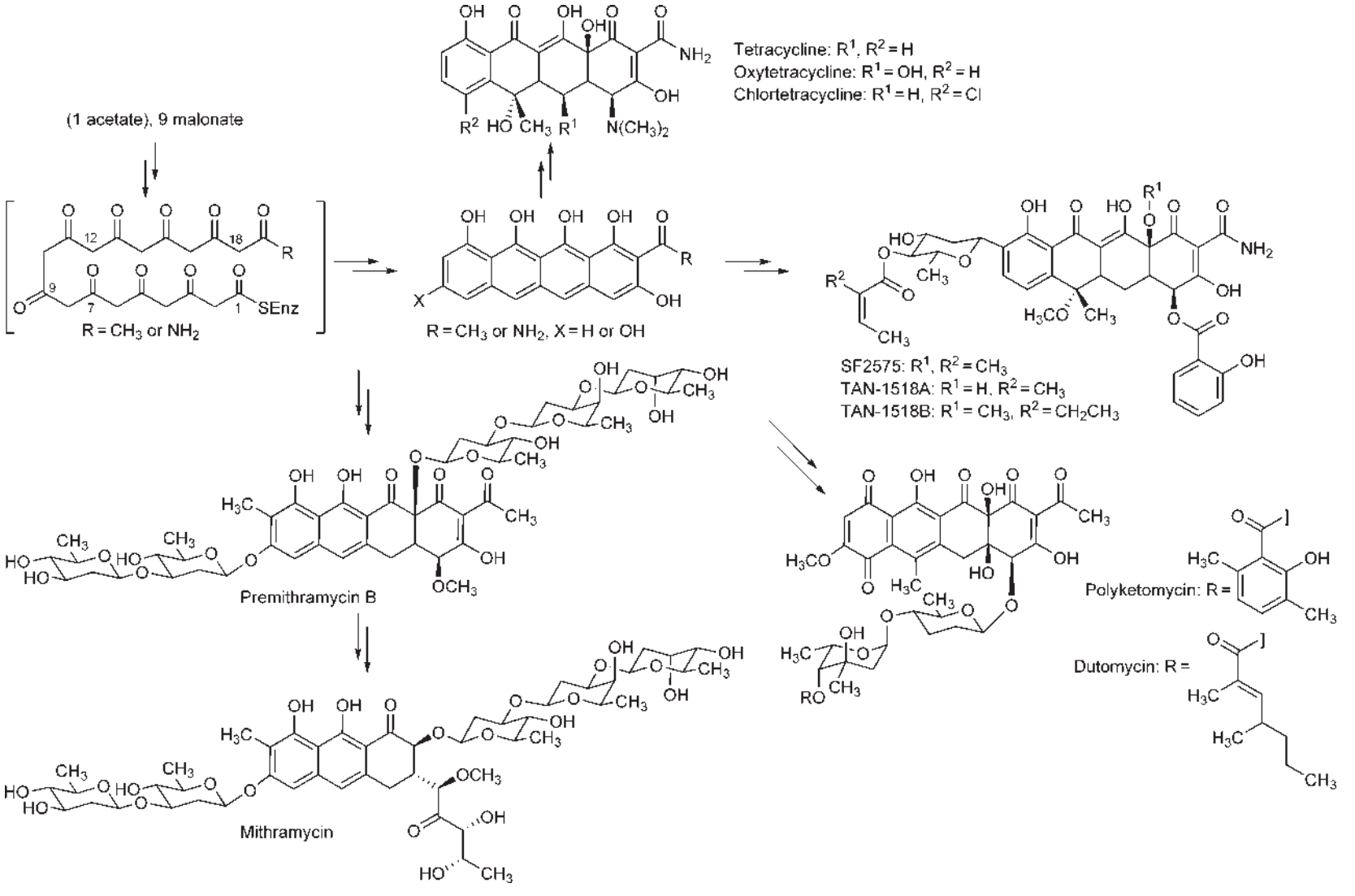
1.07.3.2.2 Tetracycline folding

Tetracyclines are one of the most important classes of broad-spectrum antibiotics.[237] Their framework is type II PKS derived, and various natural (tetracycline, chlortetracycline, oxytetracycline, and demeclocycline) and semisynthetic tetracyclines (e.g., doxycycline, minocycline, rolitetracycline, and tigecycline) are in clinical use. Their antibacterial mode of action is interruption of protein biosynthesis by hindering the docking of aminoacyl-tRNA to the mRNA–ribosome complex, which is achieved mainly by binding to the 30S ribosomal subunit of the mRNA translation complex. In addition, antibacterial activity is enhanced, as bacteria erroneously take up tetracyclines by an active transport system.[238] Furthermore, tetracyclines were also found effective as drugs to treat prion diseases[239–241] and as neuroprotective drugs.[242] In addition to antibacterial drugs, various tetracycline-related glycosidic compounds are known (e.g., SF2575,[243,244] TAN-1518B,[245,246] cervimycin,[39,247] dutomycin,[248] polyketomycin[249–251]), most of which possess antitumor activity. Finally, biosynthetic studies of the aureolic acids revealed that this group of anticancer drugs is also related to the tetracyclines. All of these natural products have in common that the initial decaketide (or nonaketide) fold is identical, with the fourth ring being closed between C-1 and C-18 (**Scheme 15**).

The tetracycline pathway fascinated researchers early on. The chlortetracycline pathway was investigated by McCormick *et al.* and Dairi and Nakano *et al.*[136,137,139,252–258] The oxytetracycline pathway,[9,101,102,259–263] studied by Hunter and coworkers and more recently by Tang *et al.*, is characterized by its unusual malonamide starter unit, whose N-atom is introduced involving amidotransferase homologue OxyD. For all cyclizations, only two enzymes (OxyK and OxyN) are needed leading to the first isolable intermediate, pretetramide.



Scheme 14 Structure of the linear polyketide metabolite resomycin C, isolated from the chaZ null mutant, and revised biosynthetic model for the formation of the chartreusin aglycone chartarin (vs. previously proposed pathway via benzo(a)pyrene X).



Scheme 15 Some examples of natural diversity of naphthacene-core-derived tetracycline-related natural products.

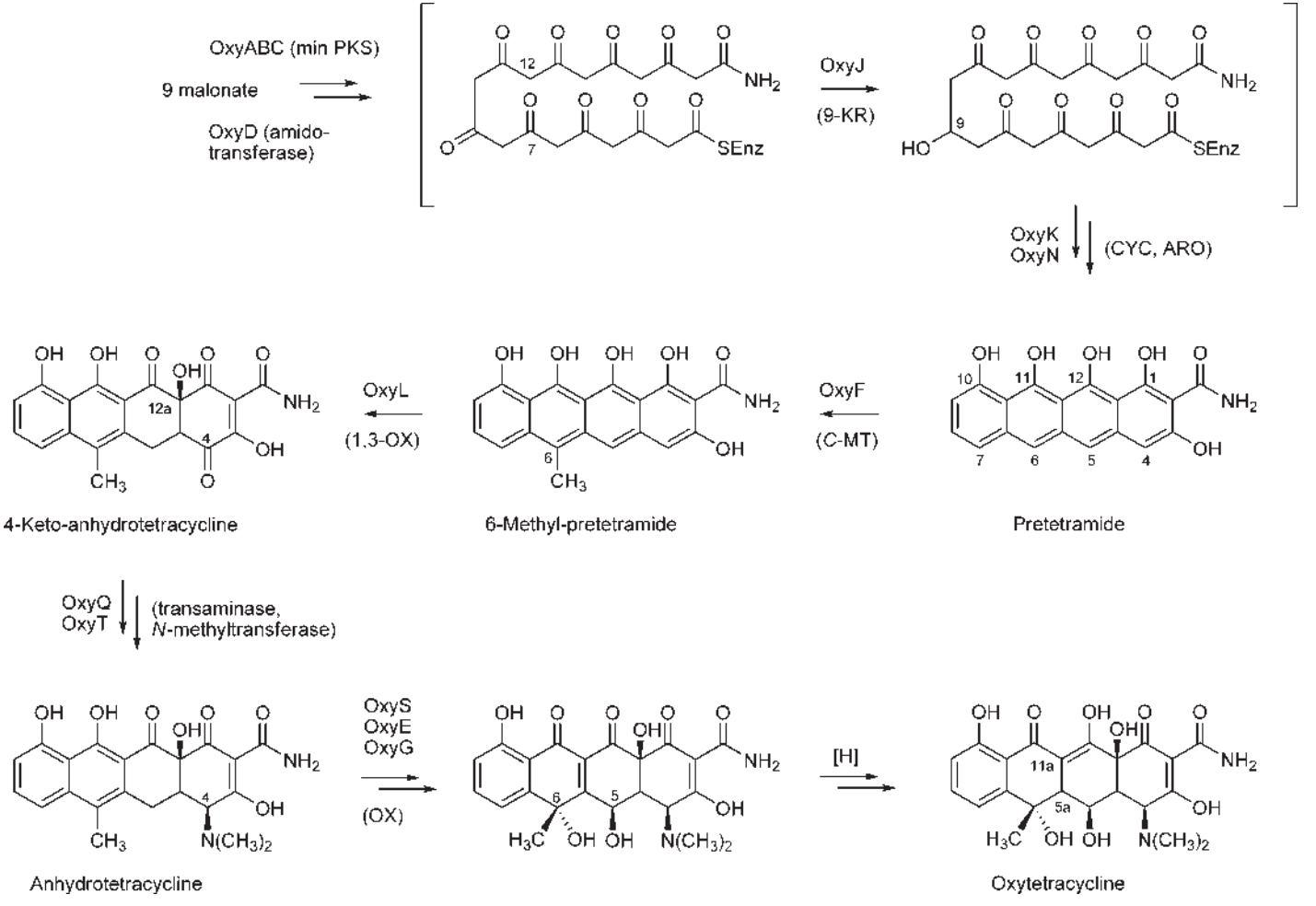
Evidence for an enzyme-tethered tricyclic intermediate was found from the analysis of minor shunt products of heterologous reconstitution constructs. Pretetramide is then C-methylated (OxyF) and further processed by 1,3-oxygenase OxyL (introducing the oxygens in the 4- and 12a-positions), generating a very labile intermediate, which is stabilized through reductive transamination by OxyQ and N,N-dimethylation by OxyT. The biosynthesis is finished by further oxidation and reduction processes (**Scheme 16**). An investigation of the oxytetracycline PKS solution structure using multidimensional heteronuclear nuclear magnetic resonance (NMR) was carried out by Crump and coworkers.[264]

The aureolic acid group of anticancer drugs consists of mithramycin, olivomycin, the chromomycins, chromocyclomycin, durhamycin, and UCH9.[265–270] The anticancer activity is achieved by crosslinking GT-rich DNA as Mg^{2+}-coordinated dimers, thereby blocking Sp (specificity protein)-dependent signal transduction pathways. Mithramycin was clinically used to treat testicular cancers and cancer-related bone diseases, including Paget's disease. Its hypocalcemic osteoclast-inhibiting effect is likely due to its ability to inhibit c-src transcription.[271–279] Durhamycin also shows anti-HIV activity; Mithramycin and chromomycin A_3 were discussed as neuroprotective drugs with significant activity against Huntington's disease.[265,280,281] All these natural aureolic acid drugs contain the same polyketide-derived aglycone, except for the 7-side chain (isobutyl instead of methyl in durhamycin and UCH9, and no 7-side chain in olivomycin). Variations occur mainly in the composition and length of the saccharide chains, which can range from one sugar (upper saccharide chain in UCH9) to four sugars (lower saccharide chains in durhamycin and UCH9). Most compounds have the disaccharide–trisaccharide arrangement found in mithramycin or chromomycin (**Scheme 17**). The biosynthetic pathways leading to chromomycin and mithramycin were initially studied by Rosazza,[282,283] and later by Mendez, Rohr, Salas *et al.*[35,103,111,270,278,284–300] (**Scheme 17**). What is remarkable about the mithramycin pathway is (1) evidence for a tricyclic intermediate (most likely still enzyme bound; see Section 1.07.4.2.3), (2) complete glycosylation sequence on tetracyclic polyketide intermediates (premithramycins) with one twice acting GT (presumably MtmGIV), and (3) the final oxidative Baeyer–Villiger rearrangement reaction, which generates the unique tricyclic scaffold of the aureolic acids. The early acting oxygenase MtmOII, the equivalent of 1,3-oxygenase OxyL of the oxytetracycline pathway (see above), is responsible for the insertion of the oxygens in the 12a-position and possibly also in the 4-position. However, MtmOI seems to be able to substitute MtmOII for the latter activity. In contrast to OxyL, MtmOII also appears to partake in the fourth cyclization step (see premithramycinone G in Section 1.07.4.2.3).

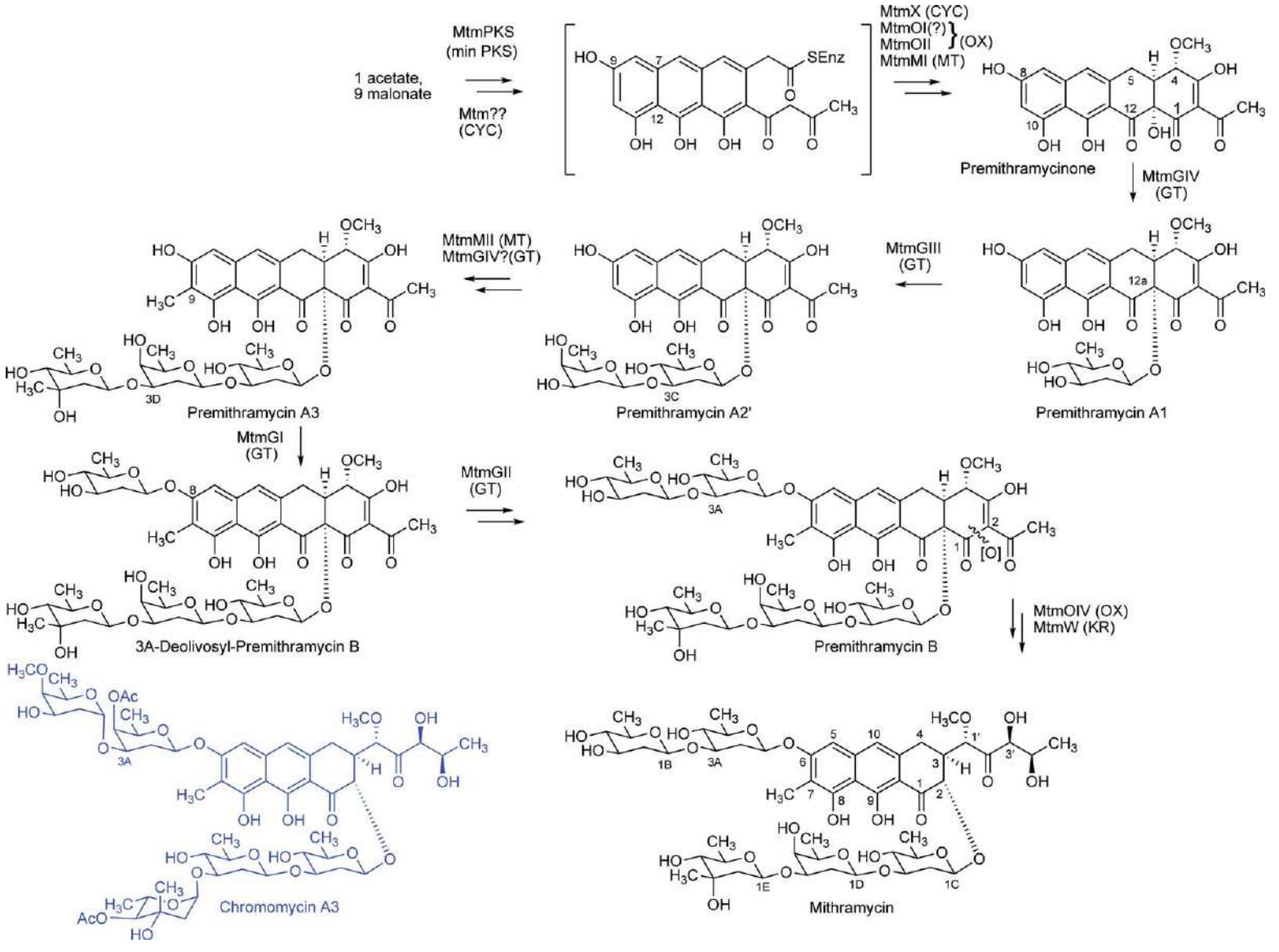
The very last step of mithramycin biosynthesis, the side chain ketoreduction by MtmW, stabilizes the labile 1,3-β-diketo side chain (see mithramycins DKA and DK, **Scheme 18**) created by the MtmOIV reaction. In the absence of MtmW, various mithramycin derivatives with shorter side chain were formed, likely through nonenzymatic β-shift rearrangements. Two of the derivatives, mithramycins SK and SDK, are interestingly significantly less toxic and therefore possess a better therapeutic index than the parent compound mithramycin itself. It is unclear whether MtmOIV partakes in their formation (**Scheme 18**). Recent emphasis focused on sugar modifications (see Section 1.07.4). The chromomycin pathway follows pretty much the same path, except for some additional sugar modifying group transfer reactions, which occur at the very end of the biosynthesis after the sugar attachment.

1.07.3.2.3 Angucycline folding

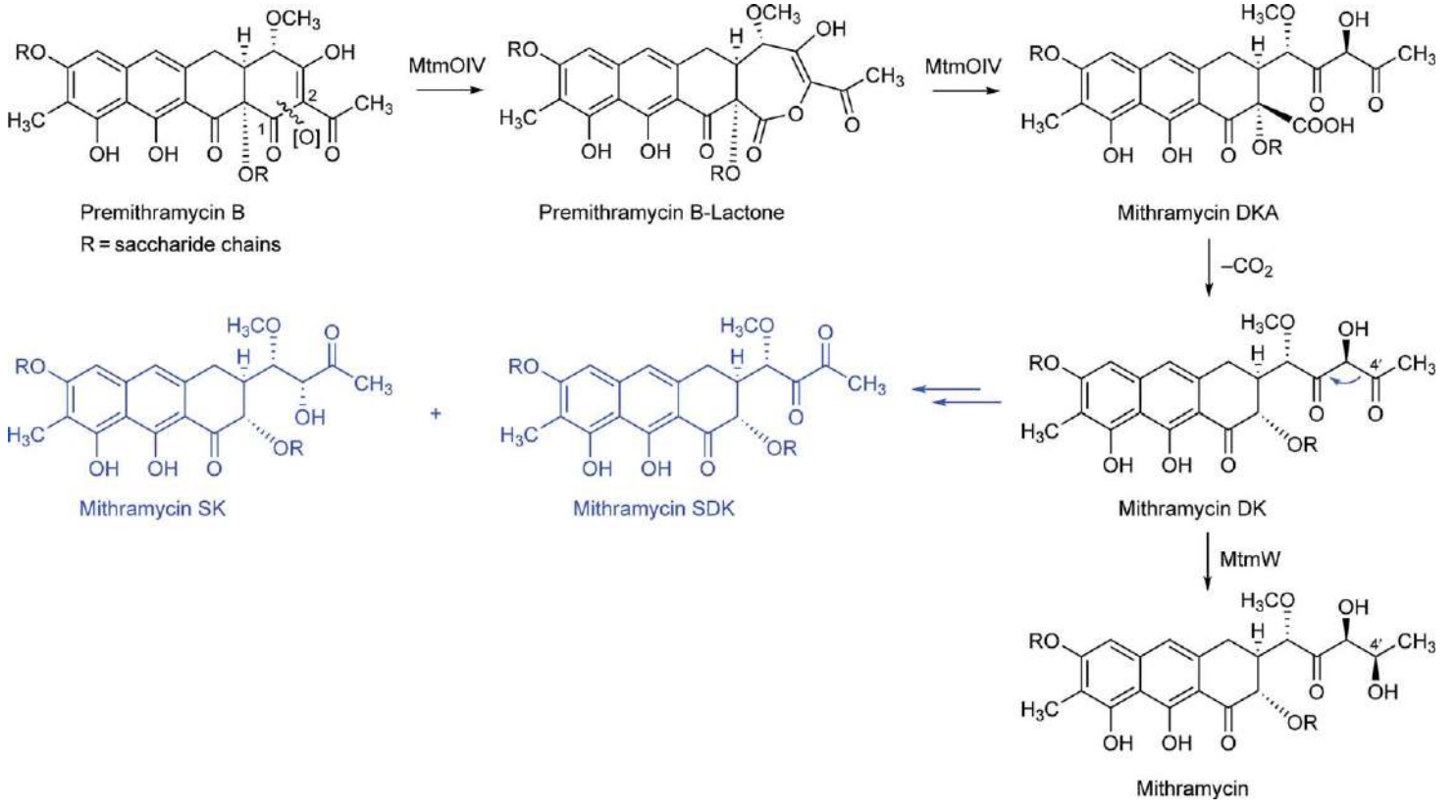
Angucyclines form by far the largest group among type II PKS–decaketide-derived natural products.[301,302] The biological activities of angucyclines or angucyclinones are quite diverse. Major biological activities were enzyme inhibitory, antibacterial, antiviral, anticancer, and platelet aggregation inhibitory effects. The biosyntheses of various typical angucyclines/angucyclinones (urdamycins, landomycins, gaudimycins, simocyclinone, and oviedomycin), whose angular benz[*a*]anthracene-derived polyketide backbones remained intact, were studied. Among these, the landomycins belong to the most active anticancer drugs. In addition, considerable focus was given to unique angucyclinone-derived natural products whose polyketide-derived scaffold is drastically modified through oxidative rearrangement reactions (jadomycins, gilvocarcins, and kinamycins).



Scheme 16 Biosynthesis of oxytetracycline in *Streptomyces rimosus*. The functions of many of the enzymes were verified through heterologous reconstitution, enzyme studies, and mutant product analysis (CYC = cyclase, ARO = aromatase, OX = oxygenase, [H] = reduction step/s).



Scheme 17 Biosynthesis of mithramycin in *Streptomyces argillaceus*. The functions of the enzymes were verified through gene inactivation experiments and enzyme studies. The sugar-decorating O-acyl and O-methyltransfer reactions in the chromomycin (blue) biosynthesis occur as last steps.



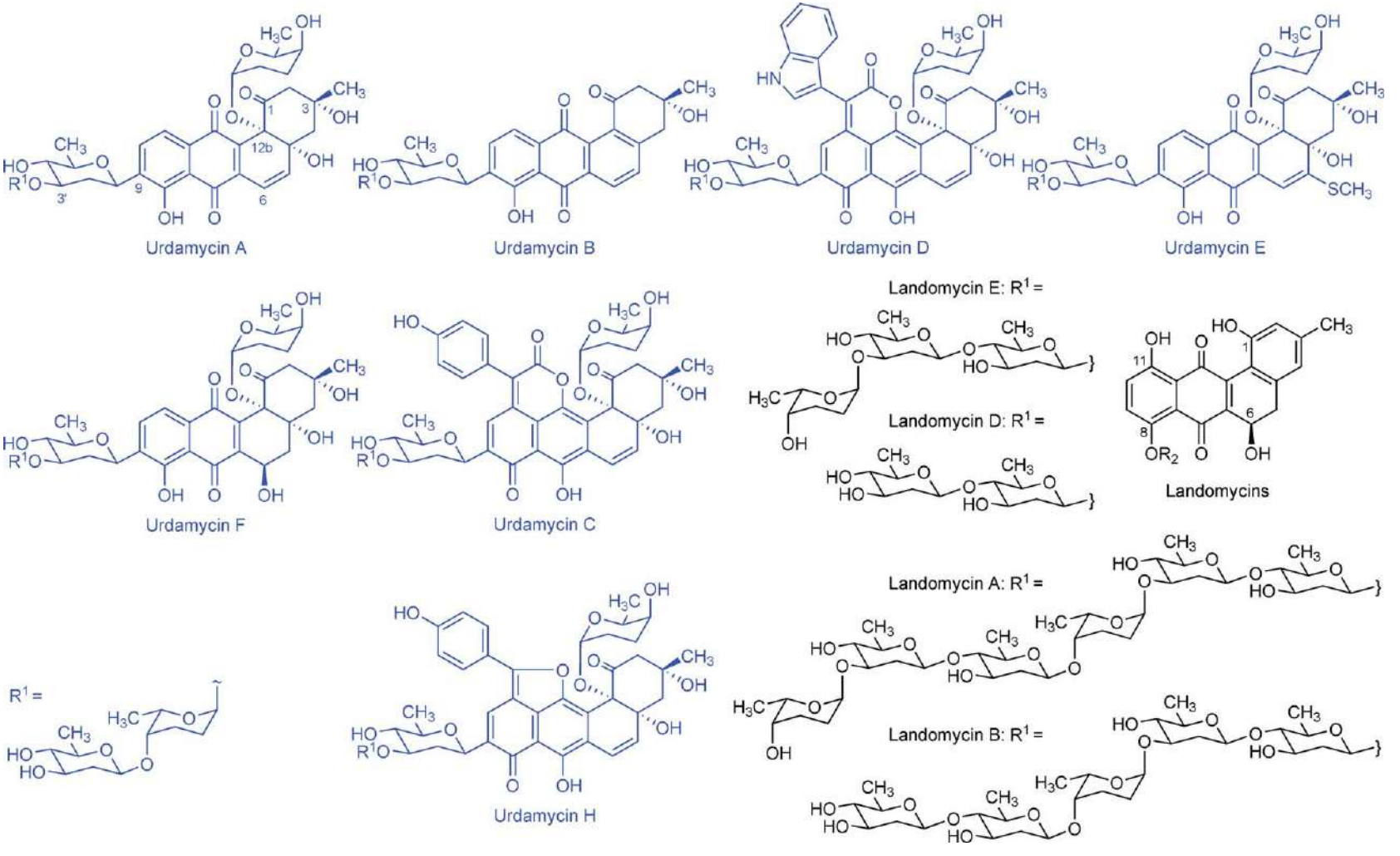
Scheme 18 Reactions of Baeyer–Villiger oxygenase MtmOIV and ketoreductase MtmW. In the absence of MtmW, through β -shift and subsequent C1-loss, rearranged compounds mithramycins SK and SDK are observed (blue).

1.07.3.2.3(i) Typical angucyclinones/angucyclines (urdamycins, landomycins, and oviedomycin) Urdamycins and landomycins (**Scheme 19**) are anticancer antibiotics, the latter group being much more potent than the former.[303–309]

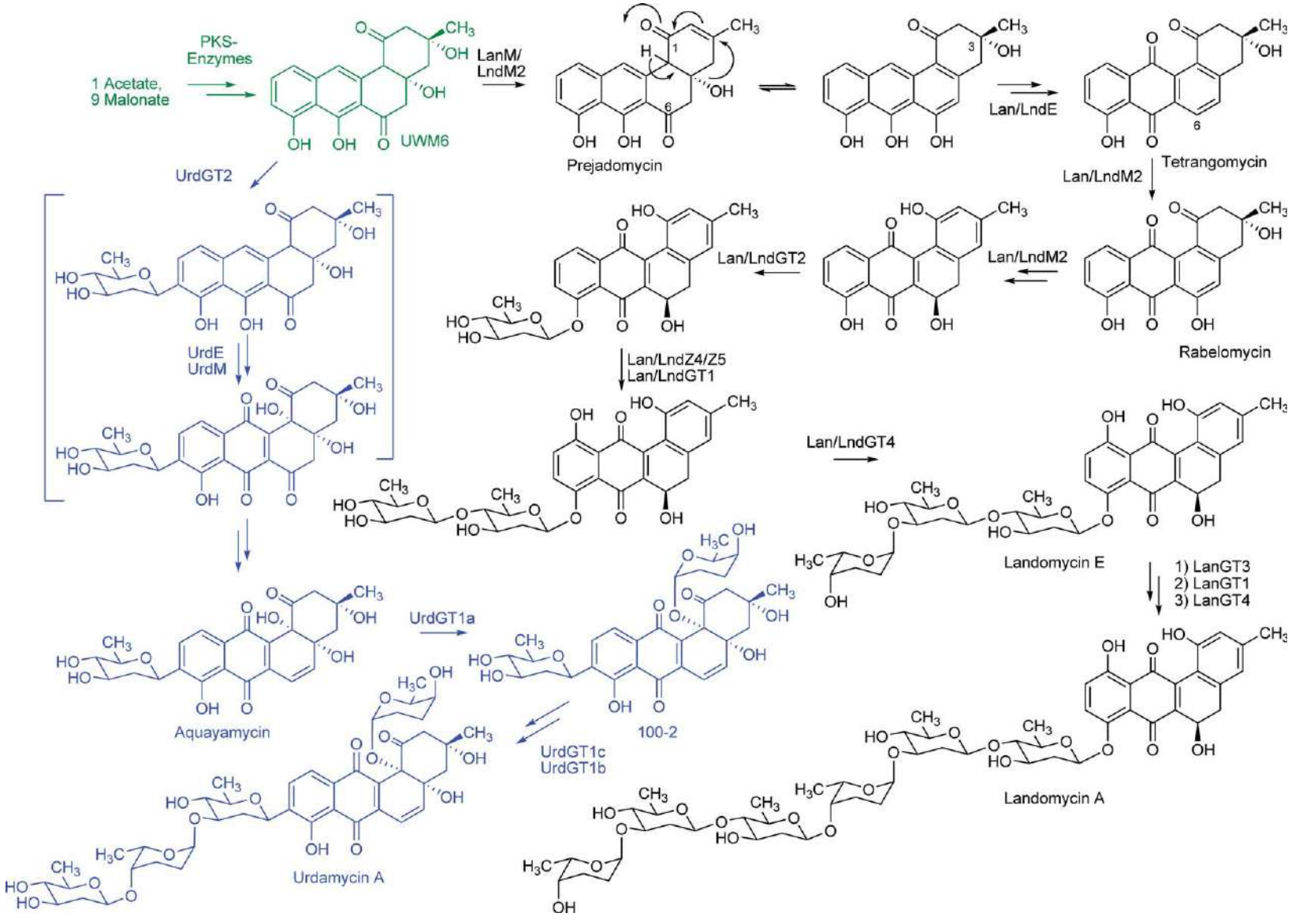
The mechanism of action of the landomycins is currently under investigation, and it appears that the drugs are not affected by multidrug resistance mechanisms.[310] Biosynthetic studies on urdamycins and landomycins range from initial incorporation experiments and mutant studies[309,311–313] to gene cluster analysis in the context of gene inactivation experiments and studies of selected enzymes.[2,15,23,124,309,311,313–332] Many experiments exploiting (predominantly deoxysugar pathway) genes of these pathways were used to generate hybrid antibiotics (see below).[333,334] The genetic studies concentrated on the assignment of oxygenase genes and on the understanding of how the complex saccharide moieties were assembled. All steps leading to the dark discolored urdamycins C, D, E, and H (**Scheme 19**), in which amino acid-derived precursors are used for the chromophore extensions, are nonenzymatic[335–337] and mostly weaken the biological activity. A center of interest of the urdamycin pathway was *C*-glycosyltransferase (GT) UrdGT2, the first systematically investigated *C*-GT,[124,314,315,317,323,333,338–341] for which a protein crystal structure was recently published.[124] The landomycin A hexasaccharide chain with a repeating D-Oliv–D-Oliv–L-Rho motif is to date the longest deoxysugar chain observed among angucyclines. Its biosynthesis requires only a relatively small set of deoxysugar pathway genes and only four GTs. It can therefore be considered as a masterpiece of nature's efficiency. Comparison of the two available gene clusters of landomycin A from *Streptomyces cyanogenus* (*lan*) and of its trisaccharidal 'little brother' landomycin E from *Streptomyces globisporus* (*lnd*) facilitated the gene assignments and an understanding of the buildup of the saccharidal chains. Important results were the recognition of the double function of olivosyltransferases Lan/LndGT1 and Lan/LndGT4 besides the unique roles of Lan/LndGT2 and LanGT3 in chain start and fourth glycosyl transfer, respectively, which distinguishes the two chains found in landomycins E and A. Another important aspect – the regulation affecting the saccharide assembly and self-resistance – is currently under investigation by the Fedorenko group.[326–328] The PKS-derived portions of the *urd* and *lan*/*lnd* pathways split quite early on, the generally found angucycline intermediate UWM6 being the last common intermediate. The urdamycin pathway is straightforward and without surprises, except for the unexpectedly early occurring *C*-GT step. In contrast, the biosynthesis of the landomycin aglycone contains unusual steps, including an oxygen rearrangement from 4a- to 3-position and seemingly fussy redox chemistry, considering that the oxygen in 6-position is first removed, but later readded, by the oxygenase function of oxidoreductase LndM2.[314,331] Various aspects of these pathways were already extensively reviewed, particularly with respect to the saccharide chain biosyntheses[1,105,113,330,342–344]; only some key steps are shown here (**Scheme 20**).

Oviedomycin is an angucyclinone with an unusually oxygen-rich ring A.[345,346] From heterologous expression experiments in *S. albus*, in which the *ovd* PKS genes were combined with different sets of *ovd* oxygenase genes, eight pathway intermediates and shunt products were found, several of which are similar or identical to those known from previous studies on jadomycins, gilvocarcins (see below), and landomycins (see above). The data allowed researchers to propose a hypothesis regarding the sequence of tailoring events involving three oxygenases,[347] one of them being a bifunctional oxygenase/dehydratase, as was found before for the jadomycin biosynthesis (**Scheme 21**). Interestingly, three of the intermediates show better *in vitro* antitumor activity than the parent drug oviedomycin, which is a weaker anticancer drug due to natural overoxidation.

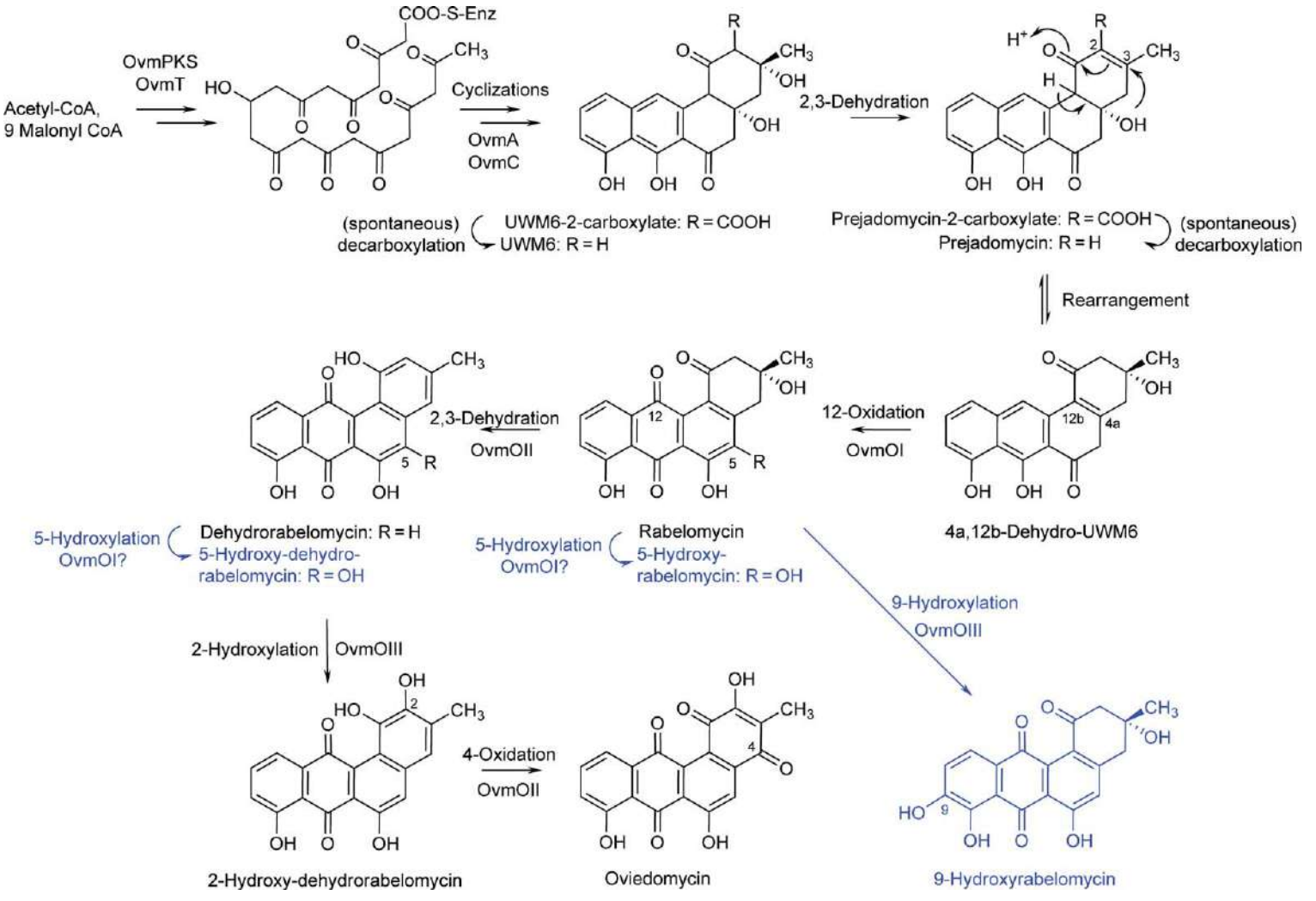
Sequencing of two cryptic angucycline-type aromatic polyketide gene clusters from *Streptomyces* sp. PGA64 and H021 by Mentsälä, Metsä-Ketelä *et al.* revealed several genes encoding oxidoreductases involved in processing the basic angucyclic carbon skeleton (*pgaE*, *pgaM*, *cabE*, *cabM*, *and cabV*); most of these were similar to those previously found for the urdamycin and landomycin pathways. The biosynthetic pathways were reconstructed in *S. lividans* TK24 by cloning and sequentially expressing the angucycline tailoring genes (*pgaE*, encoding an FAD-dependent monooxygenase, and *pgaM*, encoding a bifunctional oxygenase/reductase) with PKS genes required for the biosynthesis of the common angucycline intermediate UWM6. The expression studies showed that, after the production of UWM6, the pathways proceed through 12-oxygenation, catalyzed by the monooxygenase PgaE or CabE, possibly followed by spontaneous quinone formation. After this, reactions catalyzed by PgaM (→12b-hydroxylation, 2,3-dehydration, and 5,6-reductions) and CabMV (→6-CO-reduction and 12b-hydroxylation) finish the formation toward the new angucyclinones gaudimycins A–C (**Scheme 22**). The experiments demonstrate that genes that are either inactive or cryptic in their native host



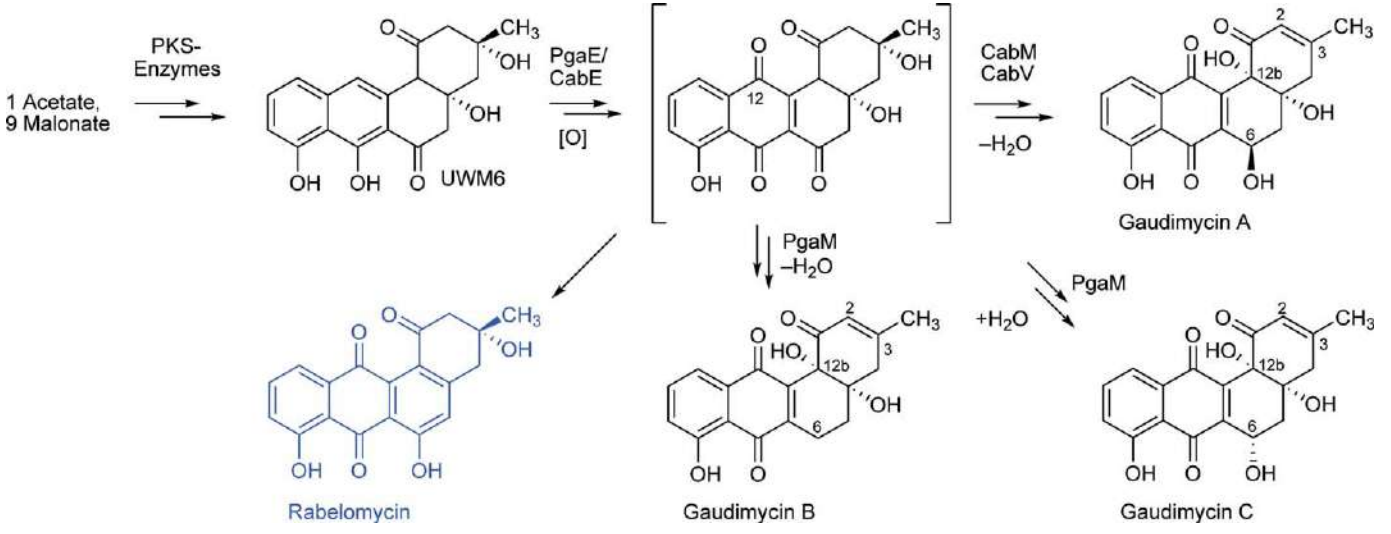
Scheme 19 Examples of natural urdamycins (blue) and landomycins (black).



Scheme 20 Biosynthesis of urdamycin A (blue) and landomycins E and A (black), both branching off from a short common pathway (green). The unusually complicated redox biochemistry in the lan/lnd pathway catalyzed by various, sometimes bifunctional, enzymes appears to be necessary to prepare a selective first glycosylation (by LanGT2), which otherwise would be hard to achieve with several very similar phenolic OH-acceptor groups found in landomycinone and its precursors. For early oxidation steps in the urdamycin pathways, see also **Scheme 23**.



Scheme 21 Proposed biosynthesis of oviedomycin. The main pathway is depicted in black and shunt products in blue. Some of the intermediates as well as the shown rearrangement are the same as previously found for the landomycin and gilvocarcin/jadomycin pathways; however, the 2-carboxy intermediates were not found anywhere else. Unspecific oxidations (→ shunt products) occurred once the oxygenase complex was disrupted.



Scheme 22 Proposed biosynthesis of the gaudimycins involving cryptic oxidoreductases. The gaudimycin pathway is depicted in black and shunt product rabelomycin in blue. Note that reductase CabV, a separate enzyme for the 6-reduction step, generates an opposite stereochemistry at C-6 than the multifunctional PgaM.

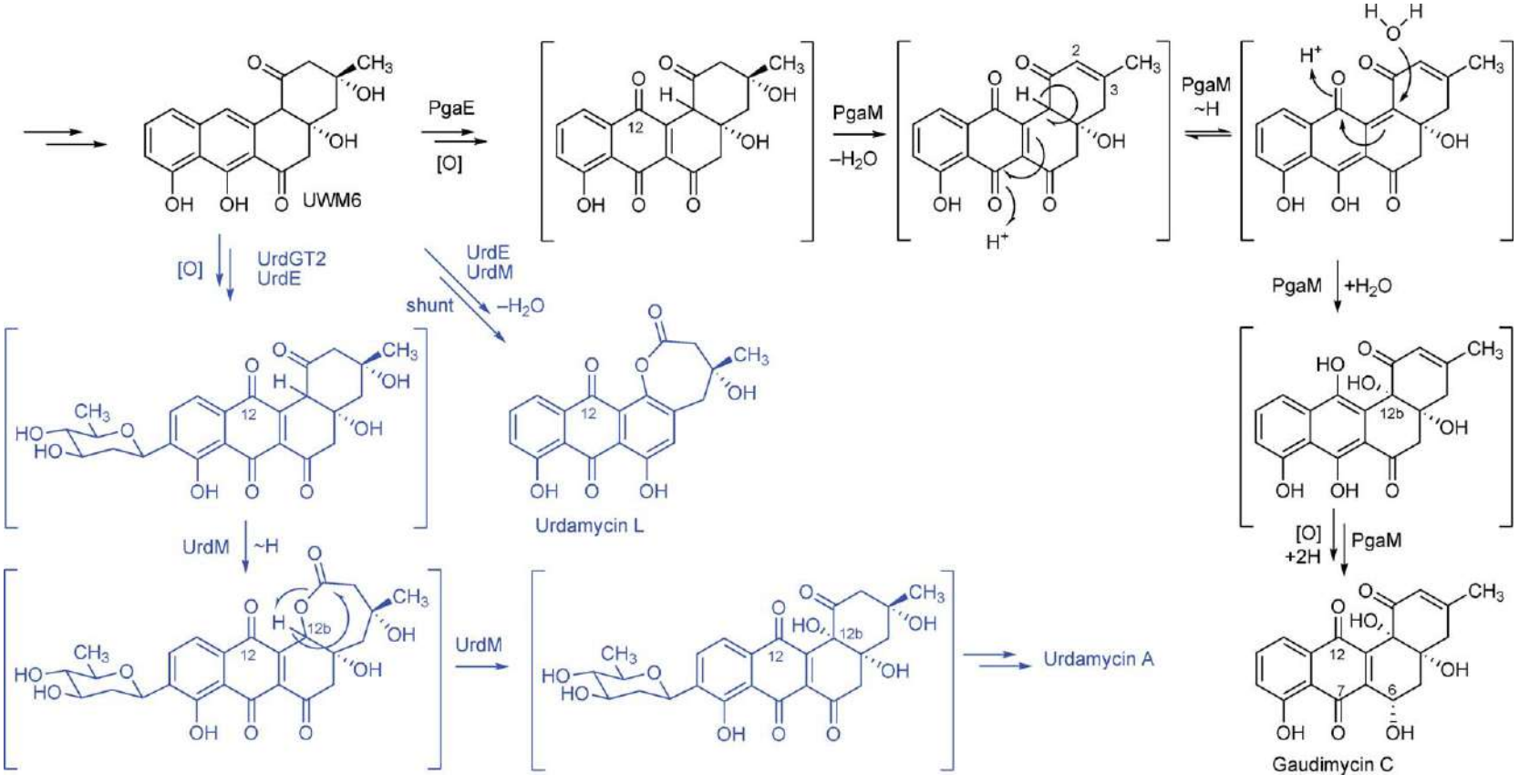
can be used as biosynthetic tools to generate new compounds.[41] In addition, enzymes PgaE and PgaM en route to gaudimycin C were further investigated, showing the requirement of several closely coupled reactions to avoid intermediate degradation. This clarifies the role of the bifunctional M-enzyme, which contains a dehydratase activity (for the formation of the 2,3-double bond) and a hydrogenation (reductase) activity (here reduction of 6-C═O) but no oxygenase activity, as it was demonstrated with incorporation experiments that the 12b-OH group of gaudimycin C derives from water,[348] which is likely inserted through a Michael-type reaction after tautomerism establishing a Δ^4 system between C-7 and C-12b (see **Scheme 23**). A H_2O-derived tertiary alcohol group was found only once before (in the tetracenomycin biosynthesis);[349] however, PgaM-related multifunctional enzymes were found previously in the pathways to the landomycins (2,3-dehydratase-6-reductase LndM2, see **Scheme 20**) and gilvocarcins/jadomycins (2,3-dehydratase-5-oxygenase GilOIV, see **Scheme 26**), the latter with a different set of activities and different mode of action. In addition, oxygenase UrdM of the urdamycin pathway, like the PgaM responsible for the introduction of the angular tertiary 12b-OH group with the same stereochemistry, employs a completely different mechanism: the 12b-OH group in the urdamycins is O_2 derived; evidence for a Baeyer–Villiger mechanism for its introduction exists in the form of shunt product urdamycin L (**Scheme 23**).[313,329] Recently, the dimeric protein structure of PgaE was solved as well.[133]

1.07.3.2.3(ii) Angucyclines of mixed biosynthetic origin The simocyclinones consist of an angucycline *C*-glycoside tethered by a tetraene dicarboxylate linker to an aminocoumarin moiety.[350–355] They were shown to decrease the proliferation of human cancer cells and were found to be antibacterial drugs, as they function as inhibitors of bacterial DNA gyrase, like the clinically used group of quinolone antibiotics. Biosynthetic studies include incorporation experiments, gene cluster analysis, and studies with selected biosynthetic enzymes.[356–360] The different congeners (A–D series) arise from slightly different angucyclinone cores and differ with respect to the degree of completion of the biosynthesis (establishment of *C*-glycosyl moiety, the tetraene linkage, coumarin linkage, and halogenation of the coumarin; **Scheme 24**). Amide synthetase SimL establishes the bond between the coumarin and the all-*trans*-deca-2,4,6,8-tetraenedioic acid linker and was used for the derivatization of other aminocoumarin antibiotics, as well as for the synthesis of a simocyclinone analogue (with methyl group instead of the Cl residue) through precursor-directed biosynthesis feeding the aminocoumarin moiety from novobiocin.[356,359]

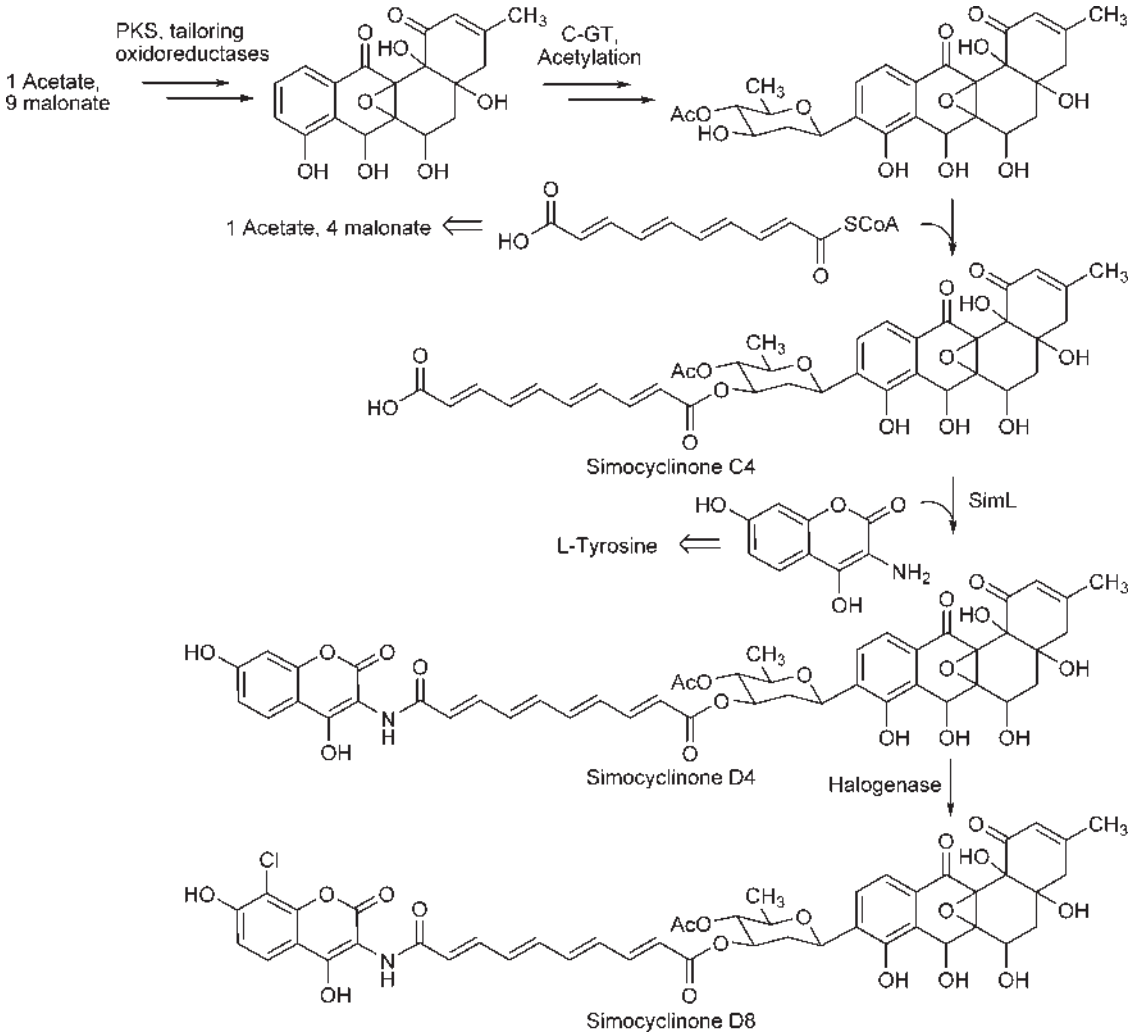
1.07.3.2.3(iii) Angucyclinone-derived natural products with significantly deviated scaffolds (gilvocarcins, jadomycins, kinamycins) The family of gilvocarcin-type anticancer drugs can be subdivided into four subgroups, each with only very few members: (1) those with C-glycosidically linked D-furanoses (e.g., gilvocarcin V, Mer-1020dD),[361–366] (2) those with C-glycosidically linked D-ravidosamine aminosugars (e.g., ravidomycin, FE35A),[367–371] (3) those with branched pyranoses (e.g., chrysomycin A, Mer-1020dC),[365,372–374] and (4) L-rhamnose-containing ones (e.g., BE12406A, polycarcin V)[375–377] (**Scheme 25**).

The gilvocarcins, particularly gilvocarcin V, show excellent antitumor activity with remarkably low toxicity through a unique mode of action. Light-dependent alkylation by [2 + 2] cycloaddition of their vinyl residues with DNA thymine residues along with binding to a core component of the histone complex, histone H3, causes crosslinking. In addition, topoisomerase II inhibition and DNA strand breaking were observed,[378–385] and gilvocarcin V was also shown to possess antiviral activity.[386]

Early biosynthetic studies by Tomita *et al.* and Carter *et al.* using incorporation experiments with ^{13}C-labeled precursors suggested that their unique polyketide-derived benzo[*d*]naphtho[1,2-*b*]pyran-6-one core structures derive from a typical angucyclinone through oxidative rearrangement.[387–389] Other biosynthetically interesting features are the *C*-glycosidic moieties (of most members) and the preference for propionate starter units otherwise not found among the angucycline group antibiotics, except for the *Nocardia* products brasiliquinones A–C.[390] Recent biosynthetic studies, combining gene inactivation, ^{18}O-incorporation, gene recombination experiments, and enzyme studies, focused mainly on the oxidative rearrangement cascade. This involves the bifunctional FAD-dependent dehydratase/oxygenase enzymes GilOIV and GilOI, ultimately responsible for the C–C bond cleavage, and cofactor-free oxygenase GilOII. All are assembled in an oxygenase complex, along with P-450 enzyme GilOIII, which is responsible for the generation of the vinyl group. The last biosynthetic



Scheme 23 Comparison of proposed mechanisms for the multifunctional enzymes PgaM, involved in gaudimycin C biosynthesis (black), and UrdM, an oxygenase/reductase involved in urdamycin biosynthesis (blue). 12b-O in the gaudimycin biosynthetic pathway derives from water, introduced by the multifunctional dehydratase/hydratase/dehydrogenase PgaM. In contrast, 12b-O was shown to derive from molecular oxygen in the urdamycin pathway, introduced by UrdM, likely through a Baeyer–Villiger oxidation mechanism, since part inactivation of UrdM yielded shunt product urdamycin L. The 2,3-dehydratase activity of PgaM resembles those of LanM/LndM2 (landomycin pathway, above) and JadF/GilOIV (jadomycin/gilvocarcin pathways, below).

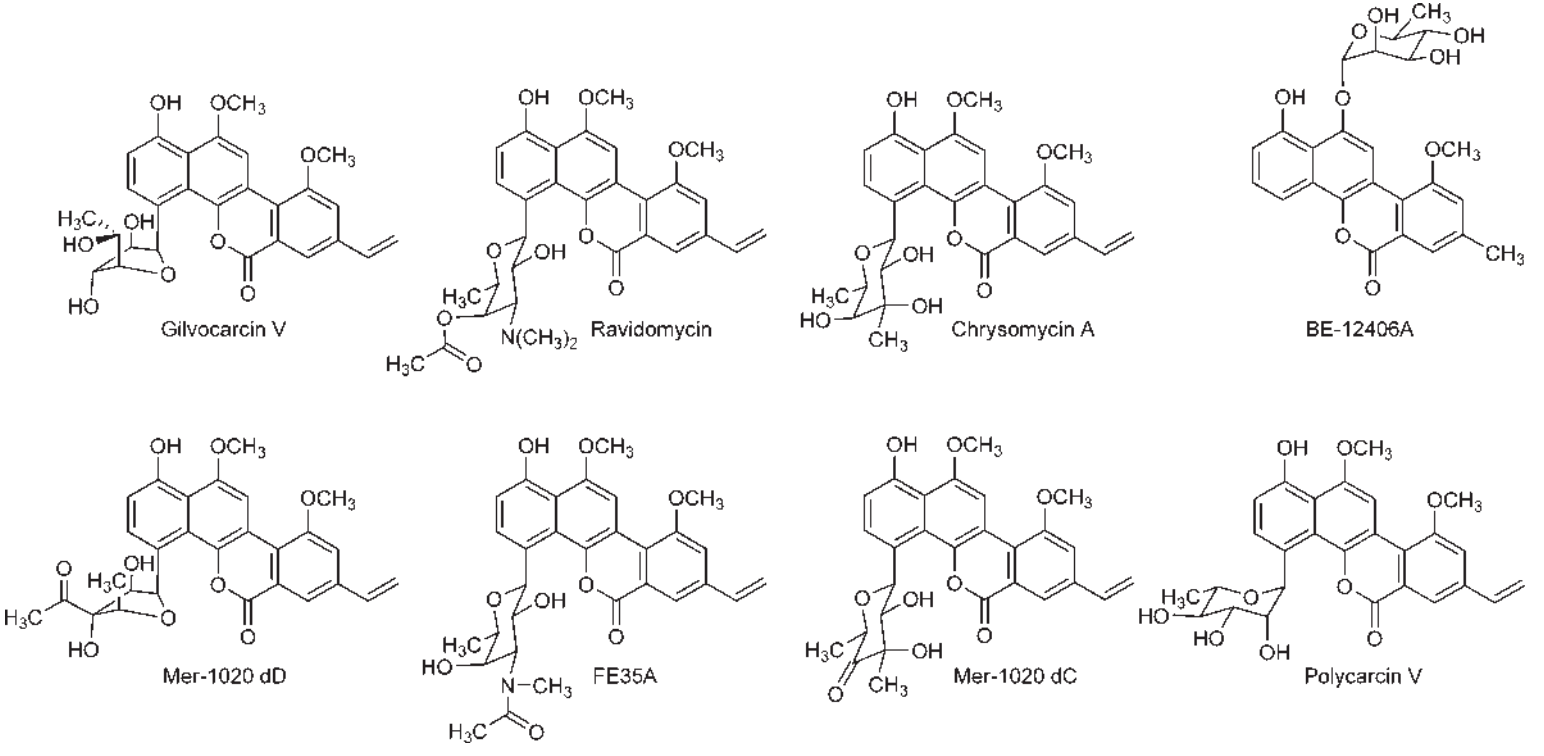


Scheme 24 Possible sequence of the biosynthesis of simocyclinone D8, the most complex simocyclinone antibiotic.

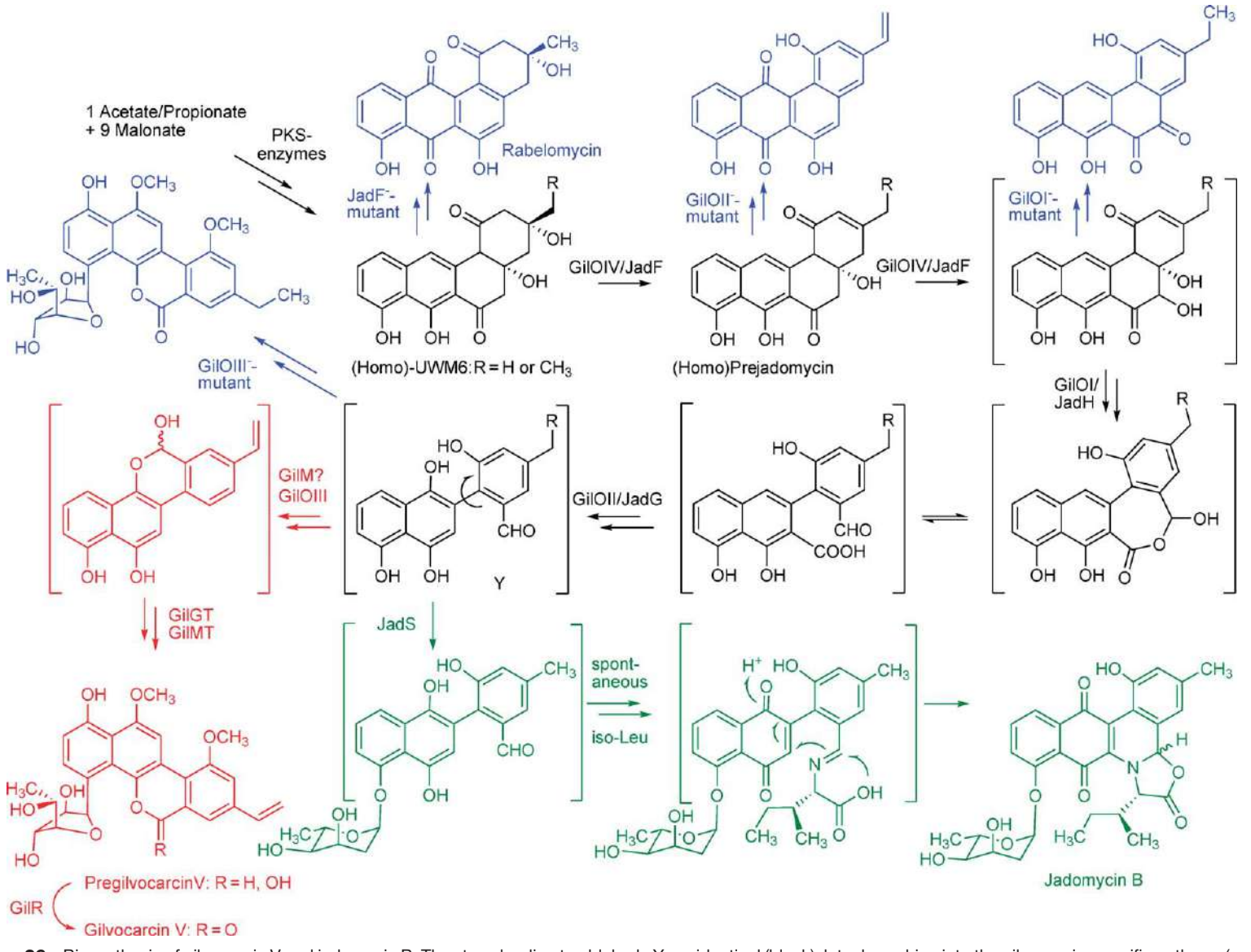
step is a dehydrogenation of pregilvocarcin V to gilvocarcin V catalyzed by GilR, an unprecedented lactone formation reaction in secondary metabolite biosynthesis[33,391–395] (**Scheme 26**).

The antibiotic and antiyeast agent jadomycin B is one of the earliest angucyclines whose gene cluster was identified and studied by the Vining group.[10,92,396–406] Inactivation of oxygenase JadF led to the accumulation of the frequently found angucyclinone rabelomycin,[405] which proved the jadomycins to be angucyclinone derived. Gene cassette assembly studies led to the recognition of the minimal set of genes to establish a benz[*a*]anthracene scaffold typical for the angucyclines and to the production of the important early precursor UWM6.[92,400] The reaction cascade (biosynthetic 'black box') catalyzed by an oxidoreductase complex, which eventually leads to the unique frames of gilvocarcins and jadomycins, yields a reactive biosynthetic intermediate, most likely aldehyde Y.[391,407,408]

For the biosynthesis of jadomycins A and B, this aldehyde reacts nonenzymatically with the amino acid isoleucine, a major component of the growth media used for jadomycin B production.[397,398,402,409] After imine formation, a series of reactions occur to finish the unique nitrogen-containing pentacyclic benz[*b*]oxazolophenanthridine scaffold of the jadomycins, and many of these probably also occur without enzyme support.[408,410,411,412] In this context, the oxazolone ring opening dynamics were discovered[412] and mechanistically studied by the Jakeman group.[413,414] The glycosyltransfer reaction necessary for the formation of



Scheme 25 Representative examples of naturally occurring gilvocarcin-type anticancer drugs.



Scheme 26 Biosynthesis of gilvocarcin V and jadomycin B. The steps leading to aldehyde Y are identical (black), later branching into the gilvocarcin-specific pathway (red) and the jadomycin-specific pathway (green). Shunt products resulting from gene inactivation experiments are depicted in blue. Some of the steps within the 'black box' (sequence with hypothetical intermediates in brackets) may follow a slightly different sequence.

jadomycin B is catalyzed by GT JadS and earlier was always proposed as occurring at a very late stage of the biosynthesis; however, more recently, it was also proposed that this might happen earlier, possibly before the nonenzymatic reaction cascade,[413] as shown here (**Scheme 26**). The gilvocarcin-specific pathway avoids spontaneous quinone formation through quick rotation and hemiacetal formation (possibly catalyzed by a rota-isomerase, most likely GilM). After O-methylation and C-glycosylation, the hemiacetal is oxidized by dehydrogenase GilR to the lactone moiety, in a unique way of lactone formation for secondary metabolites.[391,392]

The kinamycins, produced by *S. murayamaensis*, are antibiotics with activity against Gram-positive bacteria and anaerobes.[415–418] The kinamycins are structurally a unique group of diazafluorenes (originally determined wrongly as cyanocarbazoles).[22,419–422] Their biosynthesis can be envisioned following the same path as gilvocarcins and jadomycins up to aldehyde Y, after which unique steps determine the branch toward the kinamycins: decarboxylation (aided by the β-carbonyl), followed by an aldolase reaction, some oxidation steps including quinone formation leading to kinobscurinone, transamination to stealthin C, diaza group formation to prekinamycin, and further oxidation and acetylations to various kinamycins, which differ from each other in the regioselectivity and degree of acetyl transfer steps (**Scheme 27**). All of the intermediates shown were isolated as well as synthesized by Gould *et al.* and were proven to be intermediates.[2,423,424] Several natural products, for example, FL120B, cysfluoretin, momofulvenone A, and seongomycin, exemplify natural diversity following/diverting from the kinamycin path. While seongomycin was discovered after heterologous expression of parts of the kinamycin gene cluster in *S. lividans*,[425] and – similar to urdamycin E – might have been generated by nonenzymatic Michael addition of *N*-acetyl cysteine, the other compounds are produced by different strains of streptomycetes.[426–428]

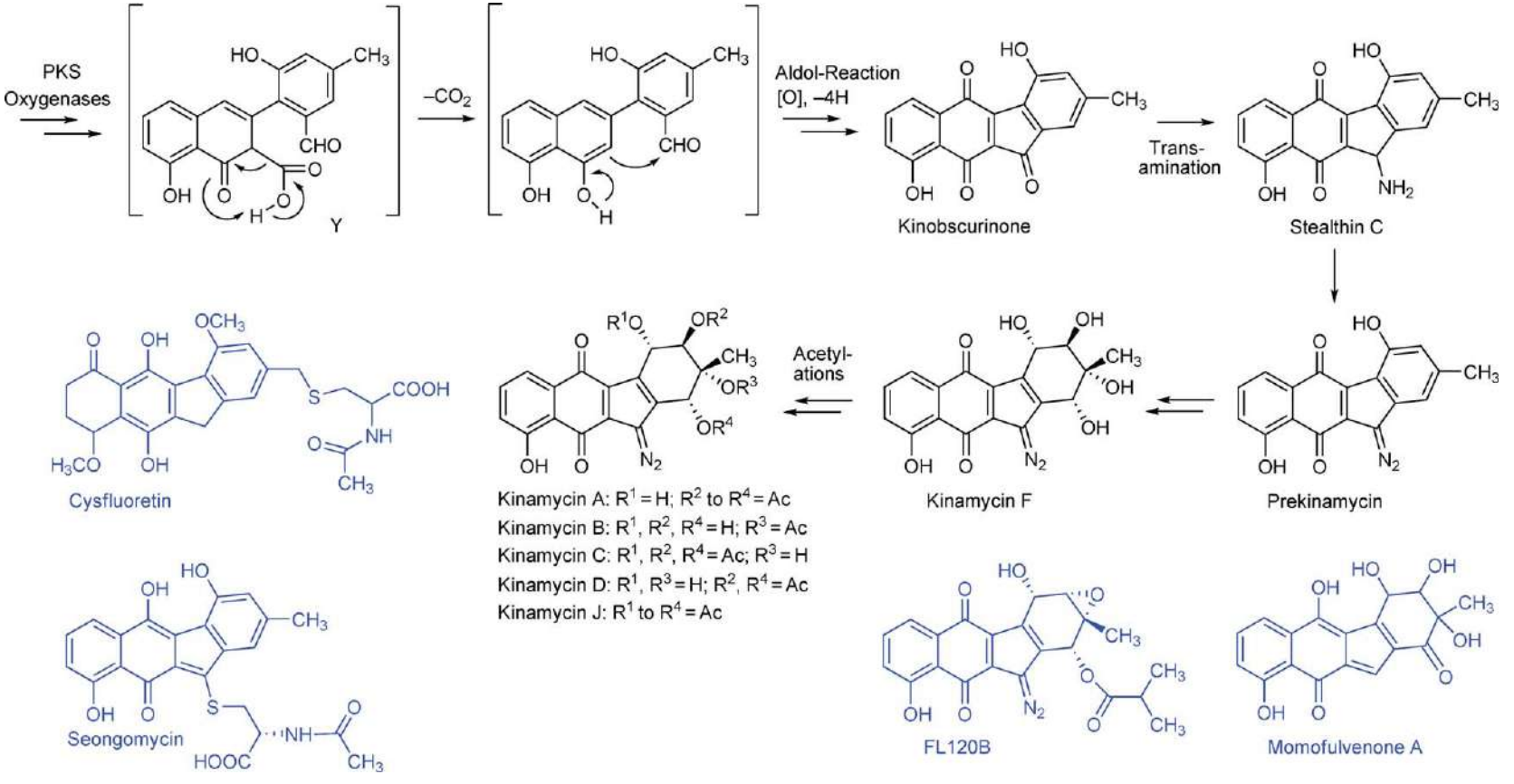
1.07.3.3 Natural Products Derived from Decaketides that Undergo Initial 9,14-Cyclization

1.07.3.3.1 Tetracenomycin folding

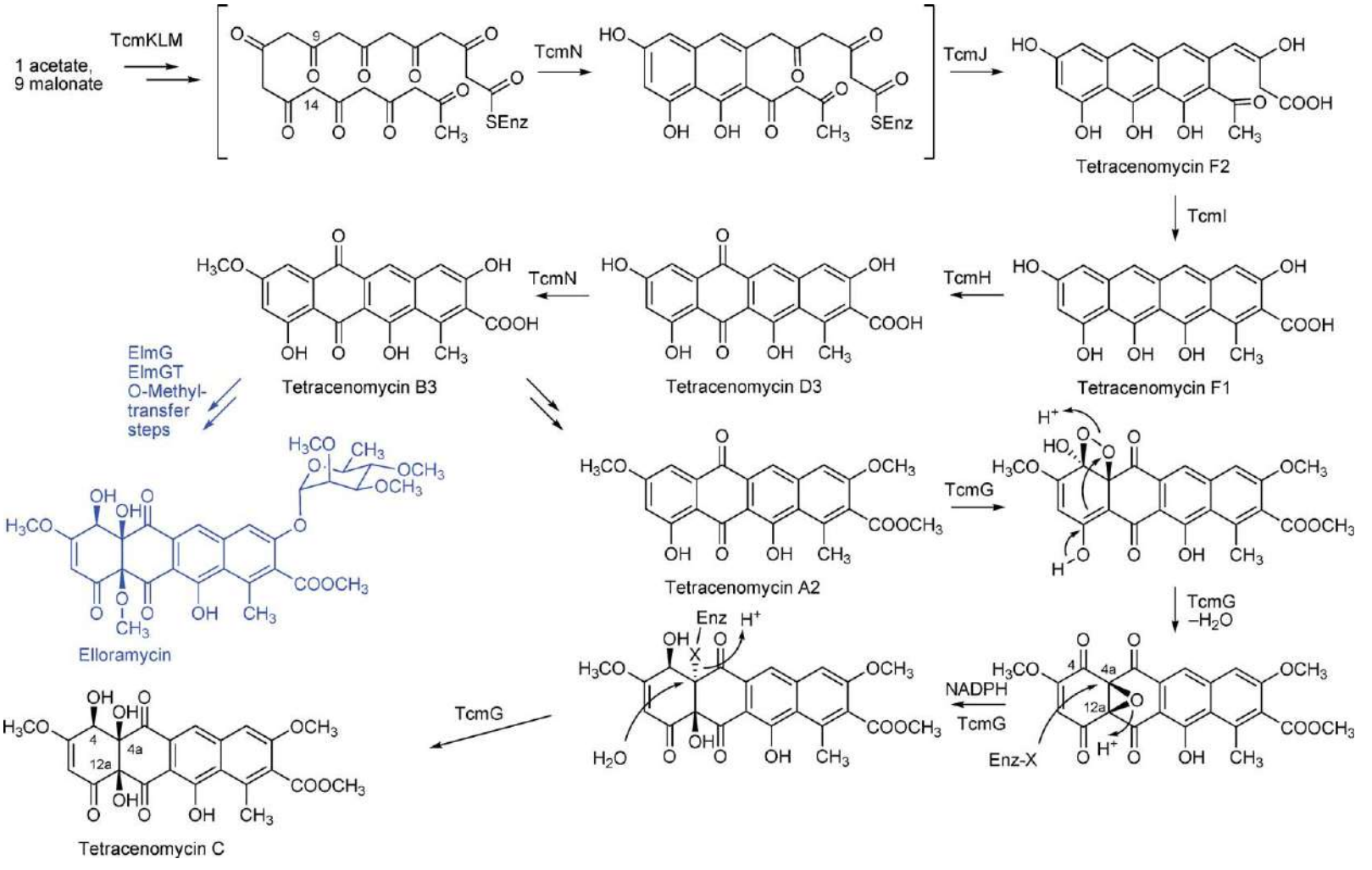
Tetracenomycin C and its related glycosidic analogues, the elloramycins, form a very small, distinct group of linearly assembled tetracyclic decaketides. They were discovered in the Zähner/Zeeck laboratories in the late 1970s/early 1980s as weak anticancer antibiotics.[429–433] Many pioneering biosynthetic/genetic studies were carried out with this class of antibiotics, mainly by the Hutchinson group,[6,57,66,93,94,434–454] and therefore the tetracenomycins became invaluable as model compounds for type II PKS-related investigations as well as for pioneering metabolic engineering studies,[92,442,455] still serving as primary models to understand polyketide cyclization, for example.[93,97,456] Also, some of the post-PKS tailoring enzymes, for example, the first discovered cofactor-free oxygenase TcmH,[457] the uniquely operating oxygenase TcmG/ElmG, and glycosyltransferase ElmGT, were thoroughly investigated,[349,458,459] the last enzyme being the first flexible GT used for combinatorial biosynthetic glycosylations[111,334,460–464] (see Section 1.07.4). Tetracenomycin C was the first aromatic polyketide for which the initial 9,14-cyclization, caused by cyclase TcmN, was discovered, and this occurs relatively rarely compared to the more common initial 7,12-cyclization. The tetracenomycin/elloramycin pathway is one of the best studied biosynthetic pathways and has already been extensively reviewed.[1–3,66,465] Thus, only a short summary with some of the highlights is given here (**Scheme 28(a)**). Cyclase/*O*-methyltransferase TcmN was one of the first bifunctional enzymes discovered, and 1,3-monooxygenase/dioxygenase TcmG has a unique mode of action introducing three oxygen atoms into the tetracyclic scaffold in the very last step of the tetracenomycin biosynthesis (first 4-OH, then the C4-O-O-C4a dioxetane, followed by rearrangement to an epoxide/ketone intermediate). The mechanism shown in **Scheme 28(a)** is consistent with the finding of two oxygens (C-4 and C-12a) derived from molecular oxygen and one (at C-4a) from water. In the elloramycin pathway, this triple hydroxylation by ElmG is followed by the L-rhamnosyl transfer step through ElmGT and sugar methylations.

1.07.3.3.2 Discoid folding

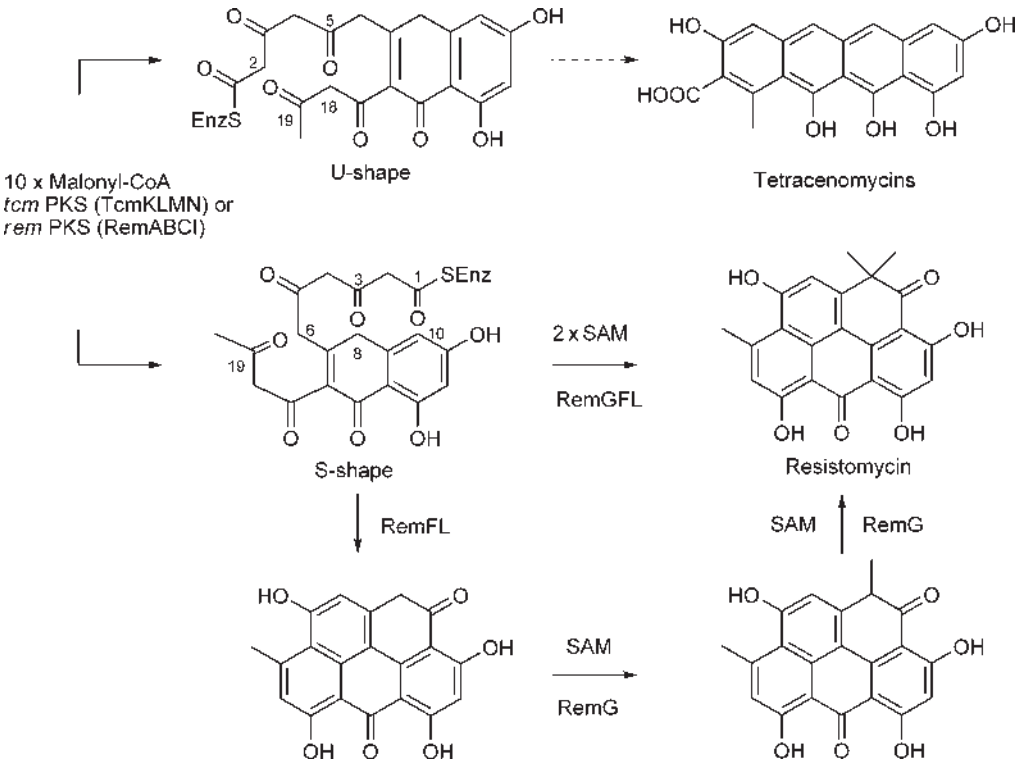
Resistomycin is an unusual naphthanthrene derivative isolated from *S. resistomycificus*[466] that exhibits a variety of pharmacologically relevant properties, for example, inhibition of HIV-1 protease,[467] as well as RNA and DNA polymerase, and activity against Gram-positive bacteria and mycobacteria.[468] Furthermore, it has also been implicated as a modulator of apoptosis.[469]



Scheme 27 Biosynthesis of the kinamycins. For early steps up to aldehyde Y, see **Scheme 26** (gilvocarcin/jadomycin pathway). Shown in blue are examples of natural relatives of the kinamycins, whose biosyntheses would require the same aldol reaction step, but different follow-up reactions.



Scheme 28 (Continued)



Scheme 28 (a) Biosynthetic pathway to tetracenomycin C/elloramycin. Particularly noteworthy is TcmG's monooxygenase mechanism that explains that the 4- and 12a-OH groups of tetracenomycin C derive from two molecules of O_2, and the 4a-OH group from a molecule of H_2O. The TcmG mechanism resembles bacterial and fungal hydroquinone epoxidizing dioxygenases as well as the mammalian vitamin K-dependent γ-glutamyl carboxylase. (b) Model for the discoid resistomycin biosynthesis and geminal bismethylation by RemG.

The structure of resistomycin differs from all other polyphenols produced by type II PKS systems in having a pentacyclic perifused (discoid) ring system. Furthermore, resistomycin features a rare geminal bis-methyl substitution pattern that is derived from two methionine carbons.[470] Resistomycin shares early biosynthetic steps with linear decaketides of the tetracenomycin (*tcm*) family, but both pathways diverge due to varying U- and S-shape cyclization of the putative decaketide.[99] The *rem* biosynthesis gene cluster was cloned, sequenced, and heterologously expressed by Jacobi and Hertweck.[36] By gene knockout, complementation, and *in vivo* reconstitution, all components of the *rem* multienzyme complex that are necessary and sufficient for resistomycin formation were identified.[99] Cross-complementation of the noncanonical minimal *rem* PKS (RemABC) and the first ring cyclase (RemI) with corresponding enzymes from the linear tetracenomycin pathway revealed that RemI and TcmN are functionally equivalent and that both pathways share early biosynthetic steps. Two additional cyclases, RemL and RemF, were identified that are required for the controlled cyclization of the reactive decaketide to give rise to the unparalleled naphthanthrene system. In contrast to various other type II PKS pathways, the *rem* cyclases cannot be attributed to individual, stepwise cyclization events. These results provide strong support for a model according to which the multienzyme complex forms a cage in which the polyketide is shaped. Thus, in the *rem* pathway, all components need to work in a highly concerted action in order to yield a polyphenol with a distinct ring topology (**Scheme 28(b)**).[99]

After the pentacyclic ring system has been assembled, the geminal methyl groups are introduced by RemG,[471] a novel type of bis-*C*-methyltransferase.[472] Further hydroxylation by RemO[473] gives rise to the boat-shaped derivative resistoflavin (**Scheme 29**).[474]

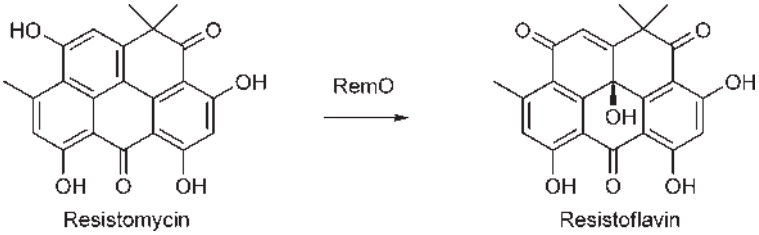
1.07.3.4 Natural Products Derived from Larger Polyketides

Benzo[*a*]naphtacenes, such as pradimicin and benastatin, also feature an angular polyphenolic ring system and may be regarded as extended angucyclines or pentangular polyphenols. The pradimicins have a broad antifungal and antiviral spectrum,[475] whereas the benastatins and bequinostatins inhibit glutathione-*S*-transferases (GSTs) and exhibit antitumoral activities.[476–478]

The pradimicin (*pdn*) biosynthesis gene cluster from *Actinomadura hibisca* was partially cloned and sequenced in 1997 by Dairi *et al.*[19] More recently, the entire 39 kb *pdn* gene locus was sequenced by Kim and coworkers, revealing 28 open reading frames (ORFs) associated with aglycone assembly, tailoring reactions, and regulation.[21] The biosynthetic pathway was partially reconstituted in a heterologous *Streptomyces* host by Tang and coworkers.[479] Accordingly, the minimal PKS PdmABC assembles the dodecaketide, and PdmDHKL furnishes the primary pentangular core before the tailoring reactions set in (**Scheme 30**).

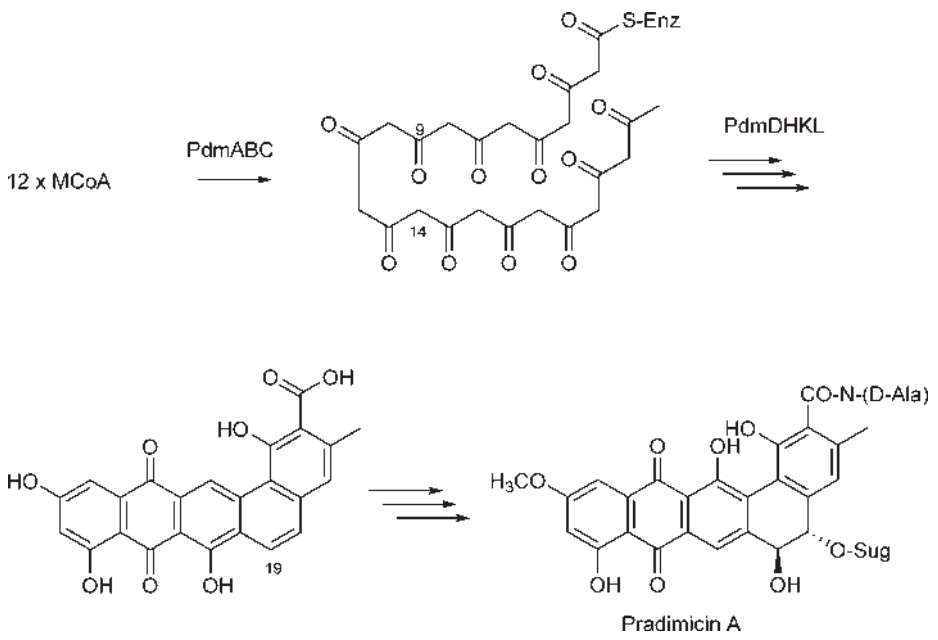
The benastatin biosynthesis gene cluster, which was cloned, sequenced, and heterologously expressed by Hertweck and coworkers, shares high similarity with the *pdm* locus. However, the *ben* pathway (**Scheme 31**) is initiated by priming the KS with a hexanoate starter unit in lieu of acetate, giving rise to the unique pentyl side chain. A designated KSIII component, BenQ, has an editing role in selecting the correct starter. In its absence, various derivatives are produced from typical short-chain linear and branched FAS starter units.[49,77]

Benastatin also deviates from most other polyphenols in featuring a rare geminal dimethyl group where usually a quinone carbonyl is found.[472] Surprisingly, a single MT (BenF) is sufficient for attaching both methionine-derived methyl groups to the anthrone methylene. However, methylation at the highly reactive benzylic position competes with oxygenation. This is evidenced by the formation of hydroxymethyl

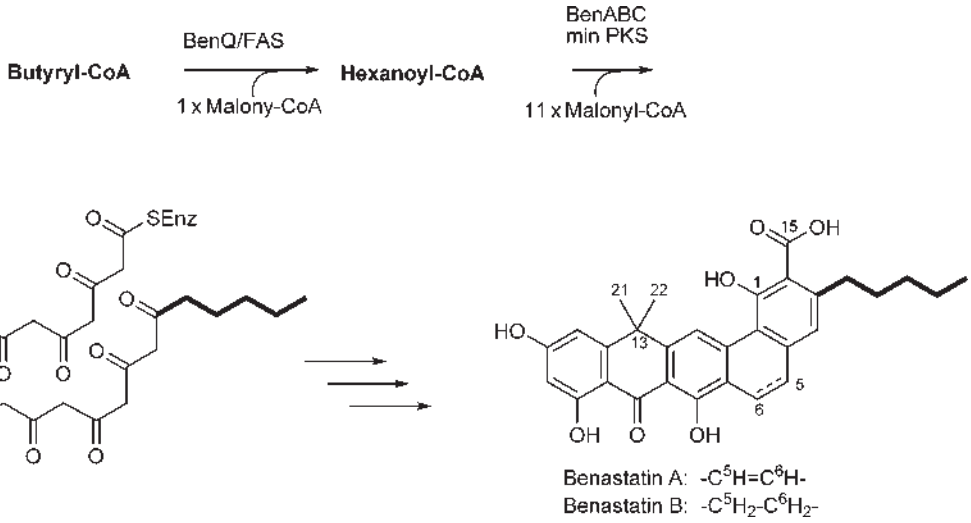


Resistomycin Resistoflavin

Scheme 29 Resistomycin results from central RemO-catalyzed hydroxylation of resistomycin.



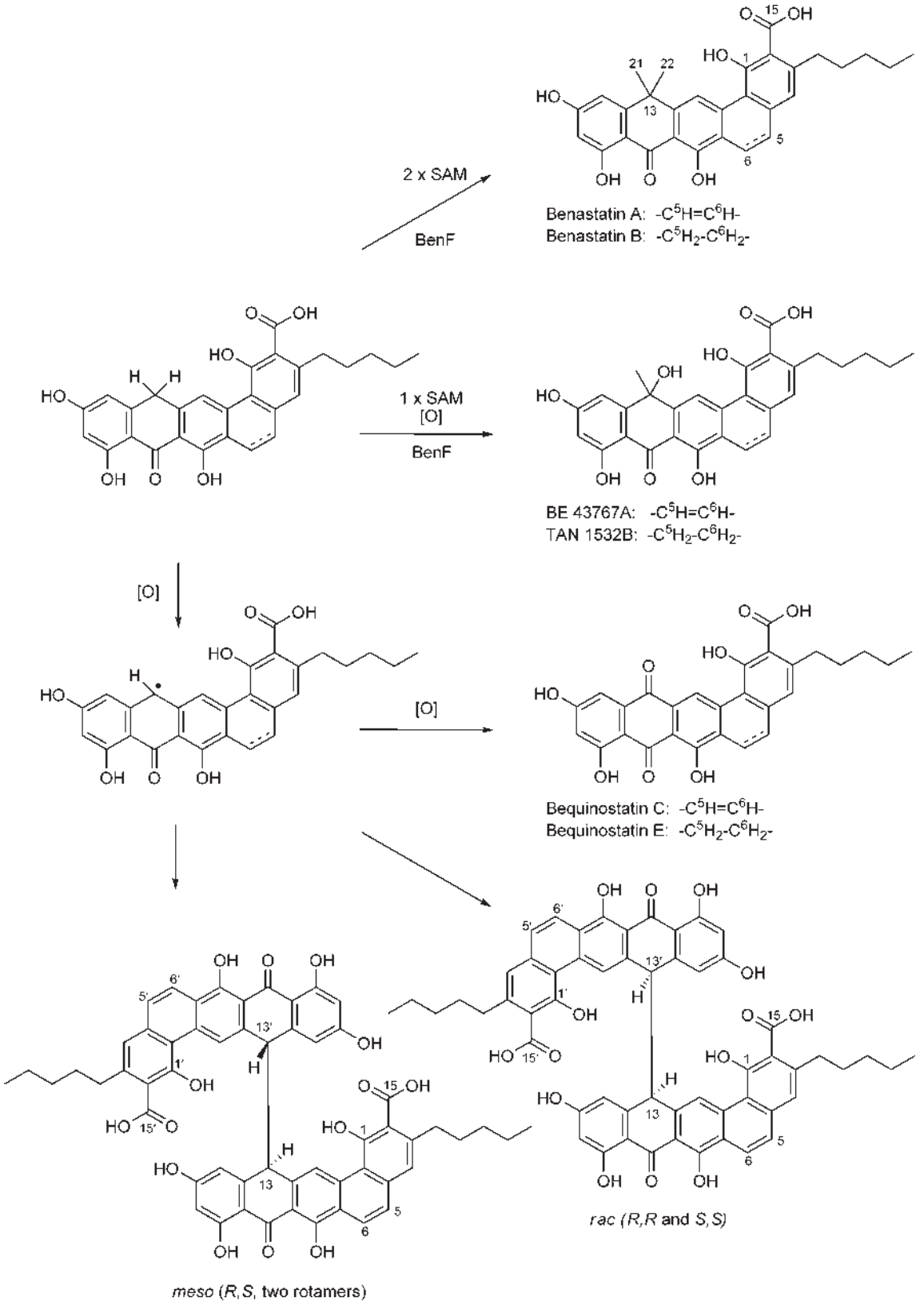
Scheme 30 Model of pradimicin biosynthesis.



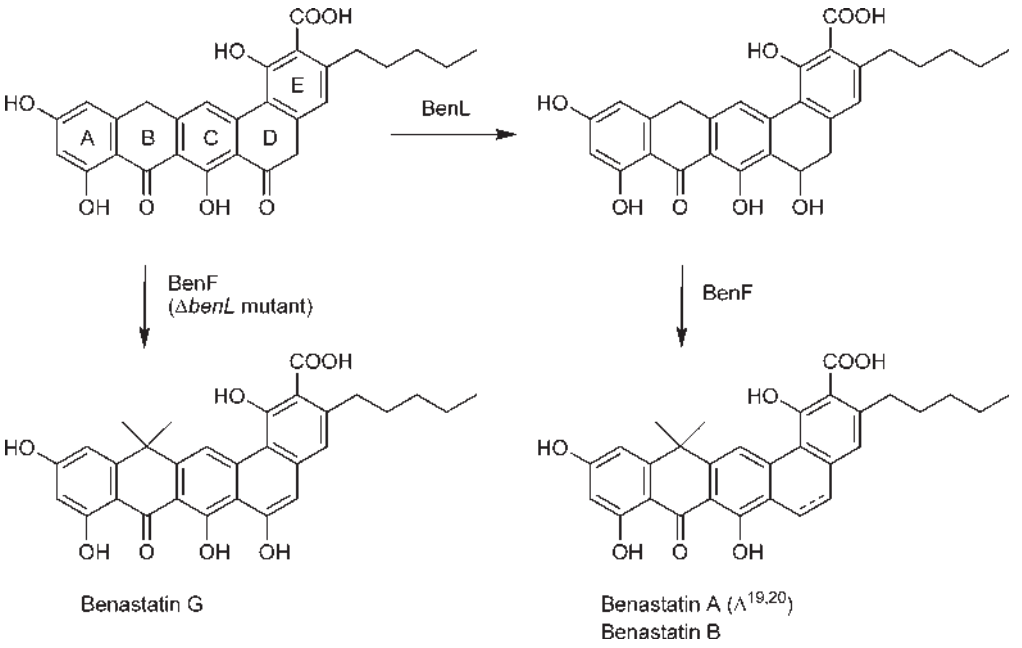
Scheme 31 Biosynthesis of benastatin from a hexanoate starter and 11 malonyl units.

derivatives when fermenting the wild type with increased aeration. A mutant lacking BenF yields anthraquinone analogues as well as unusual anthrone dimers resulting from coupling of benzylic radicals (**Schemes 32** and **33**).[472]

A hallmark of pentangular polyphenols is the reduced D ring at the 'bent' of the molecules. Mutational analyses and a phylogenetic survey conducted by the Hertweck and Piel groups revealed that ketoreduction is catalyzed by a group of ketoreductases (KRs) that are characteristic for these polyketide metabolites. A KR null mutant (ΔbenL) resulted in the formation of a series of 19-hydroxy benastatin and bequinostatin derivatives (**Scheme 33**), clearly demonstrating the function of the C-19 KR in pentangular pathways. Notably, genes coding for similar KRs were also found in gene clusters for the biosynthesis of the unusual spiro compounds rubromycin, griseorhodin, and fredericamycin, highlighting their biogenetic relationship.



Scheme 32 Geminal bismethylation in benastatin biosynthesis prevents oxygenation and dimer formation.



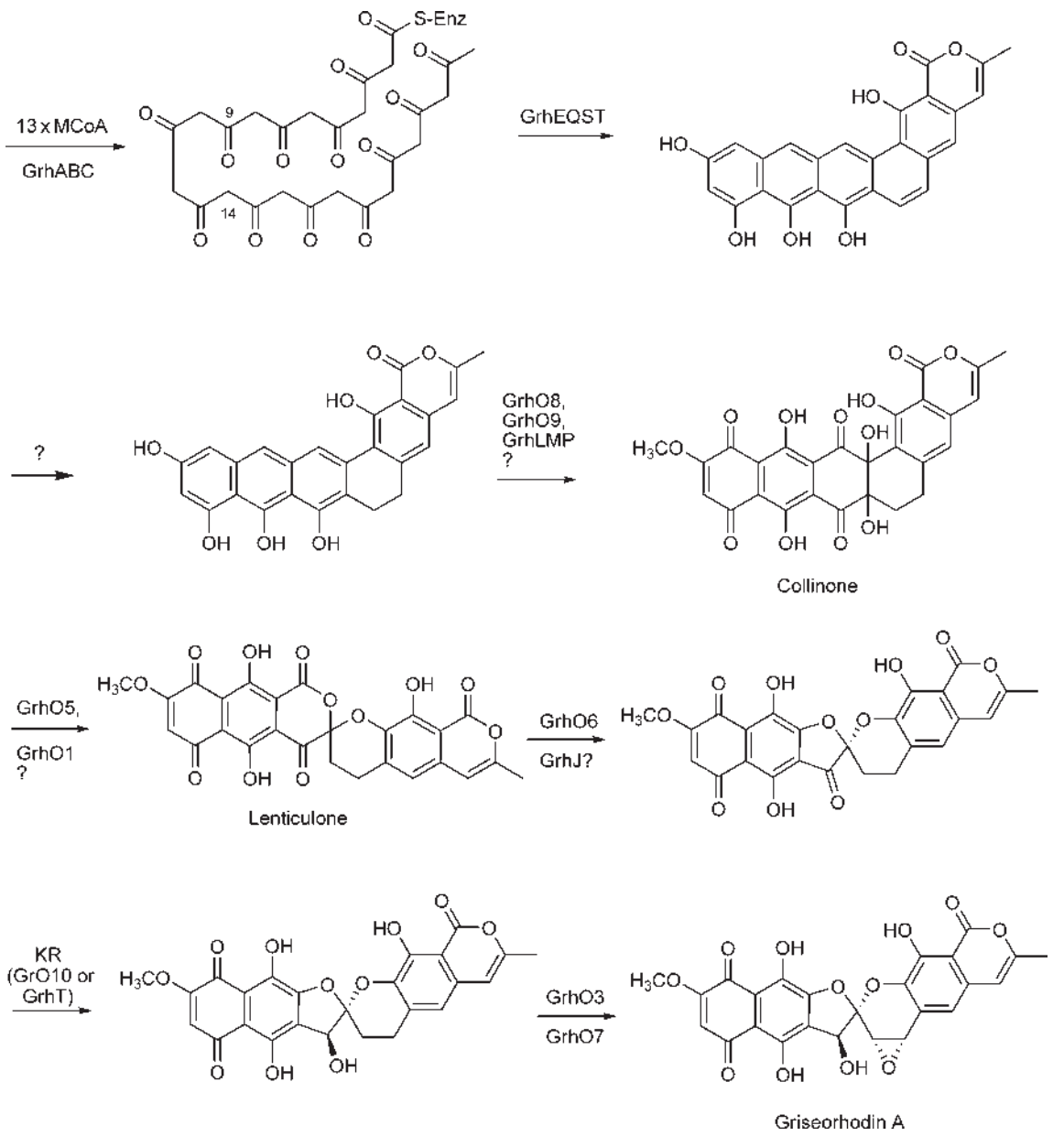
Scheme 33 BenL as a prototype for pentangular-specific D-ring ketoreductases.

Griseorhodin and rubromycin (**Scheme 34**) are telomerase inhibitors that share a spiroketal moiety. The isolation of collinone from a bacterium expressing an uncharacterized DNA segment from the rubromycin producer[28] provided circumstantial evidence that rubromycin-type compounds are biosynthetically related to pradimicins and benastatins. Li and Piel have identified gene cluster coding for griseorhodin biosynthesis in a marine *Streptomyces* sp.[30] and proposed a biosynthetic model on the basis of putative gene functions. More recently, lenticulone isolated from an engineered oxygenase mutant shed more light on the rare rearrangement process, which involves the cleavage of four carbon bonds.[480]

Fredericamycin A (**Scheme 35**) is a cytotoxic pentadecaketide from *Streptomyces griseus* featuring two sets of perihydroxy tricyclic aromatic moieties connected through a unique asymmetric carbaspiro center. Shen, Hutchinson, and coworkers have cloned and sequenced the 33 kb fredericamycin (*fdm*) biosynthesis gene cluster.[44] The fdm PKS, cyclase, and KR genes are highly similar to characterized genes involved in the biosynthesis of pentangular polyketides, thus suggesting shared early steps in the pathways. This assumption is strongly supported by the detection of the putative intermediates fredericamycins B and C_1 by Sontag and coworkers in a *S. griseus* culture.[481] The identification of fredericamycin E from a culture of the fredericamycin A producer by Shen and coworkers fills another gap in the biosynthetic model and sheds more light on this complex rearrangement sequence.[482] The transformation of the spiro[4.5]decane framework into the final spiro[4.4]nonane can be rationalized by a biosynthetic benzilic acid-like rearrangement.

1.07.3.4.1 Angucylines from angular heterocyclic folding

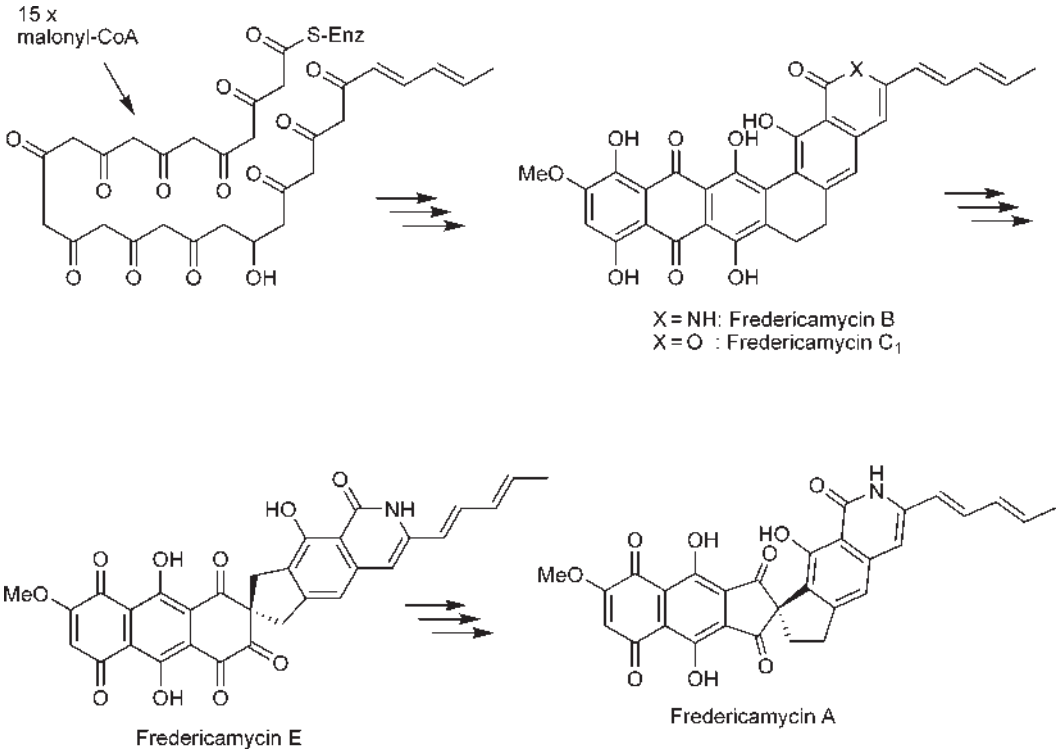
Hedamycin (**Scheme 36**), pluramycin, and altromycin represent a small group of polyphenols that share an angular heterocyclic, or more precisely 4-*H*-anthra[1,2-b]pyrane, ring system. They are DNA intercalators and alkylators, and thus potent antitumor and antimicrobial agents.[109,483] Thorson and coworkers have cloned the entire 45 kb gene cluster coding for hedamycin biosynthesis in *S. griseoruber* and proved its identity through gene inactivation.[34] Sequence analyses revealed the presence of type I (*hedT*) and minimal type II PKS (*hedCDE*) genes, as well as gene coding for a KSIII component (*hedS*). According to a biosynthetic model, HedT is an iterative type I PKS and produces the hexadienoate chain starter unit. Although the AT HedF likely functions as a shuttle, the KSIII (HedS) may regulate the priming of the type II PKS.[34] Although candidate genes for cyclization, oxygenation, sugar biosynthesis, and glycosylation have been identified, functional analyses have not yet been conducted to reveal hedamycin assembly (**Scheme 36**).



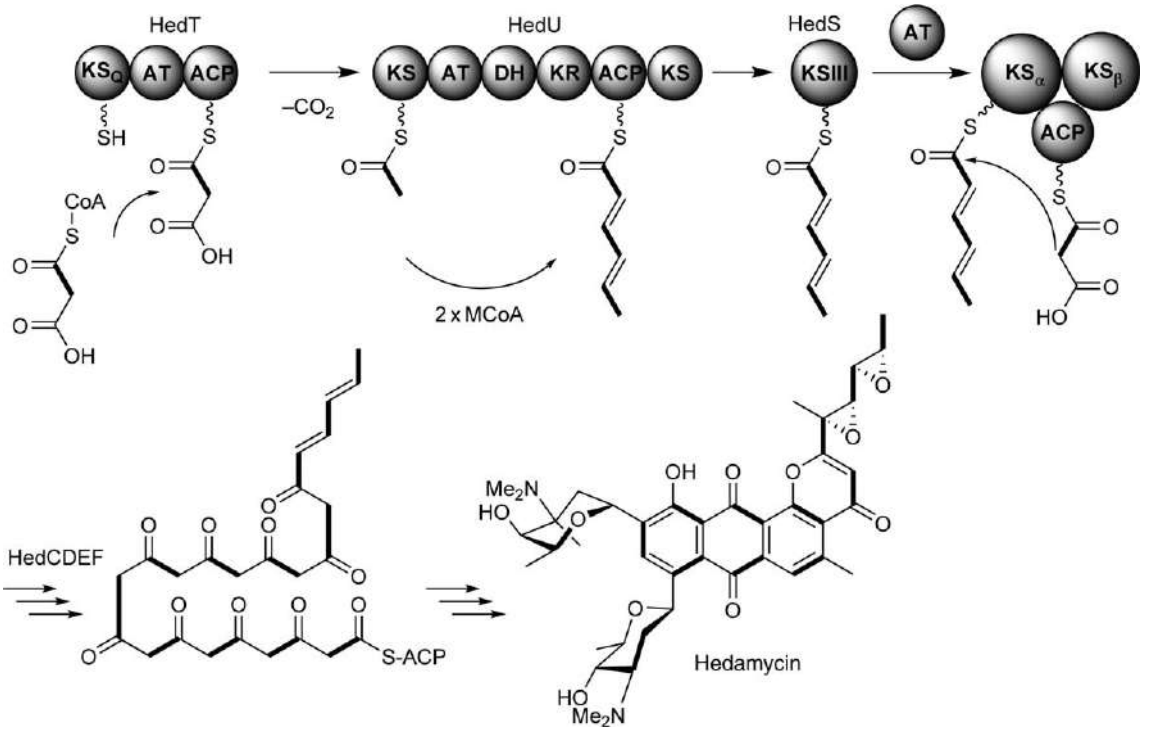
Scheme 34 Model for the biosynthesis of the spiroketal griseorhodin A.

1.07.4 Artificial Structural Diversity through Combinatorial Biosynthesis

The commonly applied term 'combinatorial biosynthesis', originally coined for any manipulation of biosynthetic pathways of natural products to generate new natural products or natural product derivatives through genetic engineering, encompasses two approaches with slightly different but clearly overlapping goals: (1) pathway engineering for drug modification or optimization and (2) generation of libraries of 'unnatural natural products' or hybrid natural products through combination of sets of genes from different pathways.[484] Although the first approach usually strives for one or a limited number of compounds, usually derivatives of a certain natural product, the second approach is aimed at libraries of compounds and/or structures with heretofore unknown structural scaffolds. Both approaches originated from early genetic engineering experiments



Scheme 35 Model for the biosynthesis of fredericamycins.



Scheme 36 Biosynthesis of hedamycins.

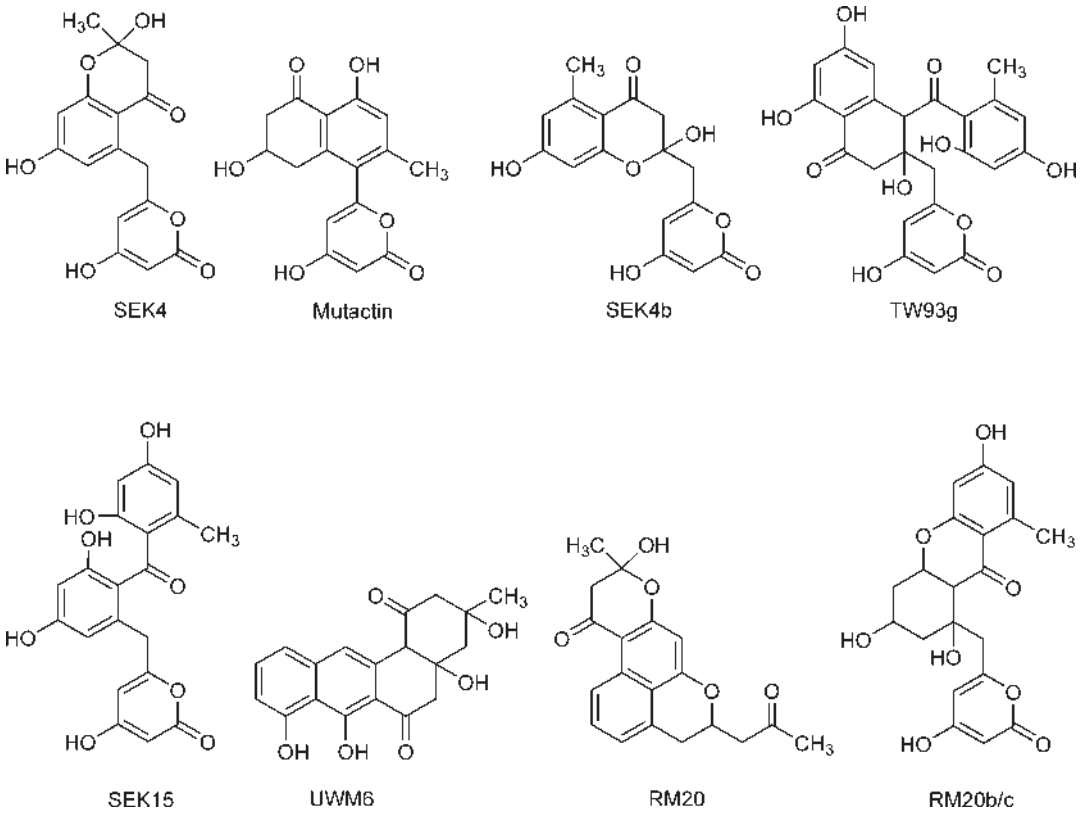
manipulating the pathways of aromatic polyketides. The first approach usually does not deviate too far from nature's molecules with evolutionary optimized biological activities and perfectly complements semisynthetic natural product derivatizations, with both methods having different limitations. Thus, usually biologically active molecules are envisaged, and first examples exist in which this approach yielded molecules with biological activities that are superior in some aspects than the parent natural drugs. The second approach can be used to generate structural diversity but comes with the drawback that the compounds generated are no longer optimized by evolution for biological activity.[484] Both approaches obviously overlap, and clear categorization is not always possible. Both approaches are also useful delineating and gaining an understanding of the complex biosynthetic pathways. In the context of aromatic polyketides, the first approach is typically aimed at post-PKS tailoring steps, whereas the second approach deals with the manipulation of the polyketide synthases themselves.

1.07.4.1 'Unnatural Natural Products' Generated by Manipulation of Type II Polyketide Synthases

As described above, aromatic (type II) polyketide synthases, similar to type II fatty acid synthases, consist of a (minimal) set of three proteins required for the polyketide chain assembly: a β-ketosynthase (KS or KS_α), a chain length factor (CLF, often also referred to as KS_β), and an ACP. This minimal set of iteratively operating proteins catalyze several (at least 18) separate reactions involved in loading, initiation, extension, cyclization, and release and is usually complemented functionally by a malonyl-CoA:ACP transacylase (MCAT, sometimes abbreviated as MAT), which is often recruited from the fatty acid biosynthesis and accelerates the otherwise rate-limiting loading.[2,3,66,465,485–487] Only in some cases can the genes encoding MCAT can be found in the biosynthetic gene clusters of aromatic polyketides (see above).[31,33,452] The type II PKSs also contain various additional subunits, such as cyclases (CYCs), aromatases (AROs), and KRs, which further process the still ACP-tethered polyketide chain into an initial (primary) frame, which is then released from the ACP. These additional subunits are also often called PKS-associated proteins, in contrast to post-PKS tailoring enzymes (see below), which act after the release from the ACP. The most important post-PKS tailoring enzymes are oxygenases (OXs), KRs, GTs, MTs, halogenases, and aminotransferases (AMTs). Many aromatic polyketides, several of which possess molecular frameworks previously not known from nature, have been generated through manipulations of type II PKS systems, for which various combinations of minimal PKS sets with PKS-associated proteins were used.[2,176] More recently, exceptional enzymes involved in the priming process of type II PKS were found[34,82] and already exploited to generate artificial aromatic polyketides through combinatorial biosynthesis and related methods.[9,80,263,488] In addition, several new natural product derivatives were generated through integration/manipulation of post-PKS tailoring enzymes.[1,106,113,148,149,268,338,343,344,489–496]

1.07.4.1.1 New aromatic polyketides by manipulation of the PKS and PKS-associated enzymes

Pioneering studies by Hopwood, Khosla, and coworkes in the early- to mid-1990s led to the development of a host (*S. coelicolor* CH999)–vector (pRM5) expression system, which allowed the construction of novel aromatic polyketides through the expression of incomplete gene sets or hybrid gene combinations of type II PKSs, predominantly recruited from the *act* and *tcm* pathways.[2,89,164,456,487,497,498] These early experiments, most of which were designed by the Khosla group, turned out to be crucial in recognizing the function of the proteins of aromatic PKSs, defining minimal PKS (consisting of KS, CLF, and ACP), including its CLF, and led to the introduction of the terms 'unnatural natural products' and 'combinatorial biosynthesis'. Many compounds, several with an unprecedented molecular scaffold, were generated by reducing the biosynthetic machinery to the minimal PKS or the minimal PKS plus additional PKS-associated enzymes, such as KRs, CYCs, or AROs/CYCs. Furthermore, experiments showed that a wide combination of such PKS elements (minimal PKS and associated enzymes) was even possible when these were recruited and coexpressed from different biosynthetic pathways. This work has already been extensively reviewed, and some representative examples of new 'unnatural' natural products generated by reduction of type II PKSs or early mixing and matching experiments of components of type II PKSs are shown in **Scheme 37**.[2,485,499–504] Depending on which minimal PKS components were used, the size of such artificially generated unnatural natural multicyclic aromatic molecules typically ranges from octaketides to dodecaketides. Their structures also depend largely on which spontaneous



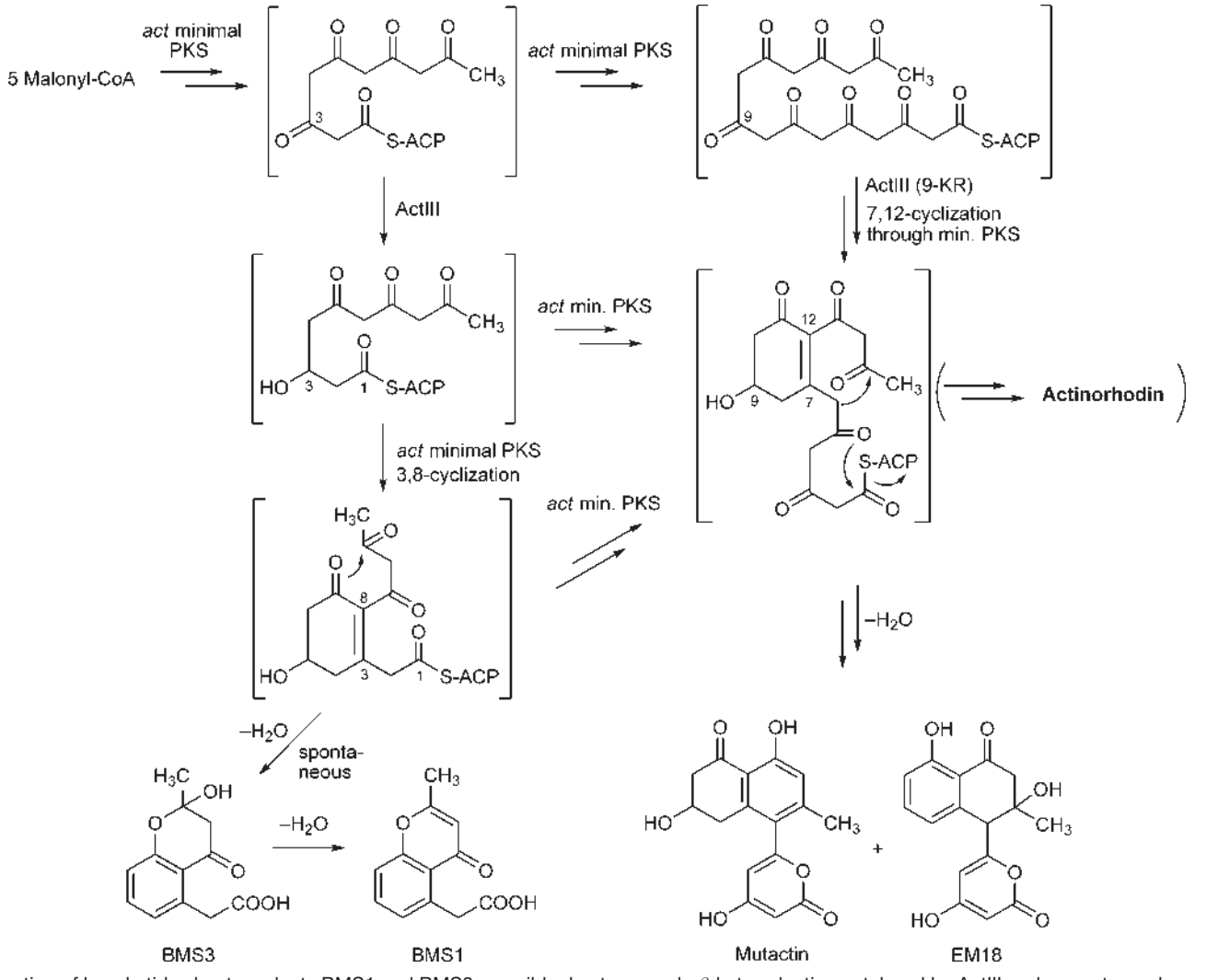
Scheme 37 Examples of early 'unnatural natural product' aromatic polyketides through selected combination of sets of PKS and PKS-associated enzymes, either from only one pathway (e.g., octaketides SEK4, SEK4b, and mutactin; decaketides SEK15 and UWM6; and dodecaketide TW93g) or hybrid compounds from mixtures of enzymes of different pathways (e.g., decaketides RM20 and RM20b/c). For details, see references cited in the text.

cyclization reactions occur first; consequently more products are formed if polyketides with longer chain length are involved.[17] A typical structural element of many of these compounds is the 4-hydroxy-2*H*-pyran-2-one ring often formed upon release from the ACP. Only recently, protein crystal structures and basic kinetic research on such minimal PKS proteins have provided a better understanding of why such a recombination of components from various biosynthetic pathways works.[9,17,72,84,88,97,485,505–514]

Similar, more recent investigations using enzyme combinations of the *act* pathway yielded previously undiscovered shunt products. Through heterologous expression of the actinorhodin PKS in the actinorhodin biosynthesis-deficient host *S. lividans* K4-114, Moore *et al.* found two new hexaketide derivatives, BSM1 and BSM3 (**Scheme 38**), besides previously known octaketide shunt products, such as mutactin or EM18. These findings, along with experiments on the related enterocin/wailupemycin pathway, suggested that the PKS-associated ketoreduction step by the endogenous KR ActIII in actinorhodin biosynthesis can occur already during and not necessarily only after completion of the octaketide chain,[487] preferably when in the β-position to the thioester carbonyl, known from the dissociated bacterial type II fatty acid synthase (FAS).[515] The same conclusions were drawn for KR EncD and the enterocin biosynthesis.[176]

1.07.4.1.2 Artificial aromatic polyketides through utilization of enzymes involved in the priming process

The nature of the primer unit can provide important structural features and influence the biological activity of natural products. Unnatural primer units were tolerated by the loading AT domains of type I PKSs, leading to a successful strategy in manipulating type I PKSs and to new variants of natural products.[516,517] Most aromatic polyketides use acetate starter units, which are usually generated by decarboxylation of malonate, but many



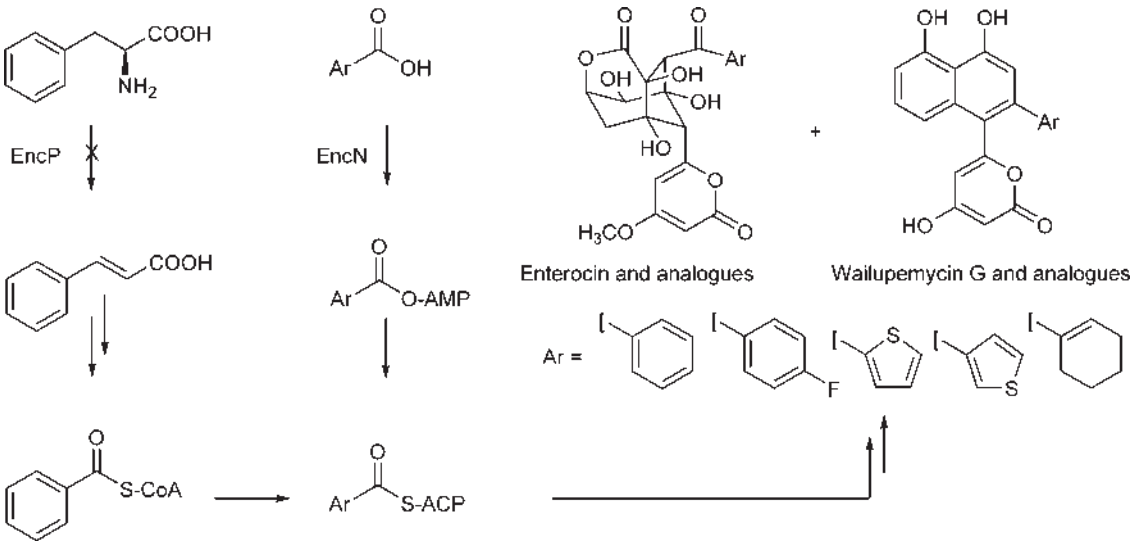
Scheme 38 Formation of hexaketide shunt products BMS1 and BMS3, possibly due to an early β-ketoreduction catalyzed by ActIII and premature release from the ACP.

exceptions were found. For instance, the biosyntheses of several aromatic polyketides utilize propionate (methylmalonate) instead of acetate as a starter unit, for example, most of the anthracyclines and some angucycline group antibiotics (brasiliquinones and gilvocarcins, see above). Often, the 'normal' acetate-primed molecules were found to be congeners. However, examples of type II PKS-encoded pathways, in which truly unusual priming processes are utilized, are the tetracycline pathway (see oxytetracycline, above), the lysolipin pathway,[518] and the enterocin (see above) pathway. Only the enterocin machinery and, most recently, the oxytetracycline pathway were exploited for the generation of new artificial aromatic polyketides. It was shown that notable variations in the frameworks of natural products and 'unnatural' natural products can be brought in when such truly unusual priming processes are involved.

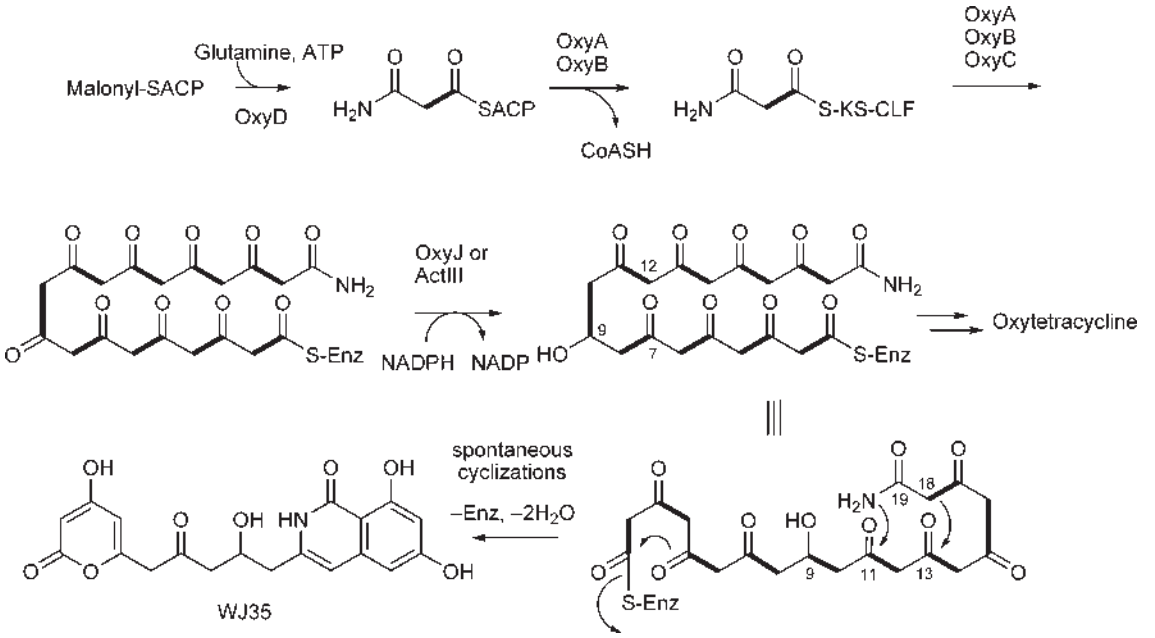
1.07.4.1.2(i) Manipulation of an NRPS-like priming mechanism involved in the enterocin pathway Based on their investigations of the enterocin pathway, the Moore group generated a series of enterocin and wailupemycin analogues through mutasynthesis. The unusual priming process of this pathway, in which benzoyl-ACP is generated either from phenylalanine via benzoyl-CoA involving phenylalanine ammonium lyase (PAL) EncP as a key enzyme or from free benzoic acid via benzoyl-AMP by benzoate:ACP ligase EncN, is flexible enough to accommodate various primer units, namely aryl acids such as monosubstituted benzoates and heteroaromatic isosters, but also cyclohex-1-ene carboxylate, when the PAL (EncP) pathway branch is inactivated (**Scheme 39**).[80,82] However, compared to mutasynthesis experiments with modular type I PKSs, enterocin type II PKS is less flexible. Studies showed that the activating benzoate:ACP ligase EncN has relatively broad substrate specificity, and limits are imposed by processing downstream enzymes.

The recently achieved enzymatic total synthesis of enterocin/wailupemycin polyketides[175] opens up further possibilities to expand the spectrum of artificial structural biodiversity, for example, by including rationally modified enzymes.

1.07.4.1.2(ii) Nitrogen-containing unnatural polyaromatic compounds using amidotransferase OxyD of the oxytetracyclin pathway The oxytetracycline (Oxy) biosynthetic pathway (see Section 1.07.3) contains two enzymes that introduce N-atoms into the Oxy scaffold: OxyD and transaminase OxyQ. The Tang group exploited one of these enzymes successfully to generate new nitrogen-containing multicyclic aromatic compounds, which resemble natural alkaloids. Through expressing amidotransferase OxyD with an extended *oxy*-encoded minimal PKS in *S. coelicolor* CH999, pyran-2-one-isoquinolone WJ35 was generated.[9] Besides the amidotransferase OxyD, which is involved priming the oxytetracycline pathway, the other crucial addition was



Scheme 39 Generation of enterocin (R = Phe) and wailupemycin (R = Phe) analogues by mutasynthesis taking advantage of flexibility of priming enzymes, particularly benzoate:ACP ligase EncN. The EncP-mediated pathway was inactivated.



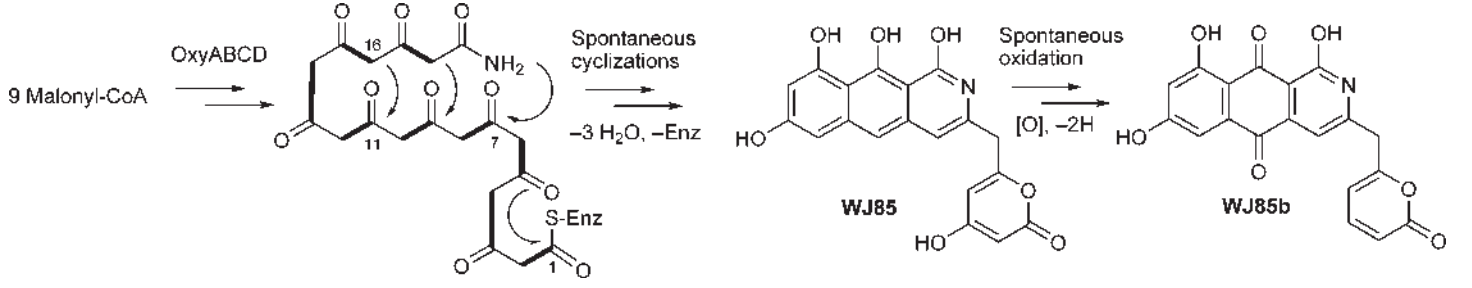
Scheme 40 Generation of isoquinolone WJ35 by the minimal oxy-encoded PKS, amidotransferase OxyD, and a 9-ketoreducase (OxyJ or ActIII).

the involvement of a PKS-associated 9-KR, either OxyJ or ActIII. In the absence of first cyclase OxyK, which normally would have catalyzed the C-7/C-12 cyclization necessary for the *oxy* pathway, a premature C-13/C-18 cyclization occurs, followed by the N-attack on carbonyl C-11. All ring cyclizations including the establishment of the N-heterocycle appear to occur spontaneously (**Scheme 40**). It remains unclear in which sequence the other rings of WJ35 are closed and when the release from the ACP occurs. The most likely scenario is C-13/C-18 cyclization while the polyketide is still tethered to the ACP OxyC, followed by release from the enzyme with concomitant pyranone ring formation. The nucleophilic attack of the amide nitrogen is likely the last reaction toward WJ35.

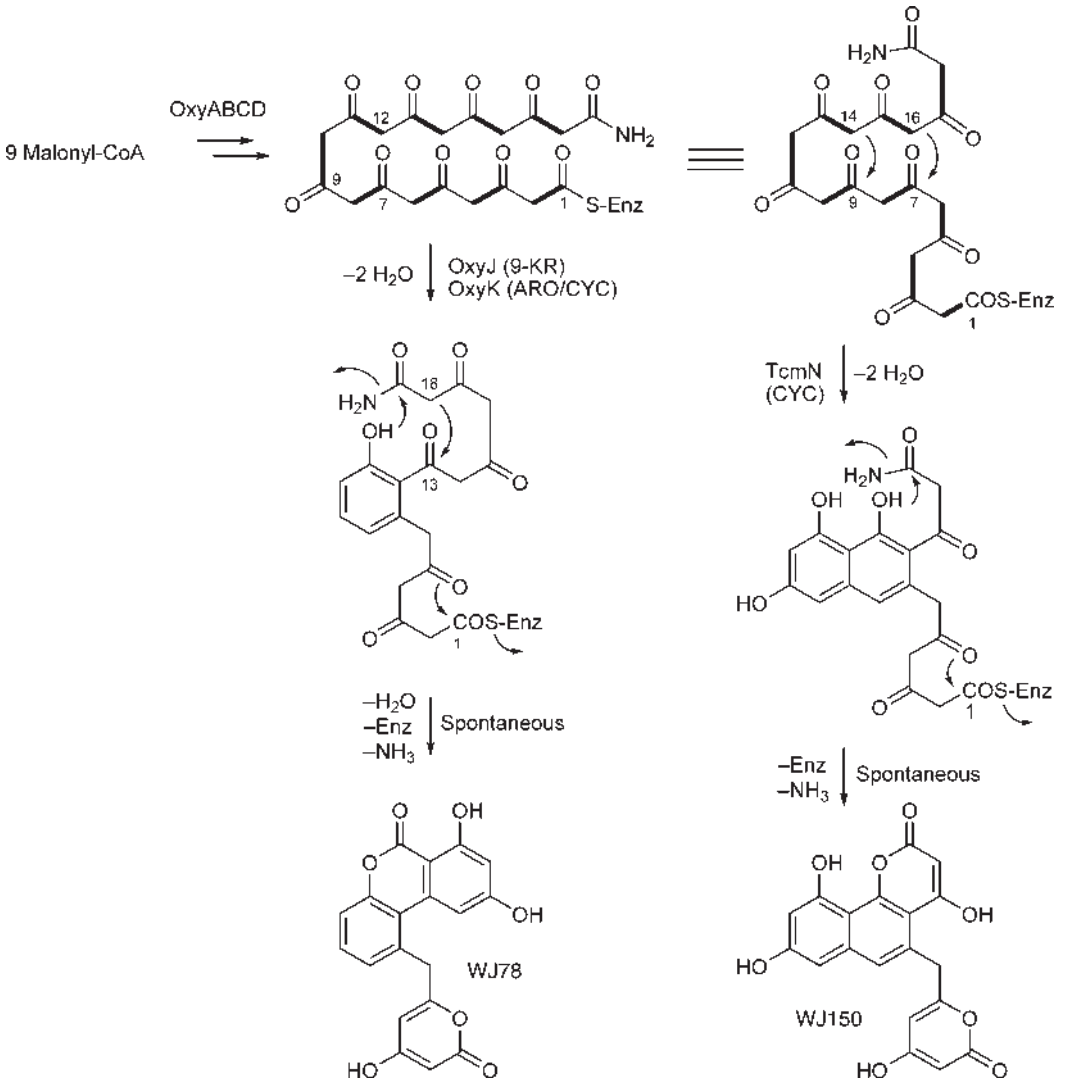
Not only did this work yield the first artificial N-containing unnatural natural product, it also allowed new conclusions regarding type II PKSs in general, that the 9-KR (here OxyJ or ActIII) reduces carbonyl 9 independent of the C-7/C-12 cyclizations observed in both the *oxy* and the *act* pathways. Attempts to heterologously pair OxyD with minimal PKSs of the *act* and *tcm* pathways failed to yield other N-containing polyketides, possibly because their acetate-primed KS-CLFs do not tolerate or cannot interact with the polar starter malonamyl ACP. In the absence of 9-KR OxyJ, spontaneous C-11/C-16 cyclization occurs initially, and benzo[*g*]isoquinoline-pyranone WJ85 is formed through various spontaneous cyclizations and release from the ACP. The formation of WJ85 is again most likely terminated by a nucleophilic attack of carbonyl C-7 by the amide. In addition, a quinone WJ85b was found, presumably generated through spontaneous oxidation (**Scheme 41**).[263]

Interestingly, the addition of cyclases, which lead to the formation of phenol rings, changes the situation drastically. The resulting phenolic OH groups attack the amide carbonyl, and benzochromenones WJ150 (involving cyclase TcmN) and WJ78 (involving 9-KR OxyJ and cyclase OxyK) are formed with the elimination of NH_3, instead of the originally desired N-heterocyclic compounds (**Scheme 42**).

Thermodynamic calculations and model reactions showed that the formation of the aromatic pyrone ring is the driving force for this unexpected outcome. Thus, a favorably positioned primary amide carbonyl can serve as an electrophilic center for the design of pyrone ring-containing unnatural natural products through combinatorial biosynthesis, a strategy nature might also have used to biosynthesize rubromycin and related pyrone-containing compounds, for example.[488]



Scheme 41 Spontaneous formation of benzo[g]isoquinoline WJ85 by the minimal oxy-encoded PKS and amidotransferase OxyD, and further oxidation to quinone WJ85b.

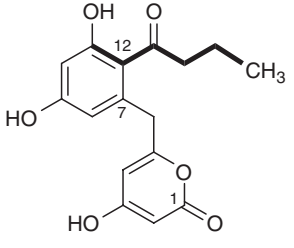


Scheme 42 Left branch: Generation of benzochromenone WJ78 by the combination of minimal oxy-encoded PKS (OxyABC), amidotransferase OxyD, and aromatase/cyclases OxyK (CYC). Tight branch: Recombination of the minimal oxy-encoded PKS, amidotransferase OxyD, and cyclase TcmN from the tetracenomycin C pathway, which catalyzes the initial 9/14-cyclization and a second cyclization, leads to benzochromenone WJ150.

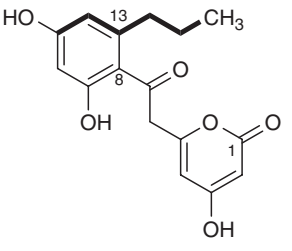
1.07.4.1.2(iii) New octaketides by rational mutation of the* act *PKS Cox, Simpson, and coworkers redesigned the minimal actinorhodin (*act*) PKS to accommodate longer fatty acids as starter units.[86] Two point mutations manipulating the measuring system of the *act* PKS, one in the CLF and another in the ACP, allowed for novel C_4–C_8 to be incorporated to yield new C_{16} aromatic polyketides with long exocyclic, mostly saturated hydrocarbon residues shown in **Scheme 43**. Some of these novel C_{16} metabolites undergo an unprecedented folding initiated by a 3,7-, 6,11-, or 8,13 first cyclization (**Scheme 43**).

1.07.4.1.2(iv) New dodecaketides by manipulation of various PKS-associated enzymes involved in benastatin biosynthesis The benastatins, pradimicins, fredericamycins, and members of the griseorhodin/rubromycin family represent a structurally and functionally diverse group of long-chain polyphenols from actinomycetes (see Section 1.07.3.4). Comparison of their biosynthetic gene clusters (ben, prm, fdm, grh, and rub) revealed that all loci harbor genes coding for a similar, yet uncharacterized, type of PKS-associated

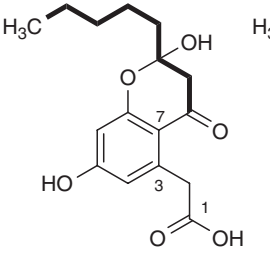
Initial 7,12-cyclization:

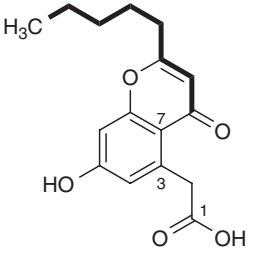


Initial 8,13-cyclization:

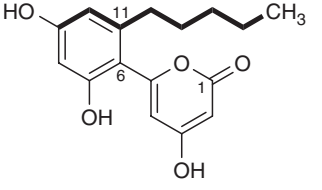


Initial 3,7-cyclization:





Initial 6,11-cyclization:



Scheme 43 Artificial aromatic polyketides through targeted manipulation of the priming process of the act octaketide PKS. 'Starter units' generated this way (here C6 or C8), partly integrated in the molecules' rings, are highlighted in bold.

ketoreductases. In a phylogenetic survey of representative KRs involved in type II PKS systems, Hertweck, Piel, and coworkers found that it is generally possible to deduce the regiospecificity of the PKS-associated KR from the amino acid sequence and thus to predict the nature (shape) of the aromatic polyketide (e.g., angucycline, anthracycline, and BIQs). A new clade of KRs is characteristic for biosynthesis of pentangular polyphenols, exemplified by BenL of the benastatin pathway, which functions as a C-19 KR and is responsible for the characteristic ring D bent found in these compounds (also see above). Inactivation of gene *benL* resulted in the formation of a series of 19-hydroxy benastatin and bequinostatin derivatives (e.g., benastatins G, K, and L; **Scheme 44**; see also **Scheme 33**). Inactivation of other early *ben* PKS-associated enzymes (e.g., KSIII-related enzyme BenQ responsible for the attachment of the hexanoate starter unit or *gem*-di-*C*-MT BenF) led to the production of various new benastatin derivatives (see **Scheme 44** for BenQ mutant productsand **Scheme 32** for BenF mutant products).[100,519,520]

1.07.4.1.3 Outlook

The few examples above illustrate that priming can be used advantageously to generate new metabolites, even for aromatic polyketides, which were once believed to allow only acetate (malonyl-CoA) priming. The usage of an iteratively acting bacterial type I polyketide synthase for the generation of a starter unit, which in turn is used to prime a type II PKS, was proposed for the hedamycin/pluramycin biosynthesis (see above).[34,521] Furthermore, it was shown that nonreducing aromatic fungal multidomain iterative PKSs, which consist of nondissociated iteratively acting megasynthases containing enzyme activities (similar to those described in this chapter on bacterial type II PKSs) to generate aromatic polyketides, such as the aflatoxins, SMA76A, or bikaverin, are quite comparable in both their mechanism of action and products to bacterial aromatic polyketides discussed here.[75,506,508,522,523] New artificial aromatic polyketides, such as SMA93, along with some previously generated through manipulations of bacterial type II PKSs, such as SEK34, SEK26, and DMAC (see also Section 1.07.4.1.1), could be generated either by using a truncated fungal megasynthase (*Gibberella fujikuroi* PKS4 without the directing thioesterase Claisen cyclase domain, TE/CLC) or combining this fungal megasynthase PKS4 with bacterial type II PKS modifying enzymes, for example, *Act*KR, *Gris*CYC, and/or *Oxy*CYC2.[524]

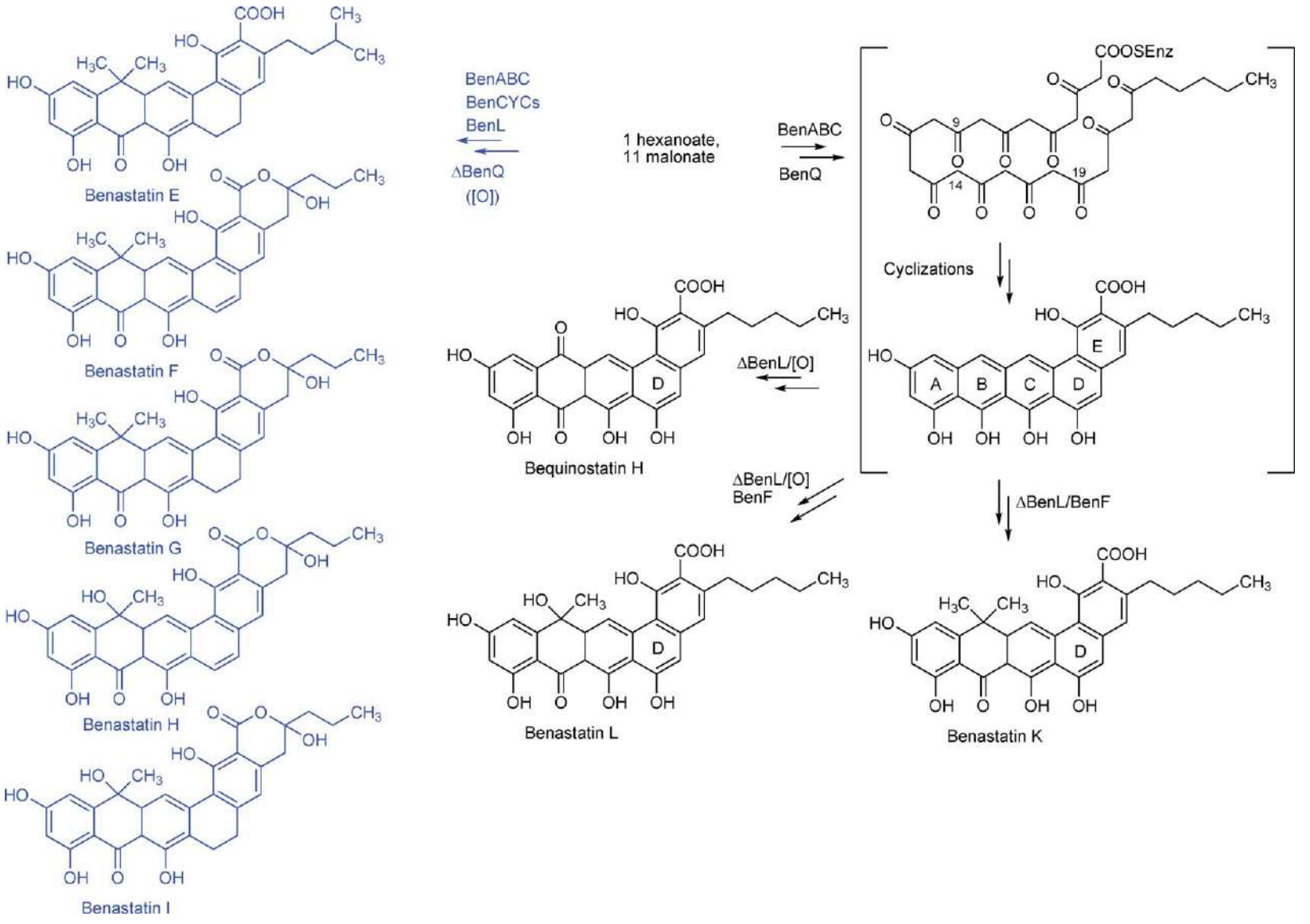
Further exploitation of this combination of fungal and bacterial aromatic PKSs allowed for the first time to generate such artificial aromatic polyketides in *Escherichia coli*.[525] Although these very recent developments could just be viewed as a further argument to acknowledge sheer unlimited natural and artificial diversity among polyketides or to abandon any classification system for PKSs,[526] they also show future perspectives for the engineering of artificial, aromatic polyketides. Thus, along with rational mutation of PKS units (as described in Section 1.07.4.1.2), both the usage of type I PKS fragments for the generation of unusual starter units for priming and the utilization of fungal PKSs are likely to be exploited in future to generate even greater structural diversity among aromatic polyketides.

1.07.4.2 Derivatives of Natural Products Generated by Recombination with Genes Encoding Post-PKS Tailoring Enzymes

With the discovery and invention of suitable molecular biological tools, many new natural products have been generated in the past decade through exploiting post-PKS tailoring enzymes. Although the structural changes obtained through usage of post-PKS tailoring enzymes usually do not deviate much from the parent natural products, that is, novel structural frameworks usually cannot be obtained, the likelihood of generating biologically active compounds is significantly increased. In fact, the exploitation of post-PKS tailoring enzymes has already led to the discovery of a few derivatives/hybrid natural products that turned out to be biologically more active, at least in some respects, than the parent natural product. This subsection mainly focuses on hybrid natural products generated after 2001/2002, when this topic was thoroughly reviewed.[106]

1.07.4.2.1 Exploiting deoxysugar pathways and glycosyltransferases

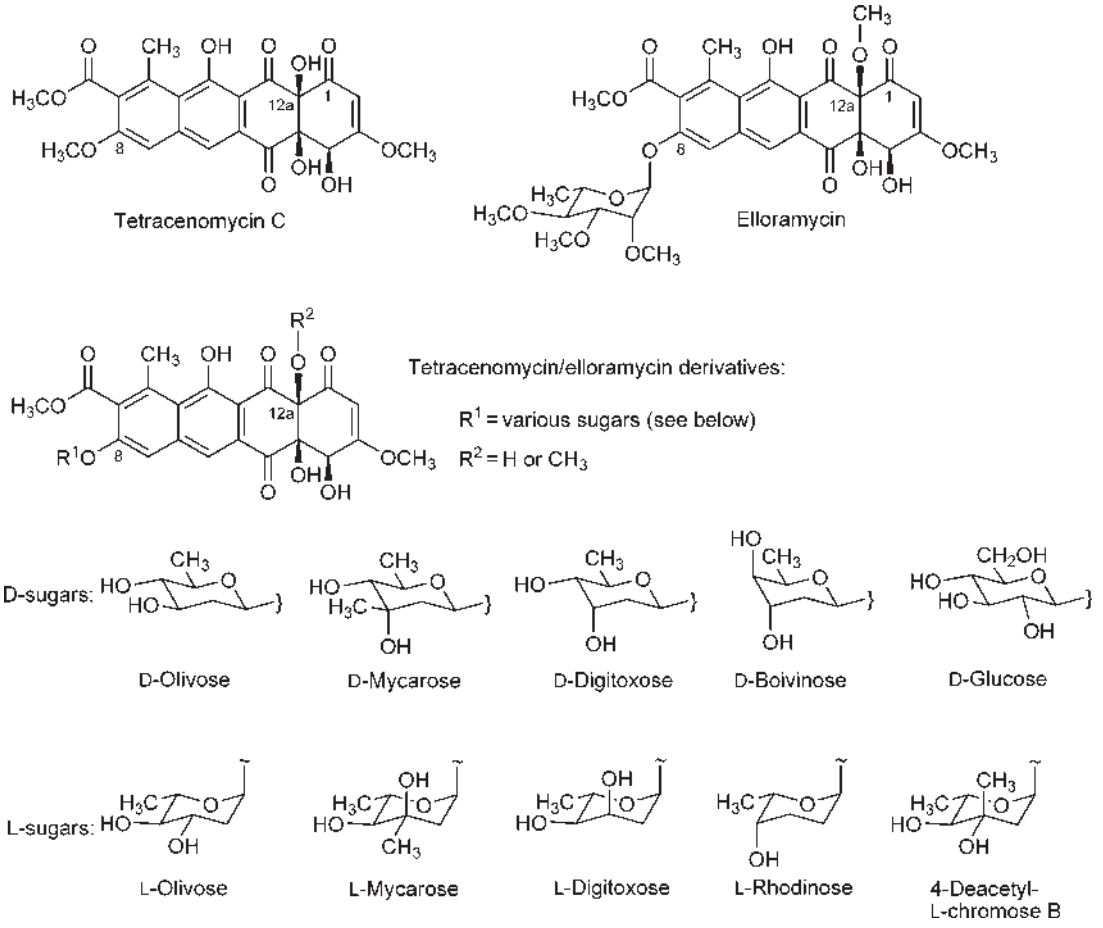
Many natural products with aromatic polyketide cores contain sugar residues, predominantly deoxysugar residues or deoxyaminosugars. These sugars play a major role in decorating and thus defining the shape of such type II PKS-derived natural products, and are in most cases essential for their biological activity. This is certainly also true for all other glycosylated natural products. Thus, it is not surprising that a lot of research in the past decade focused



Scheme 44 New benastatin analogues through inactivations of KASIII homologue BenQ (blue) or 19-KR BenL (black). Note that di-*C*-methyltransferase BenF occasionally competes with presumably nonenzymatic anthrone oxidation.

on deoxysugar pathways and enzymes involved in deoxysugar pathways and glycosyltransfer. Various, often very unusual, deoxysugar pathways were explored in detail, including their total enzymatic synthesis, and many excellent reviews including the most recent ones are available.[108,110,113,338,492,493,527–530] However, even though the first hybrid natural product involving a foreign sugar moiety had already been generated in 1995,[460] only a few deoxysugar pathways/pathway enzymes and GTs were exploited by combinatorial biosynthesis for the systematic generation of new hybrid natural products. Nevertheless, various derivatives of natural products assembled by type II PKSs were generated through manipulation of deoxysugar pathways, and some GTs turned out to be of significantly lesser substrate specificity than expected.

The first example of a hybrid natural product in which the deoxysugar moiety was modified was 8-demethyl-8-D-olivosyl-tetracenomycin C (Scheme 1.07.4.2.1.1, R_1 = D-olivose, R_2 = H).[460] This was the result of a combination of genes from the elloramycin (aglycone biosynthesis genes and glycosyltransferase gene *elmGT*) and urdamycin (D-olivose biosynthesis genes) pathways. This first success triggered the exploration of the crucial glycosyltransferase ElmGT, which turned out to be extremely flexible regarding its sugar donor substrate, but quite restricted toward its acceptor substrate. Besides its natural sugar L-rhamnose, ElmGT can also transfer various D- and L-sugars, including the highly deoxygenated sugar L-rhodinose and L-amicetose as well as a disaccharide composed of two D-olivoses. However, attempts to transfer aminosugars or 4-ketosugars failed. All of the accepted sugars were placed by ElmGT onto the 8-OH group of 8-demethyltetracenomycin C (without 12a-OCH_3 group) or elloramycinone (with 12a-OCH_3 group), and, as a result of these studies, various tetracenomycin/elloramycin derivatives were generated (**Scheme 45**).[111,334,445,461–463] Although the beginning of these studies had been marked



Scheme 45 Tetracenomycin C, elloramycin, and examples of hybrid derivatives through exploiting the sugar donor substrate flexible glycosyltransferase ElmGT and heterologously expressed enzymes of foreign deoxysugar biosynthesis.

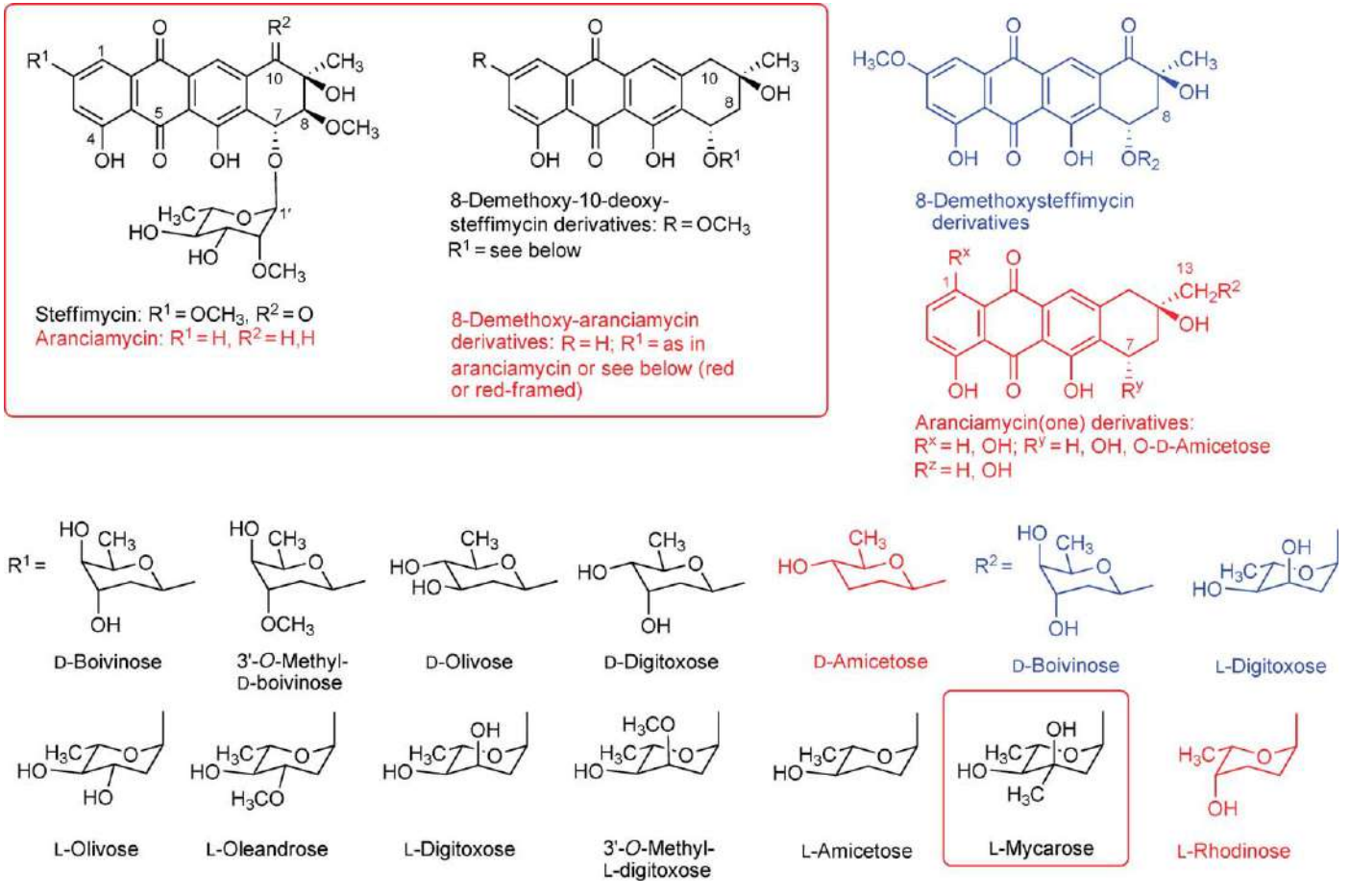
by recombination/heterologous expression of larger fragments of natural product biosynthesis pathways, more predictable attempts to generate tetracenomycin/elloramycin derivatives later employed the so-called deoxysugar plasmids, originally invented by Salas *et al.*, in which entire pathways toward desired deoxysugars from different biosynthetic pathways were recombined.[462,492,528,531]

Deoxysugars also play an important role in the DNA-binding properties of antitumor anthracycline antibiotics, and preliminary attempts to generate different deoxyaminosugars and attach them to anthracycline-type aglycones in the quest for new antitumor drugs were explored by Hutchinson *et al.*[532] Production of TDP-L-daunosamine, the aminodeoxysugar present in daunorubicin and doxorubicin, and its attachment to ε-rhodomycinone to generate rhodomycin D was achieved by bioconversion with a *S. lividans* strain bearing two plasmids, one containing genes encoding the aglycone moiety and glycosyltransfer, the other deoxysugar biosynthesis genes and resistance genes. Although the involved GT seemed to be restricted to its natural sugar donor substrate, some flexibility regarding the acceptor anthracyclinone was found. Similarly, the expression of the steffimycin gene cluster in *S. albus* in combination with various plasmids directing the biosynthesis of different neutral and branched-chain deoxyhexoses led to the formation of 12 new steffimycin derivatives (**Scheme 46**).[216] As in the case of the L-rhamnosyltransferase ElmGT (see above), these experiments demonstrated the flexibility of the involved L-rhamnosyltransferase StfG toward a broad variety of NDP-sugar donor substrates (D- and L-6-deoxyhexoses, a different degree of further deoxygenation, and branched or unbranched sugars) but also allowed for small variations regarding the type II PKS-derived acceptor substrates, that is, compounds with or without 10-keto group. The results also shed further light on the steffimycin biosynthetic pathway (glycosylation occurs prior to 8-oxygenation and 8-O-methylation). Most significantly, the analysis of the cytotoxic activities of these compounds revealed that two of the new derivatives – 3′-*O*-methylsteffimycin (not shown) and D-digitoxosyl-8-demethoxy-10-deoxysteffimycinone – possess improved antitumor activities compared with the parent drug steffimycin and allowed for some insights into the structure–activity relationships (SARs) of this class of anthracycline antibiotics. Following up on this, structurally similar analogues of anthracycline aranciamycin, which is closely related to steffimycin, were also generated, taking advantage of the flexible GT (AraGT) of the pathway (**Scheme 46**).[533,534] However, here the generation of new aranciamycin analogues was achieved by heterologous transfer of the aranciamycin gene cluster into heterologous hosts harboring natural deoxysugar pathways other than (methylated) L-rhamnose, a strategy that follows the approach used to generate the elloramycin analogues described above.

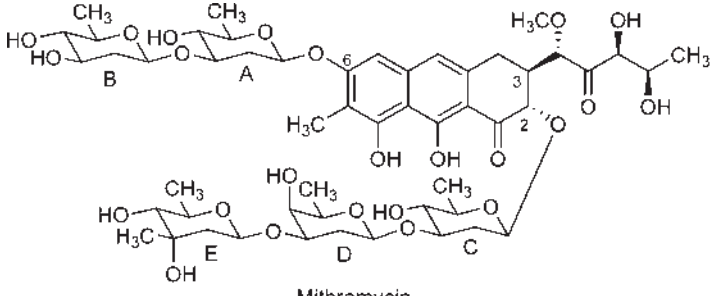
Summing up the successes with ElmGT, StfG, and AraGT, which are all L-rhamnosyl (L-Rha) transferases, it appears that L-Rha-transferases are particularly flexible for their NDP-sugar donor substrates, and pathways involving such GTs might therefore be especially suited for the generation of new analogues with modified saccharide moieties. No crystal structures have been reported so far, but one can speculate that the sugar donor substrate might be bound mainly through interactions with its activating NDP (TDP) residue.

Deoxysugar plasmids were also exploited to generate derivatives of the clinically relevant aureolic acid anticancer drug mithramycin. In contrast to the abovementioned studies, the natural producer *S. argillaceus* or its mutant *S. argillaceus* M7D1 was used as a host. This was necessary due to the significantly larger gene cluster of the *mtm* pathway. A major difference to all the studies mentioned above was the fact that here four GTs (responsible for the five glycosyltransfer steps in mithramycin biosynthesis) were challenged simultaneously in a truly combinatorial biosynthetic approach. Overall, 11 new mithramycin (MTM) derivatives (**Scheme 47**) were generated. It turned out that particularly sugar units A, D, and E could be modified, although units B (often missing, but never changed) and C (never replaced) remained unaffected. Three of the new analogues, demycarosyl-3D-β-D-digitoxosyl-MTM, deoliosyl-3C-β-D-mycarosyl-MTM, and 3A-deolivosyl-MTM, showed improved and/or significantly broadened anti–breast cancer activity compared to the parent drug mithramycin. The studies also revealed further SARs of aureolic acid antitumor drugs.[535,536]

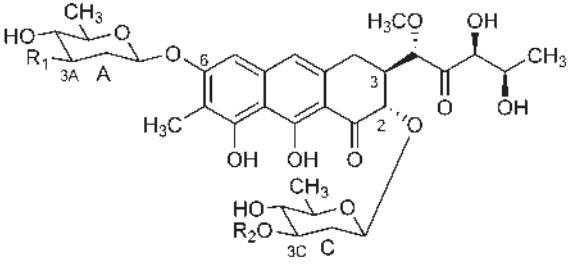
The knowledge gained about the assembly of the saccharide chains of urdamycins and landomycins was exploited to design many new analogues through pathway engineering, taking advantage of the similarities of the deoxysugar chains of urdamycins and landomycins and the GTs involved. Both the natural flexibility of olivosyltransferases LanGT1 and LanGT4 and the unique roles of LanGT2 and LanGT3 were exploited to generate naturally unknown landomycins or urdamycins – for example, the tetrasaccharidal compounds landomycin J through overexpression of LanGT3[332] and urdamycin U through heterologous expression of the rhodinosyltransferase LanGT4 in *Streptomyces fradiae* AX.[319] Various genetic experiments along these lines



Scheme 46 Steffimycin derivatives through combinatorial biosynthesis exploiting various deoxysugar plasmids and the TDP-sugar donor substrate flexible GT StfG (black or blue structures). Aranciamycin analogues (red or red framed structures) were generated by heterologous expression of the ara gene cluster in various hosts with established NDP-deoxysugar pathways.

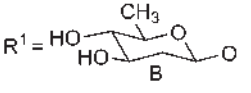


Mithramycin

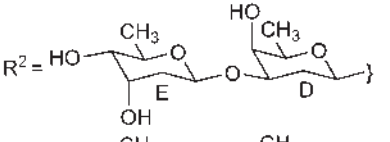


Mithramycin Analogues

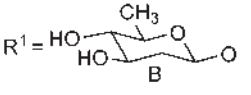
Demycarosyl-3D-β-D-digitoxosyl-MTM: R^1 =



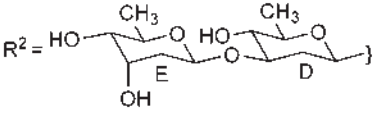
R^2 =



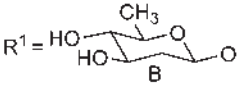
R^1 =



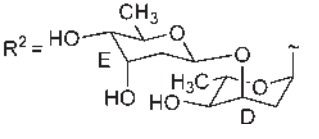
R^2 =



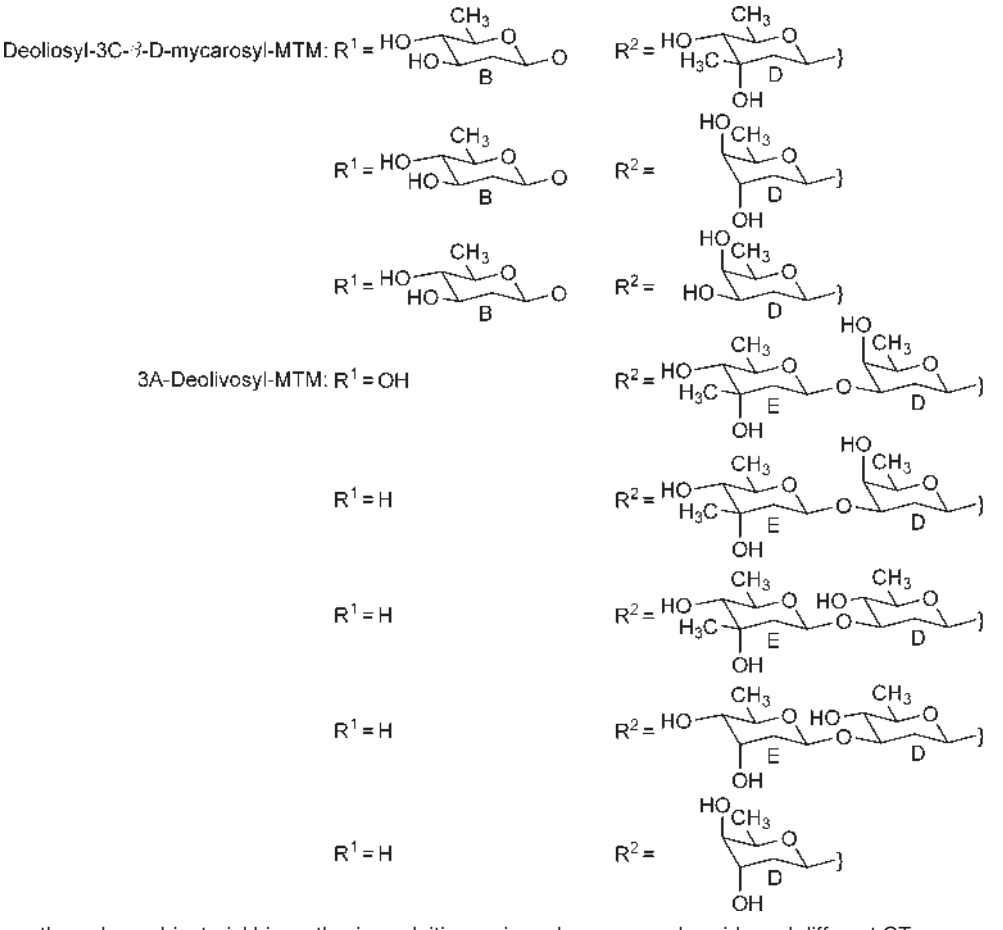
R^1 =



R^2 =



Scheme 47 (Continued)



Scheme 47 Mithramycin derivatives through combinatorial biosynthesis exploiting various deoxysugar plasmids and different GTs.

by the Bechthold group,[342] for example, inactivation or overexpression of *urd* GTs, sometimes combined with heterologous expression of *lan* GTs, led to many new urdamycins or ladamycins (hybrids containing the urdamycin aglycone, but sections of the landomycin sugar chain), so that in addition to the eight natural urdamycins A–H, almost the entire remainder of alphabet is meanwhile covered for new urdamycin A or B analogues. A gene technological highlight is certainly urdamycin P with a novel branched saccharide pattern, designed by systematic exchange of gene segments encoding the two very similar GTs UrdGT1b and UrdGT1c, leading to the new hybrid GT1707 with altered acceptor substrate specificity.[537] The flexibility of *C*-GT UrdGT2 toward both its NDP-sugar donor and acceptor substrates was exploited to generate many new urdamycin, landomycin, premithramycin, and prejadomycin derivatives. In most of these attempts, other GTs were either coexpressed, for example, LanGT1 for one of the C-glycosylated premithramycin-type molecules, or were available in the host strain to further complex the saccharide pattern.[105,113,314,333,340] For selected examples, see **Scheme 48**.

An important observation was made by the Thorson group in 2006, who discovered that GT reactions, usually perceived as unidirectional, can also be reversed in some cases.[538] Examples from the calicheamicin and vancomycin pathway were reported, and none of the 'reversible' GTs were involved in or used for decorating type II PKS-derived metabolites. However, exploiting the reversibility of natural product GT-catalyzed reactions bears enormous potential and may be used in the future to generate new analogues of type II PKS-derived natural products. Glycosyltransfer reversibility could be used for (1) generating otherwise hard to obtain NDP sugars, (2) sugar exchange at the same scaffold (aglycone), or (3) transfer from one scaffold to another (**Scheme 49**). Although this method may be generally useful in obtaining rare NDP sugars, sugar exchange and aglycone exchange depend on the NDP-donor and (aglycone) acceptor flexibility of the involved GTs, respectively. The above examples show that the former is more frequently found than the latter.

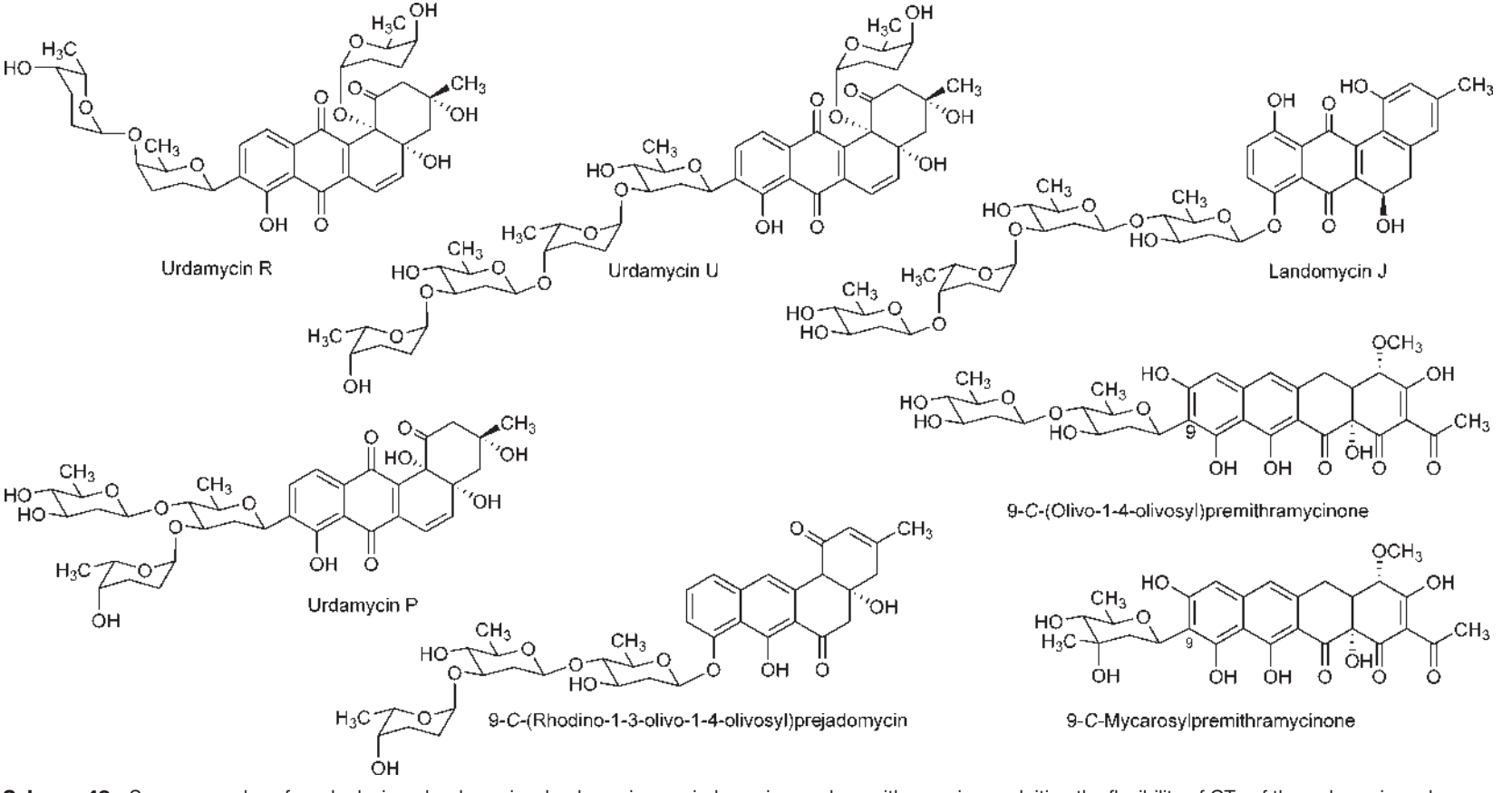
To further expand the method repertoire for exchange or modification of sugar residues in natural products, Thorson *et al.* also designed glycorandomization, later expanded to neoglycorandomization.[122,491,539–541] This powerful combinatorial approach is applicable to all kinds of natural products but has not yet been exploited to modify type II PKS-derived aromatic polyketides.

1.07.4.2.2 Exploiting group transferases other than glycosyltransferases

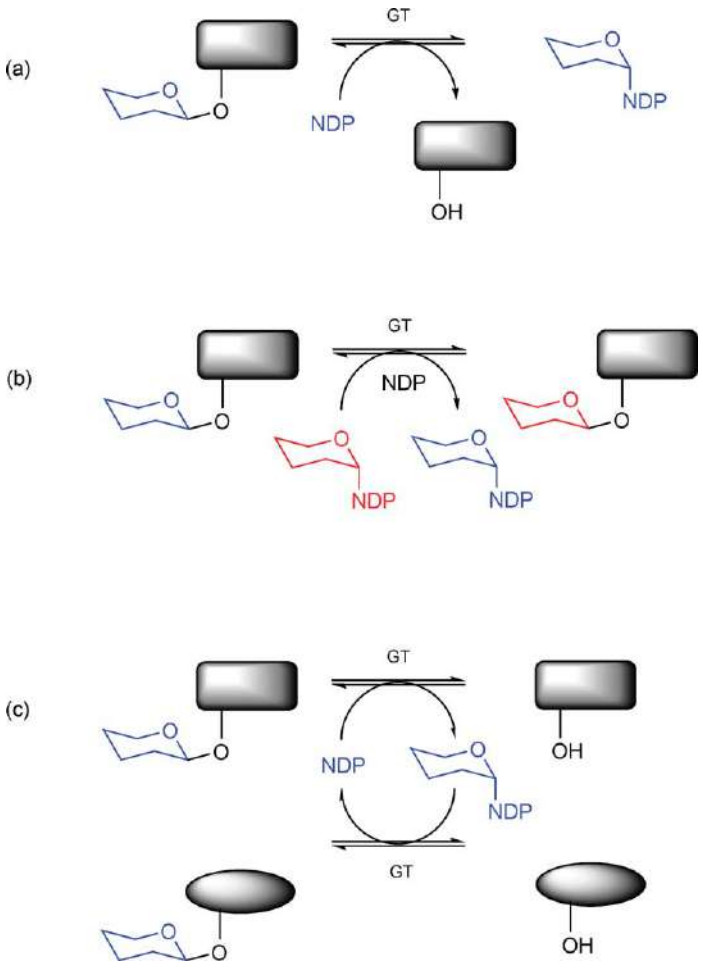
Besides GTs, only a few other types of group transferases were investigated, but not yet exploited, for combinatorial biosynthetic approaches on type II PKS-derived compounds. This includes MTs, acyltransferases, and aminotransferases. In particular, SAM-dependent MTs can be used to generate more or other O-methyl decorated analogues. A good example is the flexible 3′-*O*-methyltransferase OleY, which was used for the methylation of the deoxyhexose attached to the steffimycin aglycone through expression of the corresponding *oleY* gene in *Streptomyces steffisburgensis*, leading to the isolation of 3′-*O*-methylsteffimycin.[216]

1.07.4.2.3 Exploiting oxidoreductases

The term 'oxidoreductase' used here includes app post-PKS tailoring enzymes involved in either oxidation or reduction processes, such as oxygenases, ketoreductases, and dehydrogenases. Oxidoreductases were used early on in combinatorial biosynthetic studies. In fact, the first hybrid natural product ever generated, mederrhodin, was obtained through an oxygenase activity.[155] However, this historical example of recombining a tailoring oxygenase from one pathway with another natural product was not followed up, likely because of substrate specificity restrictions of oxygenases that act relatively late in biosynthetic pathways, that is, on pretty unique substrates. For example, recombining the mithramycin biosynthetic pathway with *tcmG*, a gene encoding a triply hydroxylating monooxygenase–dioxygenase from the tetracenomycin pathway[458,542] led to the accumulation of premithramycinone H, a shunt product of the mithramycin pathway. Formation of premithramycinone H could be explained by interference of the heterologously expressed foreign oxygenases TcmG and TcmH from the tetracenomycin C biosynthetic pathway with the mithramycin pathway, inhibiting its normal oxidative processing toward premithramycinone and on to mithramycin (see Section 1.07.3) and forcing a derailment to shunt product premithramycinone H[290] (**Scheme 50**). However, the desired triple hydroxylation event toward a mithramycin derivative with a different hydroxylation pattern did not occur, possibly because of the too-strict substrate specificity of TcmG.



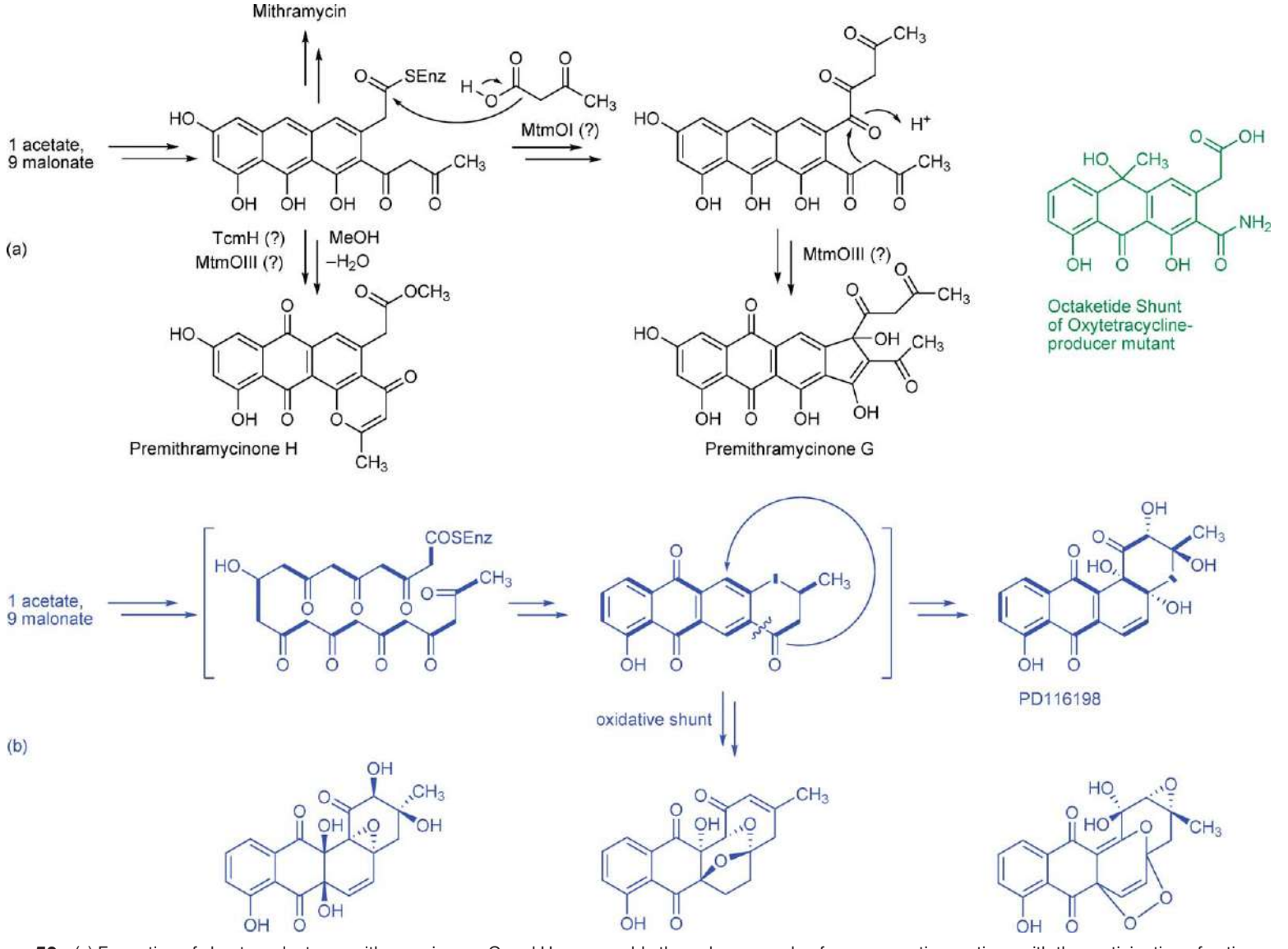
Scheme 48 Some examples of newly designed urdamycins, landomycins, prejadomycins, and premithramycins exploiting the flexibility of GTs of the urdamycin and landomycin pathways, particularly of C-GT UrdGT2.



Scheme 49 Potential for exploiting GT reversibility: (a) generation of rare NDP-activated sugars; (b) sugar moiety exchange; (c) aglycone exchange.

Various new products were found by removing (either through gene deletion or gene inactivation) of oxygenases from biosynthetic pathways, including important biosynthetic intermediates mentioned above, such as premithramycin B, a key intermediate of the mithramycin biosynthetic pathway, or prejadomycin (2,3-dehydro-UWM6), an intermediate of many angucycline pathways. In addition to intermediates, unique shunt products were produced this way, which arose from nonenzymatic cyclization reactions of reactive early products of aromatic polyketide pathways. For instance, an unexpected octaketide-derived shunt (**Scheme 50**, green) product was accumulated upon inactivation of oxygenase OtcC (=OxyS; see Section 1.07.3) of the pathway normally leading to nonaketide oxytetracyline. Apparently, the control of the chain length was lost once the post-PKS tailoring oxygenase gene was removed, which might be an essential partner of the PKS complex.[261] Similarly, premithramycinone G, derived from a polyketide with seemingly abnormal chain length accumulated upon inactivation of early acting oxygenase MtmOII. Premithramycinone G, a dodecaketide, is a metabolite with greater chain length than the normal pathway end product, the decaketide mithramycin. However, here incorporation experiments with ^{13}C-labeled precursors indicated that the chain length extension is likely due to a nonenzymatic reaction of a reactive ACP-tethered intermediate with acetone or acetoacetate (**Scheme 50a**).

Gould *et al.* found various highly oxygenated angucyclinone derivatives as minor compounds in the PD116198 producer *S. phaeochromogenes*,[543] which may be viewed as deranged shunt pathways caused by hyperactive oxygenases. The structure of the major metabolite PD116198 is an ordinary angucyclinone;

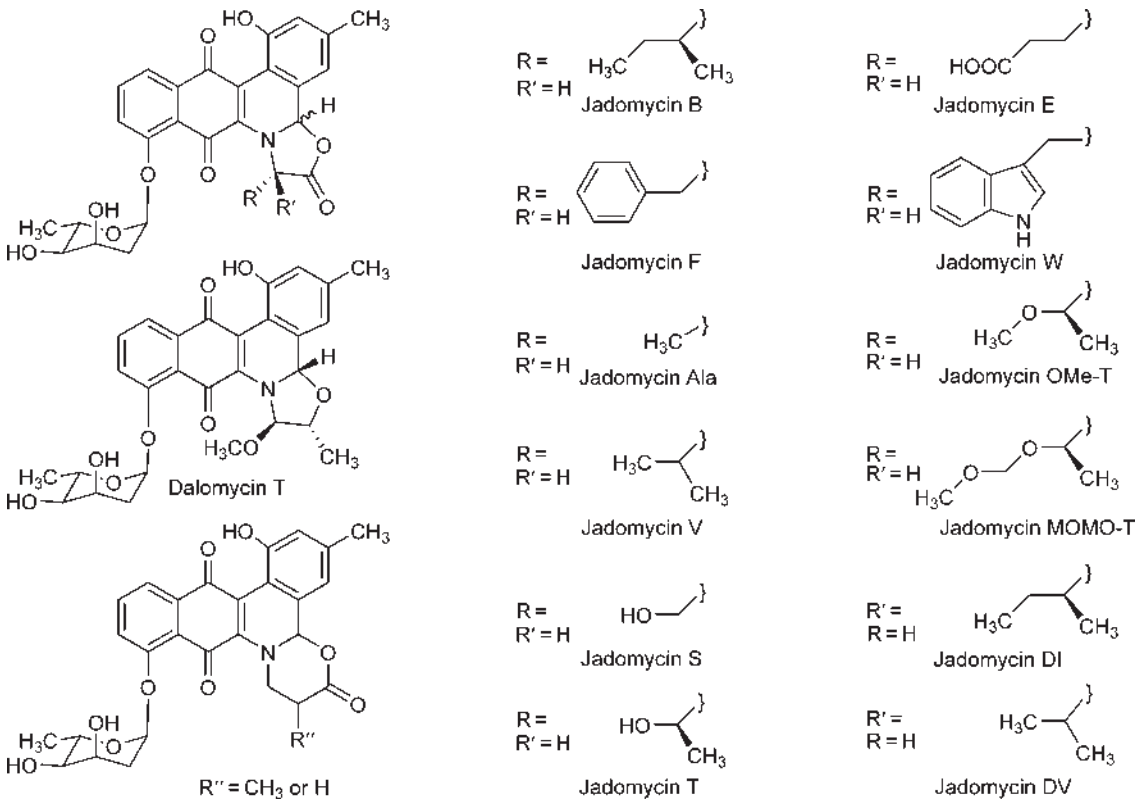


Scheme 50 (a) Formation of shunt products premithramycinones G and H, presumably through a cascade of nonenzymatic reactions with the participation of native or foreign oxygenases. Green indicates an octaketide shunt product isolated from the oxytetracycline producer after inactivation of an oxygenase. (b) Pathway to angucyclinone PD116198, possibly derived by oxidative rearrangement from an anthracyclinone intermediate, and minor congeners from oxidative shunt pathways (blue).

however, the acetate incorporation pattern suggested that it was derived through a novel oxidative rearrangement of a linear tetracyclic anthracyclinone-type precursor (**Scheme 50b**).

As discussed above, a oxidoreductase-catalyzed reaction cascade in the jadomycin biosynthesis leads to the reactive biosynthetic intermediate, aldehyde Y (see Section 1.07.3.2.3). The following nonenzymatic reaction (normally occurring with the amino acid isoleucine) could be exploited to generate a library of jadomycin analogues by replacing isoleucine with other amino acids in the growth medium, including some artificial amino acids. Essentially, all 20 naturally occurring L-amino acids, two β-amino acids (leading to a new class of jadomycin analogues with oxazinanone ring), some D-amino acids and nonnatural amino acid analogues, and even one β-amino alcohol (leading to dalomycin T) could be incorporated. Representative examples are shown in **Scheme 51**.[410–412,414] Only a few of the new analogues were assayed as anticancer drugs, and some showed slightly improved cytotoxic activity against certain human cancer cell lines compared with the original jadomycin B;[544] however, SARs could not be deduced because of the lack of knowledge about the targets. Many analogues were obtained only in miniscule amounts; for better production, a recently reengineered strain could be useful, in which the regulatory region was manipulated with a dramatic increase of jadomycin B yields.[545]

Some of the abovementioned aranciamycin analogues (see **Scheme 46**) were obviously targeted by host strain oxygenases that introduced oxygen residues into 1- and 13-positions; however, it is unclear which enzymes are responsible. Interestingly, the aranciamycin(one) derivatives with 1-OH residue showed better anticancer activity than those without.[534] Targeted oxidoreductase inactivations also led to new landomycins, with depleted oxygen patterns at the aglycone moiety, for example, landomycins F, M, and O[324]. The gene inactivation experiments affecting angucycline oxidoreductases yielded many new metabolites (see pathways



Scheme 51 Examples of jadomycin analogues generated by exploiting the nonenzymatic imine and consequent oxazolone/oxazolidine/oxazinanone ring formation, which follow an oxidative C–C bond cleavage cascade (see Section 1.07.4.2.3). Not shown are further analogues, with other side chains of the 20 naturally occurring L-amino acids and with side chains of other rare or unnatural amino acids including 4-F-Phe, phenylglycine, or 2-aminoisobutyric acid.

of jadomycins, landomycins, and oviedomycin mentioned above). Likewise, the pathways leading to gaudimycins A and B (see above) could also be viewed as gene recombination experiments involving (cryptic) oxidoreductases.

Similarly, ketoreductase functionalities were targeted when bifunctional oxygenase/reductase enzymes of angucycline pathways were inactivated, and genes coding for post-PKS tailoring enzymes, particularly those involved in sugar residue biosyntheses, were targeted early for the modification or optimization of natural products. The most often cited classical example is Hutchinson's elegant genetic engineering approach to produce epirubicin, a clinically used anthracycline anticancer agent that was formerly accessible only through a multistep semisynthetic approach. By replacing the *dmnV* gene of the doxorubicin pathway that encodes the 4-ketoreductase involved in the sugar moiety with *avrR* or *eryBIV*, the stereochemistry in 4-position of the sugar moiety was changed, yielding the desired epirubicin, which is now easily producible by the engineered mutant strain.[203]

1.07.5 Conclusions

Even though nature seems unbeatable regarding structural diversity, the use of combinatorial biosynthetic methods adds a large structural variety, both regarding new scaffolds and derivatives of existing natural themes. Yet, the generation of structural diversity in a true combinatorial fashion with respect to type II PKS-based compounds is still in the beginning, and more progress will be made when enzymes with natural broad substrate specificity are involved or when nonenzymatic reactions can be exploited. Because enzymes responsible for starter unit attachments and sugar pathway enzymes/GTs so far seemed to offer the broadest substrate specificities, true combinatorial biosynthetic derivatization appears most progressed once starter unit variation or variation regarding saccharide residues is envisaged. In contrast, oxygenases are too specific, too reactive, or operate in multienzyme clusters, which we are just beginning to understand. More and more emerging crystal structures, particularly of oxygenases of the anthracycline and angucycline pathways, may help to gain a better understanding, but a taming of oxygenases for a more systematic exploitation in a combinatorial biosynthetic fashion seems still far away. Similarly, the exploitation of methyl- and other group transferases as well as halogenases is still in the very beginning for the same reasons (specificity issues).

Acknowledgments

Work of the authors reported here was supported by the US National Institute of Health (grants CA 091901 and CA 102102) to J.R. and by the German Ministry for Education and Research (BMBF GenoMik) and the German Research Community (DFG) to C.H.

Glossary

chain length factor (CLF) (=KS$_\beta$) major component of the minimal PKS.
cyclization a ring closing reaction.
decaketide compound assembled from 10 small acid building blocks.
deoxysugar a sugar molecule in which one or more hydroxyl group(s) is eliminated or replaced by another functional group (not a hydroxyl group).
dioxygenase enzyme that introduces two oxygen atoms simultaneously.
dodecaketide compound assembled from 12 small acid building blocks.
glycosyltransferase enzyme that transfers a sugar.
group transferase enzyme that transfers a smaller molecule (unit), such as a sugar, a methyl group, or an acetyl residue, from a cofactor to a larger molecule. The transferred unit becomes a structural moiety or residue.
hybrid natural product naturally nonoccurring compound generated by manipulation of a biosynthetic machinery, usually achieved by addition of enzymatic activities to an existing machinery via gene recombination.

methyltransferase enzyme that transfers a methyl group.
minimal PKS a minimal set of iteratively operating proteins; the smallest possible complex of PKS enzymes (usually KS, CLF, and ACP) still functioning in assembling a molecule, usually a shunt product derived from a pathway intermediate.
monooxygenase enzyme that introduces one oxygen atom.
octaketide compound assembled from eight small acid building blocks.
oxidoreductase enzyme that performs oxygenation (oxygenase), dehydrogenation (dehydrogenase), or reduction (=hydrogenation).
oxygenases oxygen atom-introducing enzymes.
pentadecaketide compound assembled from 15 small acid building blocks.
polyketide synthase (PKS) enzyme complex assembling a molecular framework from short acid building blocks, such as acetate, malonate, and propionate.
post-PKS enzymes catalysts of post-PKS tailoring steps.
post-PKS tailoring steps all biosynthetic steps following the cleavage from the PKS acyl-carrier protein.
priming initiation of the PKS reaction sequence.
tetradecaketide compound assembled from 14 small acid building blocks.
tridecaketide compound assembled from 13 small acid building blocks.
unnatural natural product naturally nonoccurring compound generated by manipulation of a biosynthetic machinery, usually achieved by reducing the number of or modification of biosynthetic enzymes via gene inactivation or gene modification.

References

1. C. Hertweck; A. Luzhetskyy; Y. Rebets; A. Bechthold, *Nat. Prod. Rep.* **2007**, *24*, 162–190.
2. B. J. Rawlings, *Nat. Prod. Rep.* **1999**, *16*, 425–484.
3. J. Staunton; K. J. Weissman, *Nat. Prod. Rep.* **2001**, *18*, 380–416.
4. J. L. Caballero; E. Martinez; F. Malpartida; D. A. Hopwood, *Mol. Gen. Genet.* **1991**, *230*, 401–412.
5. M. A. Fernandez-Moreno; E. Martinez; L. Boto; D. A. Hopwood; F. Malpartida, *J. Biol. Chem.* **1992**, *267*, 19278–19290.
6. H. Motamedi; C. R. Hutchinson, *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4445–4449.
7. K. Ichinose; D. J. Bedford; D. Tornus; A. Bechthold; M. J. Bibb; W. P. Revill; H. G. Floss; D. A. Hopwood, *Chem. Biol.* **1998**, *5*, 647–659.
8. C. Binnie; M. Warren; M. J. Butler, *J. Bacteriol.* **1989**, *171*, 887–895.
9. W. Zhang; B. D. Ames; S. C. Tsai; Y. Tang, *Appl. Environ. Microbiol.* **2006**, *72*, 2573–2580.
10. L. Han; K. Yang; E. Ramalingam; R. H. Mosher; L. C. Vining, *Microbiology* **1994**, *140*, 3379–3389.
11. T. W. Yu; M. J. Bibb; W. P. Revill; D. A. Hopwood, *J. Bacteriol.* **1994**, *176*, 2627–2634.
12. J. Ye; M. L. Dickens; R. Plater; Y. Li; J. Lawrence; W. R. Strohl, *J. Bacteriol.* **1994**, *176*, 6270–6280.
13. A. Grimm; K. Madduri; A. Ali; C. R. Hutchinson, *Gene* **1994**, *151*, 1–10.
14. M. J. Bibb; D. H. Sherman; S. Omura; D. A. Hopwood, *Gene* **1994**, *142*, 31–39.
15. H. Decker; S. Haag, *J. Bacteriol.* **1995**, *177*, 6126–6136.
16. E. R. Rafanan, Jr.; L. Le; L. Zhao; H. Decker; B. Shen, *J. Nat. Prod.* **2001**, *64*, 444–449.
17. Y. Shen; P. Yoon; T. W. Yu; H. G. Floss; D. Hopwood; B. Moore, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3622–3627.
18. F. Lombo; G. Blanco; E. Fernandez; C. Mendez; J. A. Salas, *Gene* **1996**, *172*, 87–91.
19. T. Dairi; Y. Hamano; I. Yasuhiro; T. Furumai; T. Oki, *Biosci. Biotech. Biochem.* **1997**, *61*, 1445–1453.
20. T. Dairi; Y. Hamano; T. Furumai; T. Oki, *Appl. Environ. Microbiol.* **1999**, *65*, 2703–2709.
21. B. C. Kim; J. M. Lee; J. S. Ahn; B. S. Kim, *J. Microbiol. Biotechnol.* **2007**, *17*, 830–839.
22. S. J. Gould; S.-T. Hong; J. R. Carney, *J. Antibiot.* **1998**, *51*, 50–57.
23. L. Westrich; S. Domann; B. Faust; D. Bedford; D. A. Hopwood; A. Bechthold, *FEMS Microbiol. Lett.* **1999**, *170*, 381–387.
24. P. Brunker; K. McKinney; O. Sterner; W. Minas; J. E. Bailey, *Gene* **1999**, *227*, 125–135.
25. J. Piel; C. Hertweck; P. R. Shipley; D. M. Hunt; M. S. Newman; B. S. Moore, *Chem. Biol.* **2000**, *7*, 943–955.
26. T. Marti; Z. Hu; N. L. Pohl; A. N. Shah; C. Khosla, *J. Biol. Chem.* **2000**, *275*, 33443–33448.
27. S. Torkkell; T. Kunnari; K. Palmu; P. Mantsala; J. Hakala; K. Ylihonko, *Mol. Genet. Genom.* **2001**, *266*, 276–288.
28. R. Martin; O. Sterner; M. A. Alvarez; E. De Clercq; J. E. Bailey; W. Minas, *J. Antibiot.* **2001**, *54*, 239–249.
29. K. Räty; J. Kantola; A. Hautala; J. Hakala; K. Ylihonko; P. Mäntsälä, *Gene* **2002**, *293*, 115–122.
30. A. Li; J. Piel, *Chem. Biol.* **2002**, *9*, 1017–1026.
31. R. Novakova; J. Bistakova; D. Homerova; B. Rezuchova; J. Kormanec, *Gene* **2002**, *297*, 197–208.
32. K. Ichinose; M. Ozawa; K. Itou; K. Kunieda; Y. Ebizuka, *Microbiology* **2003**, *149*, 1633–1645.
33. C. Fischer; F. Lipata; J. Rohr, *J. Am. Chem. Soc.* **2003**, *125*, 7818–7819.
34. T. Bililign; C.-G. Hyun; J. S. Williams; A. M. Czisny; J. S. Thorson, *Chem. Biol.* **2004**, *11*, 959–969.

35. N. Menendez; M. Nur-e-Alam; A. F. Braña; J. Rohr; J. A. Salas; C. Méndez, *Chem. Biol.* **2004**, *11*, 21–32.
36. K. Jakobi; C. Hertweck, *J. Am. Chem. Soc.* **2004**, *126*, 2298–2299.
37. R. Novakova; J. Bistakova; J. Kormanec, *Arch. Microbiol.* **2004**, *182*, 388–395.
38. R. Novakova; J. Bistakova; D. Homerova; B. Rezuchova; L. Feckova; J. Kormanec, *DNA Seq.* **2004**, *15*, 188–195.
39. K. Herold; Z. Xu; F. A. Gollmick; U. Grafe; C. Hertweck, *Org. Biomol. Chem.* **2004**, *2*, 2411–2414.
40. M. Metsä-Ketelä; K. Ylihonko; P. Mäntsälä, *J. Antibiot.* **2004**, *57*, 502–510.
41. K. Palmu; K. Ishida; P. Mantsala; C. Hertweck; M. Metsa-Ketela, *ChemBioChem* **2007**, *8*, 1577–1584.
42. Z. Xu; K. Jakobi; K. Welzel; C. Hertweck, *Chem. Biol.* **2005**, *12*, 579–588.
43. F. Lombo; A. F. Brana; J. A. Salas; C. Mendez, *ChemBioChem* **2004**, *5*, 1181–1187.
44. E. Wendt-Pienkowski; Y. Huang; J. Zhang; B. Li; H. Jiang; H. Kwon; C. R. Hutchinson; B. Shen, *J. Am. Chem. Soc.* **2005**, *127*, 16442–16452.
45. B. Aigle; X. Pang; B. Decaris; P. Leblond, *J. Bacteriol.* **2005**, *187*, 2491–2500.
46. S. Gullon; C. Olano; M. S. Abdelfattah; A. F. Brana; J. Rohr; C. Mendez; J. A. Salas, *Appl. Environ. Microbiol.* **2006**, *72*, 4172–4183.
47. A. Yoo; A. V. Demirev; J. S. Lee; S. D. Kim; D. H. Nam, *J. Microbiol.* **2006**, *44*, 649–654.
48. D. B. Basnet; T. J. Oh; T. T. Vu; B. Sthapit; K. Liou; H. C. Lee; J. C. Yoo; J. K. Sohng, *Mol. Cells* **2006**, *22*, 154–162.
49. Z. Xu; A. Magyar; C. Hertweck, *J. Am. Chem. Soc.* **2007**, *129*, 6022–6030.
50. A. Sandmann; J. Dikschat; H. Jenke-Kodama; B. Kunze; E. Dittmann; R. Müller, *Angew. Chem. Int. Ed.* **2007**, *46*, 2712–2716.
51. A. O. Brachmann; S. A. Joyce; H. Jenke-Kodama; G. Schwär; D. J. Clarke; H. B. Bode, *ChemBioChem* **2007**, *8*, 1721–1728.
52. X. Zhang; L. B. Alemany; H. P. Fiedler; M. Goodfellow; R. J. Parry, *Antimicrob. Agents Chemother.* **2008**, *52*, 574–585.
53. T. Oja; K. Palmu; H. Lehmussola; O. Leppäranta; K. Hännikäinen; J. Niemi; P. Mäntsälä; M. Metsä-Ketelä, *Chem. Biol.* **2008**, *15*, 1046–1057.
54. B. Shen, *Top. Curr. Chem.* **2000**, *209*, 1–51.
55. B. J. Rawlings, *Nat. Prod. Rep.* **2001**, *18*, 190–227.
56. B. J. Rawlings, *Nat. Prod. Rep.* **2001**, *18*, 231–281.
57. W. Bao; E. Wendt-Pienkowski; C. R. Hutchinson, *Biochemistry* **1998**, *37*, 8132–8138.
58. C. W. Carreras; C. Khosla, *Biochemistry* **1998**, *37*, 2084–2088.
59. J. Dreier; A. N. Shah; C. Khosla, *J. Biol. Chem.* **1999**, *274*, 25108–25112.
60. W. P. Revill; M. J. Bibb; D. A. Hopwood, *J. Bacteriol.* **1995**, *177*, 3946–3952.
61. C. J. Arthur; A. Szafranska; S. E. Evans; S. C. Findlow; S. G. Burston; P. Owen; I. Clark-Lewis; T. J. Simpson; J. Crosby; M. P. Crump, *Biochemistry* **2005**, *44*, 15414–15421.
62. T. S. Hitchman; J. Crosby; K. J. Byrom; R. J. Cox; T. J. Simpson, *Chem. Biol.* **1998**, *5*, 35–47.
63. A. L. Matharu; R. J. Cox; J. Crosby; K. J. Byrom; T. J. Simpson, *Chem. Biol.* **1998**, *5*, 699–711.
64. C. Bisang; P. F. Long; J. Cortes; J. Westcott; J. Crosby; A. L. Matharu; R. J. Cox; T. J. Simpson; J. Staunton; P. F. Leadlay, *Nature* **1999**, *401*, 502–505.
65. B. S. Moore; C. Hertweck, *Nat. Prod. Rep.* **2002**, *19*, 70–99.
66. C. R. Hutchinson, *Chem. Rev.* **1997**, *97*, 2525–2536.
67. C. Davies; R. J. Heath; S. W. White; C. O. Rock, *Structure* **2000**, *8*, 185–195.
68. G. Meurer; C. R. Hutchinson, *J. Am. Chem. Soc.* **1995**, *117*, 5899–5900.
69. V. B. Rajgarhia; W. R. Strohl, *J. Bacteriol.* **1997**, *179*, 2690–2696.
70. W. Bao; P. J. Sheldon; E. Wendt-Pienkowski; C. R. Hutchinson, *J. Bacteriol.* **1999**, *181*, 4690–4695.
71. E. S. Meadows; C. Khosla, *Biochemistry* **2001**, *40*, 14855–14861.
72. Y. Tang; T. S. Lee; S. Kobayashi; C. Khosla, *Biochemistry* **2003**, *42*, 6588–6595.
73. T. S. Lee; C. Khosla; Y. Tang, *J. Am. Chem. Soc.* **2005**, *127*, 12254–12262.
74. Y. Tang; N. Y. Lee; H. Y. Lee; C. Khosla, *Tetrahedron* **2004**, *60*, 7659–7671.
75. Y. Tang; T. S. Lee; C. Khosla, *PLoS Biol.* **2004**, *2*, E31.
76. H. Pan; S.-C. Tsai; E. S. Meadows; L. J. W. Miercke; A. T. Keatings-Clay; J. O'Connell; C. Khosla; R. M. Stroud, *Structure* **2002**, *10*, 1559–1568.
77. Z. Xu; M. Metsä-Ketelä; C. Hertweck, *J. Biotechnol.* **2008**, published online.
78. C. Hertweck; B. S. Moore, *Tetrahedron* **2000**, *56*, 9115–9120.
79. C. Hertweck; A. P. Jarvis; L. Xiang; B. S. Moore; N. J. Oldham, *ChemBioChem* **2001**, *2*, 784–786.
80. J. A. Kalaitzis; M. Izumikawa; L. Xiang; C. Hertweck; B. S. Moore, *J. Am. Chem. Soc.* **2003**, *125*, 9290–9291.
81. L. Xiang; B. S. Moore, *J. Bacteriol.* **2003**, *185*, 399–404.
82. M. Izumikawa; Q. Cheng; B. S. Moore, *J. Am. Chem. Soc.* **2006**, *128*, 1428–1429.
83. R. McDaniel; S. Ebert-Khosla; D. A. Hopwood; C. Khosla, *Science* **2005**, *375*, 549–554.
84. A. T. Keatinge-Clay; D. A. Maltby; K. F. Medzihradszky; C. Khosla; R. M. Stroud, *Nat. Struct. Mol. Biol.* **2004**, *11*, 888–993.
85. K. K. Burson; C. Khosla, *Tetrahedron* **2000**, *56*, 9401–9408.
86. T. P. Nicholson; C. Winfield; J. Westcott; J. Crosby; T. J. Simpson; R. J. Cox, *Chem. Commun.* **2003**, 686–687.
87. J. Dreier; C. Khosla, *Biochemistry* **2000**, *39*, 2088–2095.
88. Y. Tang; S.-C. Tsai; C. Khosla, *J. Am. Chem. Soc.* **2003**, *125*, 12708–12709.
89. R. McDaniel; S. Ebert-Khosla; D. A. Hopwood; C. Khosla, *Nature* **1995**, *375*, 549–554.
90. H. Petkovic; A. Thamchaipenet; L.-H. Zhou; D. Hranueli; P. Raspor; P. G. Waterman; I. S. Hunter, *J. Biol. Chem.* **1999**, *274*, 32829–32834.
91. J. Kantola; G. Blanco; A. Hautala; T. Kunnari; J. Hakala; C. Mendez; K. Ylihonko; P. Mäntsälä; J. Salas, *Chem. Biol.* **1997**, *4*, 751–755.
92. G. Meurer; M. Gerlitz; E. Wendt-Pienkowski; L. C. Vining; J. Rohr; C. R. Hutchinson, *Chem. Biol.* **1997**, *4*, 433–443.
93. B. Shen; C. R. Hutchinson, *Biochemistry* **1993**, *32*, 11149–11154.
94. B. Shen; C. R. Hutchinson, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6600–6604.
95. A. Sultana; P. Kallio; A. Jansson; J. S. Wang; J. Niemi; P. Mantsala; G. Schneider, *EMBO J.* **2004**, *23*, 1911–1921.

96. P. Kallio; A. Sultana; J. Niemi; P. Mantsala; G. Schneider, *J. Mol. Biol.* **2006**, *357*, 210–220.
97. B. D. Ames; T. P. Korman; W. Zhang; P. Smith; T. Vu; Y. Tang; S. C. Tsai, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5349–5354.
98. G. Lackner; A. Schenk; Z. Xu; K. Reinhardt; Z. S. Yunt; J. Piel; C. Hertweck, *J. Am. Chem. Soc.* **2007**, *129*, 9306–9312.
99. K. Fritzsche; K. Ishida; C. Hertweck, *J. Am. Chem. Soc.* **2008**, *130*, 8307–8316.
100. A. Schenk; Z. Xu; C. Pfeiffer; C. Steinbeck; C. Hertweck, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 7035–7038.
101. W. Zhang; K. Watanabe; C. C. Wang; Y. Tang, *J. Biol. Chem.* **2007**, *282*, 25717–25725.
102. W. Zhang; K. Watanabe; X. Cai; M. E. Jung; Y. Tang; J. Zhan, *J. Am. Chem. Soc.* **2008**, *130*, 6068–6069.
103. M. J. Fernandez-Lozano; L. L. Remsing; L. M. Quiros; A. F. Braña; E. Fernandez; C. Sanchez; C. Méndez; J. Rohr; J. A. Salas, *J. Biol. Chem.* **2000**, *275*, 3065–3074.
104. D. Rodriguez; L. M. Quiros; J. A. Salas, *J. Biol. Chem.* **2004**, *279*, 8149–8158.
105. A. Luzhetskyy; A. Bechthold, *Appl. Microbiol. Biotechnol.* **2008**, *80*, 945–952.
106. U. Rix; C. Fischer; L. L. Remsing; J. Rohr, *Nat. Prod. Rep.* **2002**, *19*, 542–580.
107. C. J. Thibodeaux; C. E. Melancon, 3rd; H. W. Liu, *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 9814–9859.
108. C. Walsh; C. L. Freel Meyers; H. C. Losey, *J. Med. Chem.* **2003**, *46*, 3425–3436.
109. T. Bililign; B. R. Griffith; J. S. Thorson, *Nat. Prod. Rep.* **2005**, *22*, 742–760.
110. C. J. Thibodeaux; C. E. Melancon; H.-W. Liu, *Nature* **2007**, *446*, 1008–1016.
111. G. Blanco; E. P. Patallo; A. F. Braña; A. Trefzer; A. Bechthold; J. Rohr; C. Méndez; J. A. Salas, *Chem. Biol.* **2001**, *8*, 253–263.
112. A. Luzhetskyy; H. Weiss; A. Charge; E. Welle; A. Linnenbrink; A. Vente; A. Bechthold, *Appl. Microbiol. Biotechnol.* **2007**, *75*, 1367–1375.
113. A. Luzhetskyy; C. Mendez; J. A. Salas; A. Bechthold, *Curr. Top. Med. Chem.* **2008**, *8*, 680–709.
114. J. A. Campbell; G. J. Davies; V. Bulone; B. Henrissat, *Biochem. J.* **1997**, *326*, 929–939.
115. J. A. Campbell; G. J. Davies; V. Bulone; B. Henrissat, (**1997**, *326*, 929) *Biochem. J.* **1998**, *329*, 719.
116. D. Kapitanov; R. K. Yu, *Glycobiology* **1999**, *9*, 961–978.
117. P. M. Coutinho; E. Deleury; G. J. Davies; B. Henrissat, *J. Mol. Biol.* **2003**, *328*, 307–317.
118. W. Klyne, *Biochem. J.* **1950**, *47*, xli–xlii.
119. Y. Bourne; B. Henrissat, *Curr. Opin. Struct. Biol.* **2001**, *11*, 593–600.
120. N. Tarbouriech; S. J. Charnock; G. J. Davies, *J. Mol. Biol.* **2001**, *314*, 655–661.
121. U. M. Unligil; J. M. Rini, *Curr. Opin. Struct. Biol.* **2000**, *10*, 510–517.
122. W. A. Barton; J. Lesniak; J. B. Biggins; P. D. Jeffrey; J. Q. Jiang; K. R. Rajashankar; J. S. Thorson; D. B. Nikolov, *Nat. Struct. Biol.* **2001**, *8*, 545–551.
123. E. S. Burgie; H. M. Holden, *Biochemistry* **2008**, *47*, 3982–3988.
124. M. Mittler; A. Bechthold; G. E. Schulz, *J. Mol. Biol.* **2007**, *372*, 67–76.
125. A. M. Mulichak; H. C. Losey; C. T. Walsh; R. M. Garavito, *Structure* **2001**, *9*, 547–557.
126. G. J. Williams; J. S. Thorson, *Nat. Protoc.* **2008**, *3*, 357–362.
127. C. Zhang; E. Bitto; R. D. Goff; S. Singh; C. A. Bingman; B. R. Griffith; C. Albermann; G. N. Phillips, Jr.; J. S. Thorson, *Chem. Biol.* **2008**, *15*, 842–853.
128. L. Poppe; L. Novak, *Selective Biocatalysis – A Synthetic Approach*, 1st ed.; VCH: Weinheim, New York, Cambridge, 1992.
129. K. S. Rein; J. Borrone, *Comp. Biochem. Physiol.* **1999**, *124*, 117–131.
130. I. Alexeev; A. Sultana; P. Mantsala; J. Niemi; G. Schneider, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 6170–6175.
131. P. Beinker; B. Lohkamp; T. Peltonen; J. Niemi; P. Mantsala; G. Schneider, *J. Mol. Biol.* **2006**, *359*, 728–740.
132. A. Jansson; J. Niemi; Y. Lindqvist; P. Mantsala; G. Schneider, *J. Mol. Biol.* **2003**, *334*, 269–280.
133. H. Koskiniemi; M. Metsa-Ketela; D. Dobritzsch; P. Kallio; H. Korhonen; P. Mäntsälä; G. Schneider; J. Niemi, *J. Mol. Biol.* **2007**, *372*, 633–648.
134. G. Sciara; S. G. Kendrew; A. E. Miele; N. G. Marsh; L. Federici; F. Malatesta; G. Schimperna; C. Savino; B. Vallone, *EMBO J.* **2003**, *22*, 205–215.
135. J. J. Goodman; M. Matrishin; R. W. Young; C. J. Mc, *J. Bacteriol.* **1959**, *78*, 492–499.
136. J. J. Goodman; M. Matrishin; J. R. McCormick, *Nature* **1963**, *198*, 1093–1094.
137. J. R. McCormick; E. R. Jensen, *J. Am. Chem. Soc.* **1969**, *91*, 206.
138. T. Dairi; T. Nakano; K. Aisaka; R. Katsumata; M. Hasegawa, *Biosci. Biotechnol. Biochem.* **1995**, *59*, 1099–1106.
139. T. Nakano; K. Miyake; H. Endo; T. Dairi; T. Mizukami; R. Katsumata, *Biosci. Biotechnol. Biochem.* **2004**, *68*, 1345–1352.
140. H. Deng; S. L. Cobb; A. R. McEwan; R. P. McGlinchey; J. H. Naismith; D. O'Hagan; D. A. Robinson; J. B. Spencer, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 759–762.
141. D. Galonic Fujimori; C. T. Walsh, *Curr. Opin. Chem. Biol.* **2007**, *11*, 1–8.
142. S. Lin; S. G. Van Lanen; B. Shen, *J. Am. Chem. Soc.* **2007**, *129*, 12432–12438.
143. C. S. Neumann; D. G. Fujimori; C. T. Walsh, *Chem. Biol.* **2008**, *15*, 99–109.
144. K. H. van Pee; C. Dong; S. Flecks; J. Naismith; E. P. Patallo; T. Wage, *Adv. Appl. Microbiol.* **2006**, *59*, 127–157.
145. K. H. van Pee; E. P. Patallo, *Appl. Microbiol. Biotechnol.* **2006**, *70*, 631–641.
146. E. Bitto; Y. Huang; C. A. Bingman; S. Singh; J. S. Thorson; G. N. Phillips, Jr., *Proteins* **2008**, *70*, 289–293.
147. C. Dong; S. Flecks; S. Unversucht; C. Haupt; K. H. van Pee; J. H. Naismith, *Science* **2005**, *309*, 2216–2219.
148. A. S. Eustaquio; B. Gust; S. M. Li; S. Pelzer; W. Wohlleben; K. F. Chater; L. Heide, *Chem. Biol.* **2004**, *11*, 1561–1572.
149. L. Heide; L. Westrich; C. Anderle; B. Gust; B. Kammerer; J. Piel, *ChemBioChem* **2008**, *9*, 1992–1999.
150. C. Sanchez; L. Zhu; A. F. Braña; A. P. Salas; J. Rohr; C. Méndez; J. A. Salas, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 461–466.
151. M. Wolpert; L. Heide; B. Kammerer; B. Gust, *ChemBioChem* **2008**, *9*, 603–612.
152. J. He; E. Roemer; C. Lange; X. Huang; A. Maier; G. Kelter; Y. Jiang; L. H. Xu; K. D. Menzel; S. Grabley; H. H. Fiebig; C. L. Jiang; I. Sattler, *J. Med. Chem.* **2007**, *50*, 5168–5175.
153. M. A. Fernandez-Moreno; E. Martinez; J. L. Caballero; K. Ichinose; D. A. Hopwood; F. Malpartida, *J. Biol. Chem.* **1994**, *269*, 24854–24863.
154. H. Fu; R. McDaniel; D. A. Hopwood; C. Khosla, *Biochemistry* **1994**, *33*, 9321–9326.

155. D. A. Hopwood; F. Malpartida; H. M. Kieser; H. Ikeda; J. Duncan; I. Fujii; B. A. Rudd; H. G. Floss; S. Omura, *Nature* **1985**, *314*, 642–644.
156. Z. H. Hu; D. A. Hopwood; C. Khosla, *Appl. Environ. Microbiol.* **2000**, *66*, 2274–2277.
157. K. Ichinose; C. Surti; T. Taguchi; F. Malpartida; K. I. Booker-Milburn; G. R. Stephenson; Y. Ebizuka; D. A. Hopwood, *Bioorg. Med. Chem. Lett.* **1999**, *9*, 395–400.
158. K. Ichinose; T. Taguchi; D. J. Bedford; Y. Ebizuka; D. A. Hopwood, *J. Bacteriol.* **2001**, *183*, 3247–3250.
159. S. G. Kendrew; S. E. Harding; D. A. Hopwood; E. N. G. Marsh, *J. Biol. Chem.* **1995**, *270*, 17339–17343.
160. S. G. Kendrew; D. A. Hopwood; E. N. Marsh, *J. Bacteriol.* **1997**, *179*, 4305–4310.
161. C. Khosla; R. McDaniel; S. Ebert-Khosla; R. Torres; D. H. Sherman; M. J. Bibb; D. A. Hopwood, *J. Bacteriol.* **1993**, *175*, 2197–2204.
162. F. Malpartida; D. A. Hopwood, *Nature* **1984**, *309*, 462–464.
163. F. Malpartida; D. A. Hopwood, *Mol. Gen. Genet.* **1986**, *205*, 66–73.
164. R. McDaniel; S. Ebert-Khosla; H. Fu; D. A. Hopwood; C. Khosla, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11542–11546.
165. T. Taguchi; Y. Ebizuka; D. A. Hopwood; K. Ichinose, *Tetrahedron Lett.* **2000**, *41*, 5253–5256.
166. T. Taguchi; K. Itou; Y. Ebizuka; F. Malpartida; D. A. Hopwood; C. M. Surti; K. I. Booker-Milburn; G. R. Stephenson; K. Ichinose, *J. Antibiot.* **2000**, *53*, 144–152.
167. T. Taguchi; Y. Ebizuka; D. A. Hopwood; K. Ichinose, *J. Am. Chem. Soc.* **2001**, *123*, 11376–11380.
168. A. Bechthold; J. K. Sohng; T. M. Smith; X. Chu; H. G. Floss, *Mol. Gen. Genet.* **1995**, *248*, 610–620.
169. G. Dräger; S. H. Park; H. G. Floss, *J. Am. Chem. Soc.* **1999**, *121*, 2611–2612.
170. H. G. Floss, *J. Ind. Microbiol. Biotechnol.* **2001**, *27*, 183–194.
171. J. J. Lee; J. P. Lee; P. J. Keller; C. E. Cottrell; C. Chang; H. Zahner; H. G. Floss, *J. Antibiot. (Tokyo)* **1986**, *39*, 1123–1134.
172. D. Tornus; H. G. Floss, *J. Antibiot.* **2001**, *54*, 91–101.
173. N. Miyairi; H. Sakai; T. Konomi; H. Imanaka, *J. Antibiot. (Tokyo)* **1976**, *29*, 227–235.
174. Y. Tokuma; N. Miyairi; Y. Morimoto, *J. Antibiot. (Tokyo)* **1976**, *29*, 1114–1116.
175. Q. Cheng; L. Xiang; M. Izumikawa; D. Meluzzi; B. S. Moore, *Nat. Chem. Biol.* **2007**, *3*, 557–558.
176. C. Hertweck; L. Xiang; J. A. Kalaitzis; Q. Cheng; M. Palzer; B. S. Moore, *Chem. Biol.* **2004**, *11*, 461–468.
177. B. S. Moore; C. Hertweck; J. N. Hopke; M. Izumikawa; J. A. Kalaitzis; G. Nilsen; T. O'Hare; J. Piel; P. R. Shipley; L. Xiang; M. B. Austin; J. P. Noel, *J. Nat. Prod.* **2002**, *65*, 1956–1962.
178. B. S. Moore; J. A. Kalaitzis; L. Xiang, *Antonie Van Leeuwenhoek* **2005**, *87*, 49–57.
179. L. Xiang; J. A. Kalaitzis; G. Nilsen; L. Chen; B. S. Moore, *Org. Lett.* **2002**, *4*, 957–960.
180. L. Xiang; B. S. Moore, *J. Biol. Chem.* **2002**, *277*, 32505–32509.
181. L. Xiang; J. A. Kalaitzis; B. S. Moore, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15609–15614.
182. P. J. Proteau, *Bioorg. Chem.* **2004**, *32*, 483–493.
183. G. Batist, *Cancer Chemother. Biol. Response Modif.* **2001**, *19*, 47–58.
184. M. Binaschi; M. Bigioni; A. Cipollone; C. Rossi; C. Goso; C. A. Maggi; G. Capranico; F. Animati, *Curr. Med. Chem. Anti-Cancer Agents* **2001**, *1*, 113–130.
185. G. Minotti; P. Menna; E. Salvatorelli; G. Cairo; L. Gianni, *Pharmacol. Rev.* **2004**, *56*, 185–229.
186. G. Minotti; S. Recalcati; P. Menna; E. Salvatorelli; G. Corna; G. Cairo, *Methods Enzymol.* **2004**, *378*, 340–361.
187. E. C. van Dalen; H. N. Caron; H. O. Dickinson; L. C. Kremer, *Cochrane Database Syst. Rev.* **2008**, CD003917.
188. W. L. Bao; P. J. Sheldon; C. R. Hutchinson, *Biochemistry* **1999**, *38*, 9752–9757.
189. M. Gerlitz; G. Meurer; E. Wendt-Pienkowski; K. Madduri; C. R. Hutchinson, *J. Am. Chem. Soc.* **1997**, *119*, 7392–7393.
190. C. R. Hutchinson; A. L. Colombo, *J. Ind. Microbiol. Biotechnol.* **1999**, *23*, 647–652.
191. A. Jansson; H. Koskiniemi; P. Mäntsälä; J. Niemi; G. Schneider, *J. Biol. Chem.* **2004**, *279*, 41149–41156.
192. S. G. Kendrew; K. Katayama; E. Deutsch; K. Madduri; C. R. Hutchinson, *Biochemistry* **1999**, *38*, 4794–4799.
193. J. Niemi; Y. Wang; K. Airas; K. Ylihonko; J. Hakala; P. Mäntsälä, *Biochim. Biophys. Acta* **1999**, *1430*, 57–64.
194. S. L. Otten; J. Ferguson; C. R. Hutchinson, *J. Bacteriol.* **1995**, *177*, 1216–1224.
195. S. L. Otten; X. C. Liu; J. Ferguson; C. R. Hutchinson, *J. Bacteriol.* **1995**, *177*, 6688–6692.
196. S. L. Otten; M. A. Gallo; K. Madduri; X. C. Liu; C. R. Hutchinson, *J. Bacteriol.* **1997**, *179*, 4446–4450.
197. S. L. Otten; C. Olano; C. R. Hutchinson, *Microbiology* **2000**, *146*, 1457–1468.
198. V. B. Rajgarhia; N. D. Priestley; W. R. Strohl, *Metab. Eng.* **2001**, *3*, 49–63.
199. R. J. Walczak; J. V. Hines; W. R. Strohl; N. D. Priestley, *Org. Lett.* **2001**, *3*, 2277–2279.
200. J. Kantola; T. Kunnari; A. Hautala; J. Hakala; K. Ylihonko; P. Mäntsälä, *Microbiology* **2000**, *146*, 155–163.
201. T. Kunnari; K. D. Klika; G. Blanco; C. Méndez; P. Mäntsälä; J. Hakala; R. Sillanpaa; P. Tahtinen; J. Salas; K. Ylihonko, *J. Chem. Soc. Perkin Trans.* **2002**, *1*, 1818–1825.
202. N. Lomovskaya; S. L. Otten; Y. Doi-Katayama; L. Fonstein; X. C. Liu; T. Takatsu; A. Inventi-Solari; S. Filippini; F. Torti; A. L. Colombo; C. R. Hutchinson, *J. Bacteriol.* **1999**, *181*, 305–318.
203. K. Madduri; J. Kennedy; G. Rivola; A. Inventi-Solari; S. Filippini; G. Zanuso; A. L. Colombo; K. M. Gewain; J. L. Occi; D. J. MacNeil; C. R. Hutchinson, *Nat. Biotechnol.* **1998**, *16*, 69–74.
204. M. Metsä-Ketelä; K. Palmu; T. Kunnari; K. Ylihonko; P. Mäntsälä, *Antimicrob. Agents Chemother.* **2003**, *47*, 1291–1296.
205. J. Niemi; K. Ylihonko; J. Hakala; R. Pärssinen; A. Kopio; P. Mäntsälä, *Microbiology* **1994**, *140*, 1351–1358.
206. J. Niemi; P. Mäntsälä, *J. Bacteriol.* **1995**, *177*, 2942–2945.
207. C. Scotti; C. R. Hutchinson, *J. Bacteriol.* **1996**, *178*, 7316–7321.
208. S. E. Wohlert; E. Wendt-Pienkowski; W. L. Bao; C. R. Hutchinson, *J. Nat. Prod.* **2001**, *64*, 1077–1080.
209. A. Hautala; S. Torkkell; K. Raty; T. Kunnari; J. Kantola; P. Mantsala; K. Ylihonko, *J. Antibiot. (Tokyo)* **2003**, *56*, 143–153.
210. H. S. Kim; Y. S. Hong; Y. H. Kim; O. J. Yoo; J. J. Lee, *J. Antibiot.* **1996**, *49*, 355–360.
211. C. Leimkuhler; M. Fridman; T. Lupoli; S. Walker; C. T. Walsh; D. Kahne, *J. Am. Chem. Soc.* **2007**, *129*, 10546–10550.
212. W. Lu; C. Leimkuhler; M. Oberthur; D. Kahne; C. T. Walsh, *Biochemistry* **2004**, *43*, 4548–4558.
213. W. Lu; C. Leimkuhler; G. J. Gatto, Jr.; R. G. Kruger; M. Oberthur; D. Kahne; C. T. Walsh, *Chem. Biol.* **2005**, *12*, 527–534.

214. T. Oki; Y. Matsuzawa; A. Yoshimoto; K. Numata; I. Kitamura; S. Hori; A. Takamatsu; H. Umezawa; M. Ishizuka; H. Naganawa; H. Suda; M. Hamada; T. Takeuchi, *J. Antibiot.* **1975**, *28*, 830–834.
215. T. Oki; A. Yoshimoto; Y. Matsuzawa; T. Takeuchi; H. Umezawa, *J. Antibiot. (Tokyo)* **1980**, *33*, 1331–1340.
216. C. Olano; M. S. Abdelfattah; S. Gullón; A. F. Braña; J. Rohr; C. Méndez; J. A. Salas, *ChemBioChem* **2008**, *9*, 624–633.
217. K. Raty; T. Kunnari; J. Hakala; P. Mantsala; K. Ylihonko, *Mol. Gen. Genet.* **2000**, *264*, 164–172.
218. K. Räty; A. Hautala; S. Torkkell; J. Kantola; P. Mäntsälä; J. Hakala; K. Ylihonko, *Microbiology* **2002**, *148*, 3375–3384.
219. A. Sultana; P. Kallio; A. Jansson; J. Niemi; P. Mantsala; G. Schneider, *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 1118–1120.
220. A. Sultana; I. Alexeev; I. Kursula; P. Mantsala; J. Niemi; G. Schneider, *Acta Crystallogr. D Biol. Crystallogr.* **2007**, *63*, 149–159.
221. S. Torkkell; T. Kunnari; K. Palmu; J. Hakala; P. Mantsala; K. Ylihonko, *Antimicrob. Agents Chemother.* **2000**, *44*, 396–399.
222. Y. L. Wang; J. Niemi; K. Airas; K. Ylihonko; J. Hakala; P. Mantsala, *Biochim. Biophys. Acta – Protein Struct. Mol. Enzymol.* **2000**, *1480*, 191–200.
223. B. E. Leach; K. M. Calhoun; L. E. Johnson; C. M. Teeters; W. G. Jackson, *J. Am. Chem. Soc.* **1953**, *75*, 4011–4012.
224. E. Simonitsch; W. Eisenhuth; O. A. Stamm; H. Schmid, *Helv. Chim. Acta* **1964**, *47*, 1459–1484.
225. J. P. McGovren; G. L. Neil; S. L. Crampton; M. I. Robinson; J. D. Douros, *Cancer Res.* **1977**, *37*, 1666–1672.
226. M. Konishi; K. Sugawara; F. Kofu; Y. Nishiyama; K. Tomita; T. Miyaki; H. Kawaguchi, *J. Antibiot.* **1986**, *39*, 784–791.
227. G. Asai; N. Yamamoto; M. Toi; E. Shin; K. Nishiyama; T. Sekine; Y. Nomura; S. Takashima; M. Kimura; T. Tominaga, *Cancer Chemother. Pharmacol.* **2002**, *49*, 468–472.
228. X. Salas; J. Portugal, *FEBS Lett.* **1991**, *292*, 223–228.
229. A. Lorico; B. H. Long, *Eur. J. Cancer* **1993**, *29A*, 1985–1991.
230. K. Kon; H. Sugi; K. Tamai; Y. Ueda; N. Yamada, *J. Antibiot.* **1990**, *43*, 372–382.
231. M. Takai; Y. Uehara; J. A. Beisler, *J. Med. Chem.* **1980**, *23*, 549–553.
232. P. Canham; L. C. Vining, *J. Chem. Soc. Chem. Commun.* **1976**, *80*, 319–320.
233. P. L. Canham; L. C. Vining; A. G. McInnes; J. A. Walter; J. L. C. Wright, *Can. J. Chem.* **1977**, *55*, 2450–2457.
234. K. S. Lam; J. A. Veitch; S. Forenza; C. M. Combs; K. L. Colson, *J. Nat. Prod.* **1989**, *52*, 1015–1021.
235. J. R. Brown; M. S. Spring; J. R. Stoker, *Phytochemistry* **1971**, *10*, 2059–2064.
236. R. P. Maskey; I. Grün-Wollny; H. Laatsch, *J. Antibiot.* **2003**, *56*, 795–800.
237. T. Simonart; M. Dramaix; V. De Maertelaer, *Br. J. Dermatol.* **2008**, *158*, 208–216.
238. L. A. Mitscher, Antibiotics and Antimicrobial Agents. In *Foye's Principles of Medicinal Chemistry*; D. A. Williams, T. L. Lemke, Eds.; Lippincott Williams & Wilkins: Philadelphia, 2002; pp 819–866.
239. U. Cosentino; D. Pitea; G. Moro; G. A. Saracino; P. Caria; R. M. Vari; L. Colombo; G. Forloni; F. Tagliavini; M. Salmona, *J. Mol. Model* **2008**, *14*, 987–994.
240. A. De Luigi; L. Colombo; L. Diomede; R. Capobianco; M. Mangieri; C. Miccolo; L. Limido; G. Forloni; F. Tagliavini; M. Salmona, *PLoS ONE* **2008**, *3*, e1888.
241. F. Tagliavini; G. Forloni; L. Colombo; G. Rossi; L. Girola; B. Canciani; N. Angeretti; L. Giampaolo; E. Peressini; T. Awan; L. D. Gioia; E. Ragg; O. Bugiani; M. Salmona, *J. Mol. Biol.* **2000**, *300*, 1309–1322.
242. D. Orsucci; M. Mancuso; G. Siciliano, *J. Physiol.* **2008**, *586*, 2427–2428.
243. M. Hatsu; T. Sasaki; S. Gomi; Y. Kodama; M. Sezaki; S. Inouye; S. Kondo, *J. Antibiot. (Tokyo)* **1992**, *45*, 325–330.
244. M. Hatsu; T. Sasaki; H. Watabe; S. Miyadoh; M. Nagasawa; T. Shomura; M. Sezaki; S. Inouye; S. Kondo, *J. Antibiot. (Tokyo)* **1992**, *45*, 320–324.
245. T. Horiguchi; K. Hayashi; S. Tsubotani; S. Iinuma; S. Harada; S. Tanida, *J. Antibiot. (Tokyo)* **1994**, *47*, 545–556.
246. T. Horiguchi; S. Tanida, *Biochem. Pharmacol.* **1995**, *49*, 1395–1401.
247. K. Herold; F. A. Gollmick; I. Groth; M. Roth; K. D. Menzel; U. Mollmann; U. Grafe; C. Hertweck, *Chemistry* **2005**, *11*, 5523–5530.
248. L. J. Xuan; S. H. Xu; H. L. Zhang; Y. M. Xu; M. Q. Chen, *J. Antibiot. (Tokyo)* **1992**, *45*, 1974–1976.
249. I. Momose; W. Chen; N. Kinoshita; H. Iinuma; M. Hamada; T. Takeuchi, *J. Antibiot. (Tokyo)* **1998**, *51*, 21–25.
250. I. Momose; W. Chen; H. Nakamura; H. Naganawa; H. Iinuma; T. Takeuchi, *J. Antibiot. (Tokyo)* **1998**, *51*, 26–32.
251. T. Paululat; A. Zeeck; J. M. Gutterer; H. P. Fiedler, *J. Antibiot. (Tokyo)* **1999**, *52*, 96–101.
252. T. Dairi; K. Aisaka; R. Katsumata; M. Hasegawa, *Biosci. Biotechnol. Biochem.* **1995**, *59*, 1835–1841.
253. T. Dairi; T. Nakano; T. Mizukami; K. Aisaka; M. Hasegawa; R. Katsumata, *Biosci. Biotechnol. Biochem.* **1995**, *59*, 1360–1361.
254. J. R. McCormick; E. R. Jensen, *J. Am. Chem. Soc.* **1965**, *87*, 1794–1795.
255. J. R. McCormick; U. H. Joachim; E. R. Jensen; S. Johnson; N. O. Sjolander, *J. Am. Chem. Soc.* **1965**, *87*, 1793–1794.
256. J. R. McCormick; E. R. Jensen, *J. Am. Chem. Soc.* **1968**, *90*, 7126–7127.
257. J. R. McCormick; E. R. Jensen; N. Arnold; H. S. Corey; U. H. Joachim; S. Johnson; P. A. Miller; N. O. Sjolander, *J. Am. Chem. Soc.* **1968**, *90*, 7127–7129.
258. J. R. McCormick; E. R. Jensen; S. Johnson; N. O. Sjolander, *J. Am. Chem. Soc.* **1968**, *90*, 2201–2202.
259. I. S. Hunter; R. A. Hill, In *Biotechnology of Antibiotics*; W. R. Strohl, Ed.; Marcel Decker Inc.: New York, 1997; pp 659–682.
260. I. S. Hunter, Tetracyclines. In *Microbial Secondary Metabolites: Biosynthesis, Genetics and Regulation*; J. F. Martin, Ed.; Research Signpost: Lucknow, 2002; pp 141–166.
261. N. Perić-Concha; B. Borovicka; P. F. Long; D. Hranueli; P. G. Waterman; I. S. Hunter, *J. Biol. Chem.* **2005**, *280*, 37455–37460.
262. H. Petkovic; A. Thamchaipenet; L.-H. Zhou; D. Hranueli; P. Raspor; P. G. Waterman; I. S. Hunter, *J. Biol. Chem.* **1999**, *274*, 32829–32834.
263. W. Zhang; K. Watanabe; C. C. Wang; Y. Tang, *J. Nat. Prod.* **2006**, *69*, 1633–1636.
264. S. C. Findlow; C. Winsor; T. J. Simpson; J. Crosby; M. P. Crump, *Biochemistry* **2003**, *42*, 8423–8433.
265. H. Jayasuriya; R. B. Lingham; P. Graham; D. Quamina; L. Herranz; O. Genilloud; M. Gagliardi; R. Danzeisen; J. E. Tomassini; D. L. Zink; Z. Q. Guan; S. B. Singh, *J. Nat. Prod.* **2002**, *65*, 1091–1095.
266. R. Katahira; M. Katahira; Y. Yamashita; H. Ogawa; Y. Kyogoku; M. Yoshida, *Nucleic Acids Res.* **1998**, *26*, 744–755.
267. R. Katahira; Y. Uosaki; H. Ogawa; Y. Yamashita; H. Nakano; M. Yoshida, *J. Antibiot.* **1998**, *51*, 267–274.
268. F. Lombo; N. Menendez; J. A. Salas; C. Mendez, *Appl. Microbiol. Biotechnol.* **2006**, *73*, 1–14.
269. H. Ogawa; Y. Yamashita; R. Katahira; S. Chiba; T. Iwasaki; T. Ashizawa; H. Nakano, *J. Antibiot.* **1998**, *51*, 261–266.
270. J. Rohr; C. Méndez; J. A. Salas, *Bioorg. Chem.* **1999**, *27*, 41–54.

271. K. M. Anderson; M. Rubenstein; H. Sky-Peck, *Oncology* **1982**, *39*, 72–77.
272. A. J. Coukell; A. Markham, *Drugs Aging* **1998**, *12*, 149–168.
273. E. G. Elias; J. T. Evans, *J. Bone Joint Surg.* **1972**, *54-A*, 1730–1736.
274. A. G. Hadjipavlou; L. N. Gaitanis; P. G. Katonis; P. Lander, *Eur. Spine J.* **2001**, *10*, 370–384.
275. T. J. Hall; M. Schaeublin; T. J. Chambers, *Biochem. Biophys. Res. Commun.* **1993**, *195*, 1245–1253.
276. G. Koutsodontis; D. Kardassis, *Oncogene* **2004**, *23*, 9190–9200.
277. T. J. Lee; E. M. Jung; J. T. Lee; S. Kim; J. W. Park; K. S. Choi; T. K. Kwon, *Mol. Cancer Ther.* **2006**, *5*, 2737–2746.
278. L. L. Remsing; H. R. Bahadori; G. M. Carbone; E. M. McGuffie; C. V. Catapano; J. Rohr, *Biochemistry* **2003**, *42*, 8313–8324.
279. T. J. Rosol; D. J. Chew; C. G. Couto; R. D. Ayl; L. A. Nagode; C. C. Capen, *Vet. Pathol.* **1992**, *29*, 223–229.
280. S. Chatterjee; K. Zaman; H. Ryu; A. Conforto; R. R. Ratan, *Ann. Neurol.* **2001**, *49*, 345–354.
281. H. Ryu; J. Lee; K. Zaman; J. Kubilis; R. J. Ferrante; B. D. Ross; R. Neve; R. R. Ratan, *J. Neurosci.* **2003**, *23*, 3597–3606.
282. S. Y. Liu; J. P. Rosazza, *Appl. Environ. Microbiol.* **1998**, *64*, 3972–3976.
283. A. Montanari; J. P. Rosazza, *J. Antibiot. (Tokyo)* **1990**, *43*, 883–889.
284. M. S. Abdelfattah; J. Rohr, *Angew. Chem. Int. Ed.* **2006**, *45*, 5685–5689.
285. V. Albertini; A. Jain; S. Vignati; S. Napoli; A. Rinaldi; I. Kwee; M. Nur-e-Alam; J. Bergant; F. Bertoni; G. M. Carbone; J. Rohr; C. V. Catapano, *Nucleic Acids Res.* **2006**, *34*, 1721–1734.
286. G. Blanco; E. Fernandez; M. J. Fernandez; A. F. Brana; U. Weissbach; E. Kunzel; J. Rohr; C. Mendez; J. A. Salas, *Mol. Gen. Genet.* **2000**, *262*, 991–1000.
287. E. Fernandez; U. Weissbach; C. Sanchez Reillo; A. F. Braña; C. Méndez; J. Rohr; J. A. Salas, *J. Bacteriol.* **1998**, *180*, 4929–4937.
288. M. Gibson; M. Nur-e-Alam; F. Lipata; M. A. Oliveira; J. Rohr, *J. Am. Chem. Soc.* **2005**, *127*, 17594–17595.
289. A. Gonzalez; L. L. Remsing; F. Lombó; M. J. Fernandez; L. Prado; A. F. Braña; E. Künzel; J. Rohr; C. Méndez; J. A. Salas, *Mol. Gen. Genet.* **2001**, *264*, 827–835.
290. F. Lombó; E. Künzel; L. Prado; A. F. Braña; K. U. Bindseil; J. Frevert; D. Bearden; C. Méndez; J. A. Salas; J. Rohr, *Angew. Chem. Int. Ed. Engl.* **2000**, *39*, 796–799.
291. M. J. Lozano; L. L. Remsing; L. M. Quiros; A. F. Brana; E. Fernandez; C. Sanchez; C. Mendez; J. Rohr; J. A. Salas, *J. Biol. Chem.* **2000**, *275*, 3065–3074.
292. N. Menendez; M. Nur-e-Alam; C. Fischer; A. F. Brana; J. A. Salas; J. Rohr; C. Mendez, *Appl. Environ. Microbiol.* **2006**, *72*, 167–177.
293. M. Nur-e-Alam; C. Méndez; J. A. Salas; J. Rohr, *ChemBioChem* **2005**, *6*, 632–636.
294. L. Prado; E. Fernandez; U. Weissbach; G. Blanco; L. M. Quiros; A. F. Brana; C. Méndez; J. Rohr; J. A. Salas, *Chem. Biol.* **1999**, *6*, 19–30.
295. L. Prado; F. Lombó; A. F. Braña; C. Méndez; J. Rohr; J. A. Salas, *Mol. Gen. Genet.* **1999**, *261*, 216–225.
296. L. L. Remsing; J. Garcia-Bernardo; A. M. Gonzalez; E. Künzel; U. Rix; A. F. Braña; D. W. Bearden; C. Méndez; J. A. Salas; J. Rohr, *J. Am. Chem. Soc.* **2002**, *124*, 1606–1614.
297. L. L. Remsing; A. M. Gonzalez; M. Nur-e-Alam; M. J. Fernandez-Lozano; A. F. Brana; U. Rix; M. A. Oliveira; C. Mendez; J. A. Salas; J. Rohr, *J. Am. Chem. Soc.* **2003**, *125*, 5745–5753.
298. J. Rohr; U. Weissbach; C. Beninga; E. Künzel; K. Siems; K. Bindseil; L. Prado; F. Lombó; A. F. Braña; C. Méndez; J. A. Salas, *Chem. Commun.* **1998**, 437–438.
299. C. Wang; M. Gibson; J. Rohr; M. A. Oliveira, *Acta Crystallogr.* **2005**, *F61*, 1023–1026.
300. S. E. Wohlert; E. Kunzel; R. Machinek; C. Mendez; J. A. Salas; J. Rohr, *J. Nat. Prod.* **1999**, *62*, 119–121.
301. K. Krohn; J. Rohr, *Top. Curr. Chem.* **1997**, *188*, 127–195.
302. J. Rohr; R. Thiericke, *Nat. Prod. Rep.* **1992**, *9*, 103–137.
303. R. T. Crow; B. Rosenbaum; R. Smith, 3rd; Y. Guo; K. S. Ramos; G. A. Sulikowski, *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1663–1666.
304. H. Depenbrock; S. Bornschlegl; R. Peter; J. Rohr; P. Schmid; P. Schweighart; T. Block; J. Rastetter; A. R. Hanauske, *Ann. Hematol.* **1996**, *73* (Suppl II), A80/316.
305. H. Drautz; H. Zahner; J. Rohr; A. Zeeck, *J. Antibiot.* **1986**, *39*, 1657–1669.
306. T. Henkel; J. Rohr; J. M. Beale; L. Schwenen, *J. Antibiot.* **1990**, *43*, 492–503.
307. J. Rohr; A. Zeeck, *J. Antibiot. (Tokyo)* **1987**, *40*, 459–467.
308. J. Rohr; A. Zeeck; H. G. Floss, *J. Antibiot. (Tokyo)* **1988**, *41*, 126–129.
309. S. Weber; C. Zolke; J. Rohr; J. M. Beale, *J. Org. Chem.* **1994**, *59*, 4211–4214.
310. A. Korynevska; P. Heffeter; B. Matselyukh; L. Elbling; M. Micksche; R. Stoika; W. Berger, *Biochem. Pharmacol.* **2007**, *74*, 1713–1726.
311. J. Rohr; J. M. Beale; H. G. Floss, *J. Antibiot. (Tokyo)* **1989**, *42*, 1151–1157.
312. J. Rohr; M. Schönewolf; G. Udvarnoki; K. Eckardt; G. Schumann; C. Wagner; J. M. Beale; S. D. Sorey, *J. Org. Chem.* **1993**, *58*, 2547–2551.
313. G. Udvarnoki; T. Henkel; R. Machinek; J. Rohr, *J. Org. Chem.* **1992**, *57*, 1274–1276.
314. I. Baig; M. Kharel; A. Kobylyanskyy; L. L. Zhu; Y. Rebets; B. Ostash; A. Luzhetskyy; A. Bechthold; V. A. Fedorenko; J. Rohr, *Angew. Chem. Int. Ed.* **2006**, *45*, 7842–7846.
315. B. Faust; D. Hoffmeister; G. Weitnauer; L. Westrich; S. Haag; P. Schneider; H. Decker; E. Künzel; J. Rohr; A. Bechthold, *Microbiology* **2000**, *146*, 147–154.
316. O. Gromyko; Y. Rebets; B. Ostash; A. Luzhetskyy; M. Fukuhara; A. Bechthold; T. Nakamura; V. Fedorenko, *J. Antibiot. (Tokyo)* **2004**, *57*, 383–389.
317. D. Hoffmeister; K. Ichinose; S. Domann; B. Faust; A. Trefzer; G. Dräger; A. Kirschning; C. Fischer; E. Künzel; D. W. Bearden; J. Rohr; A. Bechthold, *Chem. Biol.* **2000**, *7*, 821–831.
318. D. Hoffmeister; K. Ichinose; A. Bechthold, *Chem. Biol.* **2001**, *8*, 557–567.
319. D. Hoffmeister; M. Weber; G. Drager; K. Ichinose; C. Dürr; A. Bechthold, *ChemBioChem* **2004**, *5*, 369–371.
320. C. Krauth; M. Fedoryshyn; C. Schleberger; A. Luzhetskyy; A. Bechthold, *Chem. Biol.* **2009**, *16*, 28–35.
321. A. Luzhetskyy; T. Liu; M. Fedoryshyn; B. Ostash; V. Fedorenko; J. Rohr; A. Bechthold, *ChemBioChem* **2004**, *5*, 1567–1570.
322. A. Luzhetskyy; M. Fedoryshyn; C. Dürr; T. Taguchi; V. Novikov; A. Bechthold, *Chem. Biol.* **2005**, *12*, 725–729.

323. A. Luzhetskyy; T. Taguchi; M. Fedoryshyn; C. Dürr; S. E. Wohlert; V. Novikov; A. Bechthold, *ChemBioChem* **2005**, *6*, 1406–1410.
324. A. Luzhetskyy; L. Zhu; M. Gibson; M. Fedoryshyn; C. Dürr; C. Hofmann; D. Hoffmeister; B. Ostash; C. Mattingly; V. Adams; V. Fedorenko; J. Rohr; A. Bechthold, *ChemBioChem* **2005**, *6*, 675–678.
325. A. Mayer; T. Taguchi; A. Linnenbrink; C. Hofmann; A. Luzhetskyy; A. Bechthold, *ChemBioChem* **2005**, *6*, 2312–2315.
326. I. Ostash; B. Ostash; A. Luzhetskyy; A. Bechthold; S. Walker; V. Fedorenko, *FEMS Microbiol. Lett.* **2008**, *285*, 195–202.
327. I. Ostash; Y. Rebets; B. Ostash; A. Kobylyanskyy; M. Myronovskyy; T. Nakamura; S. Walker; V. Fedorenko, *Arch. Microbiol.* **2008**, *190*, 105–109.
328. Y. Rebets; L. Dutko; B. Ostash; A. Luzhetskyy; O. Kulachkovskyy; T. Yamaguchi; T. Nakamura; A. Bechthold; V. Fedorenko, *Arch. Microbiol.* **2008**, *189*, 111–120.
329. U. Rix; L. L. Remsing; D. Hoffmeister; A. Bechthold; J. Rohr, *ChemBioChem* **2003**, *4*, 109–111.
330. S.-E. Wohlert; A. Bechthold; C. Beninga; T. Henkel; M. Holzenkämpfer; A. Kirschning; C. Oelkers; M. Ries; S. Weber; U. Weissbach; L. Westrich; J. Rohr, Investigations on the Biosynthesis of Landomycin A. In *Bioorganic Chemistry*; U. Diederichsen, T. K. Lindhorst, B. Westermann, L. A. Wessjohann, Eds.; Wiley-VCH: Weinheim, New York, Chichester, Brisbane, Singapore, Toronto, 1999; pp 305–312.
331. L. Zhu; B. Ostash; U. Rix; M. Nur-e-Alam; A. Mayers; A. Luzhetskyy; C. Méndez; J. A. Salas; A. Bechthold; V. Fedorenko; J. Rohr, *J. Org. Chem.* **2005**, *70*, 631–638.
332. L. Zhu; A. Luzhetskyy; M. Luzhetska; C. Mattingly; V. Adams; A. Bechthold; J. Rohr, *ChemBioChem* **2007**, *8*, 83–88.
333. A. Trefzer; G. Blanco; L. L. Remsing; E. Künzel; U. Rix; A. F. Braña; C. Méndez; J. Rohr; A. Bechthold; J. A. Salas, *J. Am. Chem. Soc.* **2002**, *124*, 6056–6062.
334. S.-E. Wohlert; G. Blanco; F. Lombo; E. Fernandez; A. F. Brana; S. Reich; G. Udvarnoki; C. Mendez; H. Decker; J. Frevert; J. A. Salas; J. Rohr, *J. Am. Chem. Soc.* **1998**, *120*, 10596–10601.
335. J. Rohr, *J. Chem. Soc. Chem. Commun.* **1989**, 492–493.
336. J. Rohr, *J. Chem. Soc. Chem. Commun.* **1990**, 113–114.
337. J. Rohr, *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 1051–1053.
338. S. Blanchard; J. S. Thorson, *Curr. Opin. Chem. Biol.* **2006**, *10*, 263–271.
339. C. Duerr; D. Hoffmeister; S.-E. Wohlert; K. Ichinose; M. Weber; U. von Mulert; J. S. Thorson; A. Bechthold, *Angew. Chem. Int. Ed.* **2004**, *43*, 2962–2965.
340. D. Hoffmeister; G. Drager; K. Ichinose; J. Rohr; A. Bechthold, *J. Am. Chem. Soc.* **2003**, *125*, 4678–4679.
341. E. Künzel; B. Faust; C. Oelkers; U. Weissbach; D. W. Bearden; G. Weitnauer; L. Westrich; A. Bechthold; J. Rohr, *J. Am. Chem. Soc.* **1999**, *121*, 11058–11062.
342. A. Bechthold; G. Weitnauer; A. Luzhetskyy; M. Berner; C. Bihlmeier; R. Boll; C. Durr; A. Frerich; C. Hofmann; A. Mayer; I. Treede; A. Vente; M. Luzhetskyy, *Ernst Schering Res. Found. Workshop* **2005**, 147–163.
343. A. Luzhetskyy; A. Vente; A. Bechthold, *Mol. Biosyst.* **2005**, 117–126.
344. A. Luzhetskyy; S. Pelzer; A. Bechthold, *Curr. Opin. Invest. Drugs* **2007**, *8*, 608–613.
345. F. Lombó; A. F. Braña; J. A. Salas; C. Méndez, *ChemBioChem* **2004**, *5*, 1181–1187.
346. C. Méndez; E. Künzel; F. Lipata; F. Lombó; W. Cotham; M. Walla; D. W. Bearden; A. F. Braña; J. A. Salas; J. Rohr, *J. Nat. Prod.* **2002**, *65*, 779–782.
347. F. Lombo; M. S. Abdelfattah; A. F. Brana; J. A. Salas; J. Rohr; C. Mendez, *ChemBioChem* **2009**, *10*, 296–303.
348. P. Kallio; Z. Liu; P. Mantsala; J. Niemi; M. Metsa-Ketela, *Chem. Biol.* **2008**, *15*, 157–166.
349. G. W. C. Udvarnoki; R. Machinek; J. Rohr, *Angew. Chem. Int. Ed.* **1995**, *34*, 565–567.
350. R. H. Flatman; A. J. Howells; L. Heide; H. P. Fiedler; A. Maxwell, *Antimicrob. Agents Chemother.* **2005**, *49*, 1093–1100.
351. M. Holzenkampfer; M. Walker; A. Zeeck; J. Schimana; H. P. Fiedler, *J. Antibiot.* **2002**, *55*, 301–307.
352. A. A. Sadiq; M. R. Patel; B. A. Jacobson; M. Escobedo; K. Ellis; L. M. Oppegard; H. Hiasa; R. A. Kratzke, *Invest. New Drugs* **2009**.
353. J. Schimana; H. P. Fiedler; I. Groth; R. Sussmuth; W. Beil; M. Walker; A. Zeeck, *J. Antibiot.* **2000**, *53*, 779–787.
354. J. Schimana; M. Walker; A. Zeeck; P. Fiedler, *J. Ind. Microbiol. Biotechnol.* **2001**, *27*, 144–148.
355. U. Theobald; J. Schimana; H. P. Fiedler, *Antonie Van Leeuwenhoek* **2000**, *78*, 307–313.
356. U. Galm; J. Schimana; H. P. Fiedler; J. Schmidt; S. M. Li; L. Heide, *Arch. Microbiol.* **2002**, *178*, 102–114.
357. M. Holzenkampfer; A. Zeeck, *J. Antibiot. (Tokyo)* **2002**, *55*, 341–342.
358. T. Luft; S. M. Li; H. Scheible; B. Kammerer; L. Heide, *Arch. Microbiol.* **2005**, *183*, 277–285.
359. M. Pacholec; C. L. Freel Meyers; M. Oberthur; D. Kahne; C. T. Walsh, *Biochemistry* **2005**, *44*, 4949–4956.
360. A. Trefzer; S. Pelzer; J. Schimana; S. Stockert; C. Bihlmaier; H. P. Fiedler; K. Welzel; A. Vente; A. Bechthold, *Antimicrob. Agents Chemother.* **2002**, *46*, 1174–1182.
361. D. M. Balitz; F. A. O'Herron; J. Bush; D. M. Vyas; D. E. Nettleton; R. E. Grulich; W. T. Bradner; T. W. Doyle; E. Arnold; J. Clardy, *J. Antibiot.* **1981**, *34*, 1544–1555.
362. N. Hirayama; K. Takahashi; K. Shirahata; Y. Ohashi; Y. Sasada, *Bull. Chem. Soc. Jpn.* **1981**, *54*, 1338–1342.
363. T. Hosoya; E. Takashiro; T. Matsumoto; K. Suzuki, *J. Am. Chem. Soc.* **1994**, *116*, 1004–1015.
364. H. Nakano; Y. Matsuda; K. Ito; S. Ohkubo; M. Morimoto; F. Tomita, *J. Antibiot.* **1981**, *34*, 266–270.
365. T. Nakashima; T. Fujii; K. Sakai; S. Tomohiro; H. Kumagai; T. Yoshioka, Chrysomycin Derivative Compounds and Use as Antitumor Agents. U.S. Patent 6,030,951; Mercian Corporation (JP): United States, 2000.
366. K. Takahashi; M. Yoshida; F. Tomita; K. Shirahata, *J. Antibiot.* **1981**, *34*, 271–275.
367. J. A. Findlay; J. S. Liu; L. Radics; S. Rakhit, *Can. J. Chem.* **1981**, *59*, 3018–3020.
368. D.-S. Hsu; T. Matsumoto; K. Suzuki, *Synlett* **2005**, 801–804.
369. S. N. Sehgal; H. Czerkawski; A. Kudelski; K. Pandev; R. Saucier; C. Vezina, *J. Antibiot.* **1983**, *35*, 355–361.
370. K. Suzuki, *Pure Appl. Chem.* **2000**, *72*, 1783–1786.
371. N. Yamashita; K. Shin-ya; K. Furihata; Y. Hayakawa; H. Seto, *J. Antibiot.* **1998**, *51*, 1105–1108.
372. D. J. Hart; G. H. Merriman; D. G. J. Young, *Tetrahedron* **1996**, *52*, 14437–14458.
373. T. T. Wei; K. M. Byrne; D. Warnick-Pickle; M. Greenstein, *J. Antibiot.* **1982**, *35*, 545–548.
374. U. Weiss; K. Yoshihira; R. J. Highet; R. J. White; T. T. Wei, *J. Antibiot.* **1982**, *35*, 1194–1201.

375. K. Kojiri; H. Arakawa; F. Satoh; K. Kawamura; A. Okura; H. Suda; M. Okanishi, *J. Antibiot. (Tokyo)* **1991**, *44*, 1054–1060.
376. Y. Q. Li; X. S. Huang; K. Ishida; A. Maier; G. Kelter; Y. Jiang; G. Peschel; K. D. Menzel; M. G. Li; M. L. Wen; L. H. Xu; S. Grabley; H. H. Fiebig; C. L. Jiang; C. Hertweck; I. Sattler, *Org. Biomol. Chem.* **2008**, *6*, 3601–3605.
377. S. Nakajima; K. Kojiri; H. Suda; M. Okanishi, *J. Antibiot. (Tokyo)* **1991**, *44*, 1061–1064.
378. A. E. Alegria; L. Zayas; N. Guevara, *Photochem. Photobiol.* **1995**, *62*, 409–415.
379. R. Arce; R. Oyola; A. E. Alegria, *Photochem. Photobiol.* **1998**, *68*, 25–31.
380. R. K. Elespuru; S. K. Gonda, *Science* **1984**, *223*, 69–71.
381. A. Matsumoto; Y. Fujiwara; R. K. Elespuru; P. C. Hanawalt, *Photochem. Photobiol.* **1994**, *60*, 225–230.
382. A. Matsumoto; P. C. Hanawalt, *Cancer Res.* **2000**, *60*, 3921–3926.
383. L. R. McGee; R. Misra, *J. Am. Chem. Soc.* **1990**, *112*, 2386–2389.
384. M. J. Peak; J. G. Peak; C. M. Blaumueller; R. K. Elespuru, *Chem. Biol. Interact.* **1988**, *67*, 267–274.
385. F. Tomita; K. Takahashi; T. Tamaoki, *J. Antibiot.* **1982**, *35*, 1038–1041.
386. L. E. Bockstahler; R. K. Elespuru; V. M. Hitchins; P. G. Carney; K. M. Olvey; C. D. Lytle, *Photochem. Photobiol.* **1990**, *51*, 477–479.
387. G. T. Carter; A. A. Fantini; J. C. James; D. B. Borders; R. J. White, *Tetrahedron Lett.* **1984**, *25*, 255–258.
388. G. T. Carter; A. A. Fantini; J. C. James; D. B. Borders; R. J. White, *J. Antibiot.* **1985**, *38*, 242–248.
389. K. Takahashi; F. Tomita, *J. Antibiot.* **1983**, *36*, 1531–1535.
390. M. Tsuda; H. Sato; Y. Tanaka; K. Yazawa; Y. Mikami; T. Sasaki; J. Kobayashi, *J. Chem. Soc. Perkin Trans. I* **1996**, 1773–1775.
391. M. K. Kharel; L. Zhu; T. Liu; J. Rohr, *J. Am. Chem. Soc.* **2007**, *129*, 3780–3781.
392. M. K. Kharel; P. Pahari; H. Lian; J. Rohr, *ChemBioChem* **2009**, *10*, 1305–1308.
393. T. Liu; C. Fischer; C. Beninga; J. Rohr, *J. Am. Chem. Soc.* **2004**, *126*, 12262–12263.
394. T. Liu; M. K. Kharel; C. Fischer; A. McCormick; J. Rohr, *ChemBioChem* **2006**, *7*, 1070–1077.
395. T. Liu; M. K. Kharel; L. Zhu; S. A. Bright; C. Mattingly; V. R. Adams; J. Rohr, *ChemBioChem* **2009**, *10*, 278–286.
396. S. W. Ayer; A. G. McInnes; P. Thibault; J. A. Walter; J. L. Doull; T. Parnell; L. C. Vining, *Tetrahedron Lett.* **1991**, *32*, 6301–6304.
397. J. L. Doull; S. W. Ayer; A. K. Singh; P. Thibault, *J. Antibiot.* **1993**, *46*, 869–871.
398. J. L. Doull; A. K. Singh; M. Hoare; S. W. Ayer, *J. Ind. Microbiol.* **1994**, *13*, 120–125.
399. L. Han; K. Yang; K. Kulowski; E. Wendt-Pienkowski; C. R. Hutchinson; L. C. Vining, *Microbiology* **2000**, *146*, 903–910.
400. K. Kulowski; E. Wendt-Pienkowski; L. Han; K. Q. Yang; L. C. Vining; C. R. Hutchinson, *J. Am. Chem. Soc.* **1999**, *121*, 1786–1794.
401. J. McVey, *The Downstream Genes for Jadomycin Biosynthesis in* Streptomyces venezuelae; Department of Biology, Dalhousie University: Halifax, NS, 1998.
402. L. Wang; R. L. White; L. C. Vining, *Microbiology* **2002**, *148*, 1091–1103.
403. L. R. Wang; J. McVey; L. C. Vining, *Microbiology* **2001**, *147*, 1535–1545.
404. K. Yang; L. Han; L. C. Vining, *J. Bacteriol.* **1995**, *177*, 6111–6117.
405. K. Yang; L. Han; S. W. Ayer; L. C. Vining, *Microbiology* **1996**, *142*, 123–132.
406. K. Q. Yang; L. Han; J. Y. He; L. R. Wang; L. C. Vining, *Gene* **2001**, *279*, 165–173.
407. Y. H. Chen; C. C. Wang; L. Greenwell; U. Rix; D. Hoffmeister; L. C. Vining; J. Rohr; K. Q. Yang, *J. Biol. Chem.* **2005**, *280*, 22508–22514.
408. U. Rix; C. Wang; Y. Chen; F. M. Lipata; L. L. Remsing Rix; L. M. Greenwell; L. C. Vining; K. Yang; J. Rohr, *ChemBiochem* **2005**, *6*, 838–845.
409. L. Wang; L. C. Vining, *Microbiology* **2003**, *149*, 1991–2004.
410. D. L. Jakeman; S. Farrell; W. Young; R. J. Doucet; S. C. Timmons, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1447–1449.
411. D. L. Jakeman; C. L. Graham; T. R. Reid, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 5280–5283.
412. U. Rix; J. Zheng; L. L. Remsing Rix; L. Greenwell; K. Yang; J. Rohr, *J. Am. Chem. Soc.* **2004**, *126*, 4496–4497.
413. C. N. Borissow; C. L. Graham; R. T. Syvitski; T. R. Reid; J. Blay; D. L. Jakeman, *ChemBioChem* **2007**, *8*, 1198–1203.
414. R. T. Syvitski; C. N. Borissow; C. L. Graham; D. L. Jakeman, *Org. Lett.* **2006**, *8*, 697–700.
415. M. C. Cone; P. J. Seaton; K. A. Halley; S. J. Gould, *J. Antibiot. (Tokyo)* **1989**, *42*, 179–188.
416. H. C. Lin; S. C. Chang; N. L. Wang; L. R. Chang, *J. Antibiot. (Tokyo)* **1994**, *47*, 675–680.
417. S. Omura; A. Nakagawa; H. Yamada; T. Hata; A. Furusaki, *Chem. Pharm. Bull. (Tokyo)* **1973**, *21*, 931–940.
418. W. Zeng; T. Eric Ballard; A. G. Tkachenko; V. A. Burns; D. L. Feldheim; C. Melander, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5148–5151.
419. S. J. Gould; N. Tamayo; C. R. Melville; M. C. Cone, *J. Am. Chem. Soc.* **1994**, *116*, 2207–2208.
420. S. J. Gould, *Chem. Rev.* **1997**, *97*, 2499–2509.
421. S. Mithani; G. Weeratunga; N. J. Taylor; G. I. Dmitrienko, *J. Am. Chem. Soc.* **1994**, *116*, 2209–2210.
422. X. H. Yin; B. Mahadevan; L. Grochowski; P. J. Proteau, Molecular cloning and sequence of the kinamycin angucycline type II polyketide synthase gene cluster from *Streptomyces murayamaensis.* Unpublished gene bank access no. AY228175.1, 2003.
423. S. J. Gould; C. R. Melville, *Bioorg. Med. Chem. Lett.* **1995**, *5*, 51–54.
424. S. J. Gould; C. R. Melville; M. C. Cone; J. Chen; J. R. Carney, *J. Org. Chem.* **1997**, *62*, 320–324.
425. J. R. Carney; S.-T. Hong; S. J. Gould, *Tetrahedron Lett* **1997**, *38*.
426. T. Aoyama; W. Zhao; F. Kojima; Y. Muraoka; H. Naganawa; T. Takeuchi; T. Aoyagi, *J. Antibiot.* **1993**, *46*, 1471–1474.
427. C. Volkmann; E. Rossner; M. Metzler; H. Zahner; A. Zeeck, **1995**.
428. J. J. Young; S.-N. Ho; W.-M. Ju; L.-R. Chnag, *J. Antibiot.* **1994**, *47*, 681–687.
429. H. Drautz; P. Reuschenbach; H. Zähner; J. Rohr; A. Zeeck, *J. Antibiot.* **1985**, *38*, 1291–1301.
430. E. Egert; M. Noltemeyer; J. Siebers; J. Rohr; A. Zeeck, *J. Antibiot.* **1992**, *45*, 1190–1192.
431. J. Rohr; S. Eick; A. Zeeck; P. Reuschenbach; H. Zahner; H. P. Fiedler, *J. Antibiot.* **1988**, *41*, 1066–1073.
432. J. Rohr; A. Zeeck, *J. Antibiot.* **1990**, *43*, 1169–1178.
433. W. Weber; H. Zähner; J. Siebers; K. Schroder; A. Zeeck, *Arch. Microbiol.* **1979**, *121*, 111–116.
434. M. J. Bibb; S. Biro; H. Motamedi; J. F. Collins; C. R. Hutchinson, *EMBO J.* **1989**, *8*, 2727–2736.
435. H. Decker; C. R. Hutchinson, *J. Bacteriol.* **1993**, *175*, 3887–3892.
436. H. Decker; H. Motamedi; C. R. Hutchinson, *J. Bacteriol.* **1993**, *175*, 3876–3886.

437. H. Decker; R. G. Summers; C. R. Hutchinson, *J. Antibiot.* **1994**, *47*, 54–63.
438. H. Decker; J. Rohr; H. Motamedi; H. Zähner; C. R. Hutchinson, *Gene* **1995**, *166*, 121–126.
439. H. C. Gramajo; J. White; C. R. Hutchinson; M. J. Bibb, *J. Bacteriol.* **1991**, *173*, 6475–6483.
440. P. G. Guilfoile; C. R. Hutchinson, *J. Bacteriol.* **1992**, *174*, 3651–3658.
441. P. G. Guilfoile; C. R. Hutchinson, *J. Bacteriol.* **1992**, *174*, 3659–3666.
442. C. R. Hutchinson; I. Fujii, *Annu. Rev. Microbiol.* **1995**, *49*, 201–238.
443. G. Meurer; C. R. Hutchinson, *J. Bacteriol.* **1995**, *177*, 477–481.
444. H. Motamedi; E. Wendt-Pienkowski; C. R. Hutchinson, *J. Bacteriol.* **1986**, *167*, 575–580.
445. A. Ramos; F. Lombo; A. F. Brana; J. Rohr; C. Mendez; J. A. Salas, *Microbiology* **2008**, *154*, 781–788.
446. B. Shen; R. G. Summers; H. Gramajo; M. J. Bibb; C. R. Hutchinson, *J. Bacteriol.* **1992**, *174*, 3818–3821.
447. B. Shen; C. R. Hutchinson, *Science* **1993**, *262*, 1535–1540.
448. B. Shen; H. Nakayama; C. R. Hutchinson, *J. Nat. Prod.* **1993**, *56*, 1288–1293.
449. B. Shen; R. G. Summers; E. Wendt-Pienkowski; C. R. Hutchinson, *J. Am. Chem. Soc.* **1995**, *117*, 6811–6821.
450. R. G. Summers; E. Wendt-Pienkowski; H. Motamedi; C. R. Hutchinson, *J. Bacteriol.* **1992**, *174*, 1810–1820.
451. R. G. Summers; E. Wendt-Pienkowski; H. Motamedi; C. R. Hutchinson, *J. Bacteriol.* **1993**, *175*, 7571–7580.
452. R. G. Summers; A. Ali; B. Shen; W. A. Wessel; C. R. Hutchinson, *Biochemistry* **1995**, *34*, 9389–9402.
453. T. B. Thompson; K. Katayama; K. Watanabe; C. R. Hutchinson; I. Rayment, *J. Biol. Chem.* **2004**, *279*, 37956–37963.
454. S. Yue; H. Motamedi; E. Wendt-Pienkowski; C. R. Hutchinson, *J. Bacteriol.* **1986**, *167*, 581–586.
455. P. J. Kramer; R. J. X. Zawada; R. McDaniel; C. R. Hutchinson; D. A. Hopwood; C. Khosla, *J. Am. Chem. Soc.* **1997**, *119*, 635–639.
456. R. McDaniel; C. R. Hutchinson; C. Khosla, *J. Am. Chem. Soc.* **1995**, *117*, 6805–6810.
457. B. Shen; C. R. Hutchinson, *Biochemistry* **1993**, *32*, 6656–6663.
458. E. R. Rafanan, Jr.; C. R. Hutchinson; B. Shen, *Org. Lett.* **2000**, *2*, 3225–3227.
459. B. Shen; C. R. Hutchinson, *J. Biol. Chem.* **1994**, *269*, 30726–30733.
460. H. Decker; S. Haag; G. Udvarnoki; J. Rohr, *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 1107–1110.
461. C. Fischer; L. Rodríguez; E. P. Patallo; F. Lipata; A. F. Braña; C. Méndez; J. A. Salas; J. Rohr, *J. Nat. Prod.* **2002**, 1685–1689.
462. F. Lombó; M. Gibson; L. Greenwell; A. F. Braña; J. Rohr; J. A. Salas; C. Méndez, *Chem. Biol.* **2004**, *11*, 1709–1718.
463. M. Pérez; F. Lombó; L. Zhu; M. Gibson; A. F. Braña; J. Rohr; J. A. Salas; C. Méndez, *Chem. Commun. (Camb)* **2005**, 1604–1606.
464. L. Rodriguez; C. Oelkers; I. Aguirrezabalaga; A. F. Braña; J. Rohr; C. Méndez; J. A. Salas, *J. Mol. Microbiol. Biotechnol.* **2000**, *2*, 271–276.
465. B. J. Rawlings, *Nat. Prod. Rep.* **1997**, *14*, 523–557.
466. H. Brockmann; T. Reschke, *Tetrahedron Lett.* **1968**, *27*, 3167–3170.
467. B. E. Roggo; F. Petersen; R. Delmendo; H. B. Jenny; H. H. Peter; J. Roesel, *J. Antibiot.* **1994**, *47*, 136–142.
468. I. Haupt; U. Waehnert; C. Pitra; G. Loeber; G. Luck; K. Eckardt, *Z. Allg. Mikrobiol.* **1975**, *15*, 411–421.
469. Y. Shiono; N. Shiono; S. Seo; S. Oka; Y. Yamazaki, *Z. Naturforsch.* **2002**, *57c*, 923–929.
470. G. Höfle; H. Wolf, *Liebigs Ann. Chem.* **1983**, 835–843.
471. K. Ishida; K. Fritzsche; C. Hertweck, *J. Am. Chem. Soc.* **2007**, *129*, 12648–12649.
472. A. Schenk; Z. Xu; C. Pfeiffer; C. Steinbeck; C. Hertweck, *Angew. Chem. Int. Ed.* **2007**, *46*, 7035–7038.
473. K. Ishida; K. Maksimenka; K. Fritzsche; K. Scherlach; G. Bringmann; C. Hertweck, *J. Am. Chem. Soc.* **2006**, *128*, 14619–14624.
474. K. Eckardt; G. Bradler; D. Tresselt; H. Fritzsche, *Adv. Antimicrob. Antineoplast. Chemother.* **1972**, *1*, 1025–1027.
475. M. Tsunakawa; M. Nishio; H. Ohkuma; T. Tsuno; M. Konishi; T. Naito; T. Oki; H. Kawaguchi, *J. Org. Chem.* **1989**, *54*, 2532–2536.
476. T. Aoyagi; T. Aoyama; F. Kojima; N. Matsuda; M. Maruyama; M. Hamada; T. Takeuchi, *J. Antibiot.* **1992**, *45*, 1385–1390.
477. T. Aoyama; F. Kojima; F. Abe; Y. Muraoka; H. Naganawa; T. Takeuchi; T. Aoyagi, *J. Antibiot.* **1993**, *46*, 914–920.
478. T. Aoyama; F. Kojima; T. Yamazaki; T. Tatee; F. Abe; Y. Muraoka; H. Naganawa; T. Aoyagi; T. Takeuchi, *J. Antibiot.* **1993**, *46*, 712–718.
479. J. Zhan; K. Watanabe; Y. Tang, *ChemBioChem* **2008**, *9*, 1710–1715.
480. Z. Yunt; K. Reinhardt; A. Li; M. Engeser; H. M. Dahse; M. Gutschow; T. Bruhn; G. Bringmann; J. Piel, *J. Am. Chem. Soc.* **2009**, *131*, 2297–2305.
481. B. Sontag; J. G. Muller; F. G. Hansske, *J. Antibiot.* **2004**, *57*, 823–828.
482. Y. Chen; Y. Luo; J. Ju; E. Wendt-Pienkowski; S. R. Rajski; B. Shen, *J. Nat. Prod.* **2008**, *71*, 431–437.
483. E. A. Owen; G. A. Burley; J. A. Carver; G. Wickham; M. A. Keniry, *Biochem. Biophys. Res. Commun.* **2002**, *290*, 1602–1608.
484. H. G. Floss, *J. Biotechnol.* **2006**, *124*, 242–257.
485. P. Beltran-Alvarez; R. J. Cox; J. Crosby; T. J. Simpson, *Biochemistry* **2007**, *46*, 14672–14681.
486. D. A. Hopwood, *Chem. Rev.* **1997**, *97*, 2465–2497.
487. J. A. Kalaitzis; B. S. Moore, *J. Nat. Prod.* **2004**, *67*, 1419–1422.
488. W. Zhang; B. I. Wilke; J. Zhan; K. Watanabe; C. N. Boddy; Y. Tang, *J. Am. Chem. Soc.* **2007**, *129*, 9304–9305.
489. C. L. Freel Meyers; M. Oberthur; L. Heide; D. Kahne; C. T. Walsh, *Biochemistry* **2004**, *43*.
490. A. Freitag; C. Mendez; J. A. Salas; B. Kammerer; S. M. Li; L. Heide, *Metab. Eng.* **2006**, *8*, 653–661.
491. J. M. Langenhan; B. R. Griffith; J. S. Thorson, *J. Nat. Prod.* **2005**, *68*, 1696–1711.
492. C. Méndez; J. A. Salas, *Trends Biotechnol.* **2001**, *19*, 449–456.
493. J. A. Salas; C. Mendez, *Trends Microbiol.* **2007**, *15*, 219–232.
494. S. C. Timmons; J. S. Thorson, *Curr. Opin. Chem. Biol.* **2008**, *12*, 297–305.
495. G. J. Williams; R. Gantt; J. S. Thorson, *Curr. Opin. Chem. Biol.* **2008**.
496. H. Xu; R. Kahlich; B. Kammerer; L. Heide; S.-M. Li, *Microbiology* **2003**, *149*, 2183–2191.
497. R. McDaniel; S. Ebert-Khosla; D. A. Hopwood; C. Khosla, *Science* **1993**, *262*, 1546–1550.
498. R. McDaniel; P. Licari; C. Khosla, *Adv. Biochem. Eng. Biotechnol.* **2001**, *73*, 31–52.
499. C. R. Hutchinson, *Curr. Opin. Microbiol.* **1998**, *1*, 319–329.
500. C. Khosla; R. J. X. Zawada, *Trends Biotechnol.* **1996**, *14*, 335–341.

501. C. Khosla, *Metab. Eng.* **1999**, *24*, 203–225.
502. J. Rohr, *Angew. Chem. Int. Ed.* **1995**, *34*, 881–885.
503. W. R. Strohl, *Metab. Eng.* **2001**, *3*, 4–14.
504. C. J. Tsoi; C. Khosla, *Chem. Biol.* **1995**, *2*, 355–362.
505. C. J. Arthur; A. E. Szafranska; J. Long; J. Mills; R. J. Cox; S. C. Findlow; T. J. Simpson; M. P. Crump; J. Crosby, *Chem. Biol.* **2006**, *13*, 587–596.
506. J. M. Crawford; P. M. Thomas; J. R. Scheerer; A. L. Vagstad; N. L. Kelleher; C. A. Townsend, *Science* **2008**, *320*, 243–246.
507. J. M. Crawford; A. L. Vagstad; K. C. Ehrlich; D. W. Udwary; C. A. Townsend, *ChemBioChem* **2008**, *9*, 1559–1563.
508. J. M. Crawford; A. L. Vagstad; K. P. Whitworth; K. C. Ehrlich; C. A. Townsend, *ChemBioChem* **2008**, *9*, 1019–1023.
509. J. Dreier; Q. Li; C. Khosla, *Biochemistry* **2001**, *40*, 12407–12411.
510. A. T. Hadfield; C. Limpkin; W. Teartasin; T. J. Simpson; J. Crosby; M. P. Crump, *Structure* **2004**, *12*, 1865–1875.
511. A. T. Keatinge-Clay; A. A. Shelat; D. F. Savage; S. C. Tsai; L. J. Miercke; J. D. O'Connell, 3rd; C. Khosla; R. M. Stroud, *Structure* **2003**, *11*, 147–154.
512. T. P. Korman; J. A. Hill; T. N. Vu; S. C. Tsai, *Biochemistry* **2004**, *43*, 14529–14538.
513. T. P. Korman; Y. H. Tan; J. Wong; R. Luo; S. C. Tsai, *Biochemistry* **2008**, *47*, 1837–1847.
514. T. W. Yu; Y. M. Shen; R. McDaniel; H. G. Floss; C. Khosla; D. A. Hopwood; B. S. Moore, *J. Am. Chem. Soc.* **1998**, *120*, 7749–7759.
515. C. O. Rock; S. Jackowski, *Biochem. Biophys. Res. Commun.* **2002**, *292*, 1155–1166.
516. J. R. Jacobsen; C. R. Hutchinson; D. E. Cane; C. Khosla, *Science* **1997**, *277*, 367–369.
517. K. Kinoshita; P. G. Williard; C. Khosla; D. E. Cane, *J. Am. Chem. Soc.* **2001**, *123*, 2495–2502.
518. H. Bockholt; G. Udvarnoki; J. Rohr; U. Mocek; J. M. Beale; H. G. Floss, *J. Org. Chem.* **1994**, *59*, 2064–2069.
519. G. Lackner; A. Schenk; Z. Xu; K. Reinhardt; Z. S. Yunt; J. Piel; C. Hertweck, *J. Am. Chem. Soc.* **2007**, *129*, 9306–9312.
520. Z. Xu; A. Schenk; C. Hertweck, *J. Am. Chem. Soc.* **2007**, *129*, 6022–6030.
521. J. A. Salas, *Chem. Biol.* **2004**, *11*, 892–894.
522. S. M. Ma; J. Zhan; K. Watanabe; X. Xie; W. Zhang; C. C. Wang; Y. Tang, *J. Am. Chem. Soc.* **2007**.
523. R. E. Minto; C. A. Townsend, *Chem. Rev.* **1997**, *97*, 2537–2555.
524. S. M. Ma; J. Zhan; X. Xie; K. Watanabe; Y. Tang; W. Zhang, *J. Am. Chem. Soc.* **2008**, *130*, 38–39.
525. W. Zhang; Y. Li; Y. Tang, *Proc. Natl. Acad. Sci. U.S.A.* **2008**.
526. R. Muller, *Chem. Biol.* **2004**, *11*, 4–6.
527. A. Luzhetskyy; A. Bechthold, *Mol. Microbiol.* **2005**, *58*, 3–5.
528. M. Pérez; F. Lombó; I. Baig; A. F. Braña; J. Rohr; J. A. Salas; C. Méndez, *Appl. Environ. Microbiol.* **2006**, *72*, 6644–6652.
529. J. S. Thorson; T. J. Hosted; J. Jiang; J. B. Biggins; J. Ahlert, *Curr. Org. Chem.* **2001**, *5*, 139–167.
530. A. Trefzer; J. A. Salas; A. Bechthold, *Nat. Prod. Rep.* **1999**, *16*, 283–299.
531. L. Rodriguez; I. Aguirrezabalaga; N. Allende; A. F. Braña; C. Méndez; J. A. Salas, *Chem. Biol.* **2002**, *9*, 721–729.
532. C. Olano; N. Lomovskaya; L. Fonstein; J. T. Roll; C. R. Hutchinson, *Chem. Biol.* **1999**, *6*, 845–855.
533. A. Luzhetskyy; A. Mayer; J. Hoffmann; S. Pelzer; M. Holzenkamper; B. Schmitt; S. E. Wohlert; A. Vente; A. Bechthold, *ChemBioChem* **2007**, *8*, 599–602.
534. A. Luzhetskyy; J. Hoffmann; S. Pelzer; S. E. Wohlert; A. Vente; A. Bechthold, *Appl. Microbiol. Biotechnol.* **2008**, *80*, 15–19.
535. I. Baig; M. Pérez; A. F. Braña; R. Gomathinayagam; C. Damodaran; J. A. Salas; C. Méndez; J. Rohr, *J. Nat. Prod.* **2008**, *71*, 199–207.
536. M. Pérez; I. Baig; A. F. Braña; J. A. Salas; J. Rohr; C. Méndez, *ChemBioChem* **2008**, *9*, 2295–2304.
537. D. Hoffmeister; B. Wilkinson; G. Foster; P. J. Sidebottom; K. Ichinose; A. Bechthold, *Chem. Biol.* **2002**, *9*, 287–295.
538. C. Zhang; B. R. Griffith; Q. Fu; C. Albermann; X. Fu; I.-K. Lee; L. Li; J. S. Thorson, *Science* **2006**, *313*, 1291–1294.
539. B. R. Griffith; J. M. Langenhan; J. S. Thorson, *Curr. Opin. Biotechnol.* **2005**, *16*, 622–630.
540. P. G. Hultin, *Curr. Top. Med. Chem.* **2005**, *5*, 1299–1331.
541. J. Yang; D. Hoffmeister; L. Liu; X. Fu; J. S. Thorson, *Bioorg. Med. Chem.* **2004**, *12*, 1577–1584.
542. J. Beynon; E. R. Rafanan, Jr.; B. Shen; A. J. Fisher, *Acta Crystallogr. D Biol. Crystallogr.* **2000**, *56*, 1647–1651.
543. S. J. Gould; X. C. Cheng, *J. Org. Chem.* **1994**, *59*, 400–405.
544. J. T. Zheng; U. Rix; L. Zhao; C. Mattingly; V. Adams; Q. Chen; J. Rohr; K. Q. Yang, *J. Antibiot. (Tokyo)* **2005**, *58*, 405–408.
545. J. T. Zheng; S. L. Wang; K. Q. Yang, *Appl. Microbiol. Biotechnol.* **2007**, *76*, 883–888.

Biographical Sketches

Jürgen Rohr was born in Hameln, Germany, and graduated from the University of Göttingen, Germany, in 1984 with a Ph.D. based on structure elucidation and derivatization of novel natural anticancer drugs. After a postdoctoral stay in Heinz Floss' laboratory in Columbus, OH, USA, he returned to Göttingen, Germany as an assistant professor (habilitation), and became an associate professor of Organic Chemistry in 1992. In 1997, he was appointed associate professor of Medicinal Chemistry at the Medical University of South Carolina in Charleston, SC, USA, and in 2002 a Full Professor at the College of Pharmacy of the University of Kentucky, Lexington, KY, USA. Since 2007, Dr. Rohr is the Director of the Division of Drug Discovery of the University of Kentucky's College of Pharmacy. Dr. Rohr's research focuses on natural product drugs, that is, antibiotics, anticancer drugs, drugs against bone diseases, and herbal drugs. It includes elucidation of complex multistep biosynthetic pathways carried out by bacteria, fungi, or plants, with particular emphasis on biosynthetic enzyme mechanisms. The results of these studies are used to generate modified natural product drugs through genetic engineering (pathway engineering and combinatorial biosynthesis).

Christian Hertweck (born 1969) studied chemistry at the University of Bonn and completed his Ph.D. under the supervision of Professor Boland at the Max Planck Institute for Chemical Ecology. In 1999, he became a Humboldt postdoctoral fellow of Professors Floss and Moore at the University of Washington, SA, USA. He then set up an independent research group at the HKI in Jena. Since 2006, he has held a chair of natural product chemistry at the Friedrich Schiller University Jena and is the Head of the Department of Biomolecular Chemistry at the Leibniz Institute for Natural Product Research and Infection Biology. His current research involves natural products from bacteria and fungi with emphasis on the investigation and manipulation of biosynthetic pathways.

1.08 Structural Enzymology of Polyketide Synthase: The Structure–Sequence–Function Correlation

Tyler Paz Korman, Brian Ames, and Shiou-Chuan (Sheryl) Tsai, University of California, Irvine, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

1.08.1	**Introduction**	305
1.08.2	**Fatty Acid Biosynthesis and Fatty Acid Synthase**	307
1.08.2.1	The Porcine Type I FAS Structure	308
1.08.3	**The Type II PKS**	310
1.08.3.1	Similarities and Differences between FAS and PKS	314
1.08.4	**Structural Enzymology of Individual Type II PKS Enzymes**	314
1.08.4.1	Malonyl-Coenzyme A:ACP transacetylase	314
1.08.4.1.1	Crystal structure of MAT	316
1.08.4.1.2	Proposed mechanism of MAT	317
1.08.4.1.3	Proposed molecular basis of substrate specificity of MAT	317
1.08.4.2	The ZhuH Priming Ketosynthase	318
1.08.4.2.1	Crystal structure of ZhuH	318
1.08.4.2.2	Proposed mechanism of ZhuH	320
1.08.4.2.3	Proposed molecular basis of substrate specificity of ZhuH	320
1.08.4.3	The Elongating Ketosynthase/Chain Length Factor Complex	321
1.08.4.3.1	Crystal structure of the actinorhodin KS/CLF	322
1.08.4.3.2	Proposed mechanism of KS/CLF	322
1.08.4.3.3	Proposed molecular basis of substrate specificity of KS/CLF	323
1.08.4.4	The Ketoreductase	323
1.08.4.4.1	Crystal structure of the actinorhodin ketoreductase	325
1.08.4.4.2	Proposed mechanism of ketoreductase	325
1.08.4.4.3	Proposed molecular basis of ketoreductase substrate specificity	327
1.08.4.4.4	The stereochemistry of ketoreductase	327
1.08.4.5	The Aromatase/Cyclase	329
1.08.4.5.1	Crystal structure of Tcm ARO/CYC	332
1.08.4.5.2	Proposed mechanism of Tcm ARO/CYC	334
1.08.4.5.3	The cyclization specificity of ARO/CYC	336
1.08.4.6	The Fourth Ring Cyclases	336
1.08.4.7	Acyl Carrier Protein	338
1.08.5	**Protein–Protein Interactions and Transport of Polyketide Intermediates between Enzymes**	339
1.08.6	**Summary and Future Prospects**	340
References		342

1.08.1 Introduction

Polyketides are one of the largest classes of natural products that were first identified by acetate labeling experiments, demonstrating their close relationship to fatty acid biosynthesis.[1,2] Polyketides display intricate chemical structures and a wide range of pharmaceutical activities contributing to their striking tendency to be developed into therapeutics (>0.3% of current known polyketides) compared to other natural product or synthetic compounds (<0.1%)[3] and highlighting their potential for drug discovery. Current examples of polyketide natural products include antibiotics (such as erythromycin,[4] actinorhodin (act),[5] and

tetracenomycin (Tcm)[6]), chemotherapeutics (such as doxorubicin,[7] resistomycin,[8] and mithramycin[9]), cholesterol-lowering drugs (such as lovastatin[10]), antioxidants (such as resveratrol[11]), as well as toxins (T-toxin[12]) and carcinogens (aflatoxin B1[13]) (**Figure 1**). Such an impact on health care, in conjunction with growing academic interest, underscores the significance of polyketide biosynthesis.

At the center of polyketide biosynthesis is the polyketide synthase (PKS). PKS is a multidomain or multiprotein enzyme complex that produces a huge variety of polyketides via a controlled variation of building blocks and modification reactions such as chain reduction and cyclization and is structurally and functionally related to fatty acid synthase (FAS). In spite of the widely appreciated importance of polyketides and the vigorous efforts to identify genes of PKS,[14] there is still a critical gap in the knowledge base about how different PKSs biosynthesize different polyketide products. The structures and enzymatic studies of PKS, as multidomain complexes as well as individual proteins, will help elucidate the molecular basis of PKS substrate specificity during chain elongation, reduction, and cyclization.

There are at least four distinct types of PKS: modular type I, iterative type I, type II, and type III that can be distinguished by their architecture (**Figure 2**).[15] The modular type I PKSs are huge, multimodular complexes involved in the biosynthesis of macrolides such as erythromycin,[16] with each module containing multiple domains that are typically used only once in an 'assembly line' fashion. In contrast, type II PKSs consist of stand-alone enzymes that are used iteratively to biosynthesize aromatic polyketides. Finally, type III PKSs from plants consist of a single protein that catalyzes a finite number of chain elongations using coenzyme A (CoA)-bound substrates. In the past decade, the structures of type II PKS enzymes responsible for chain

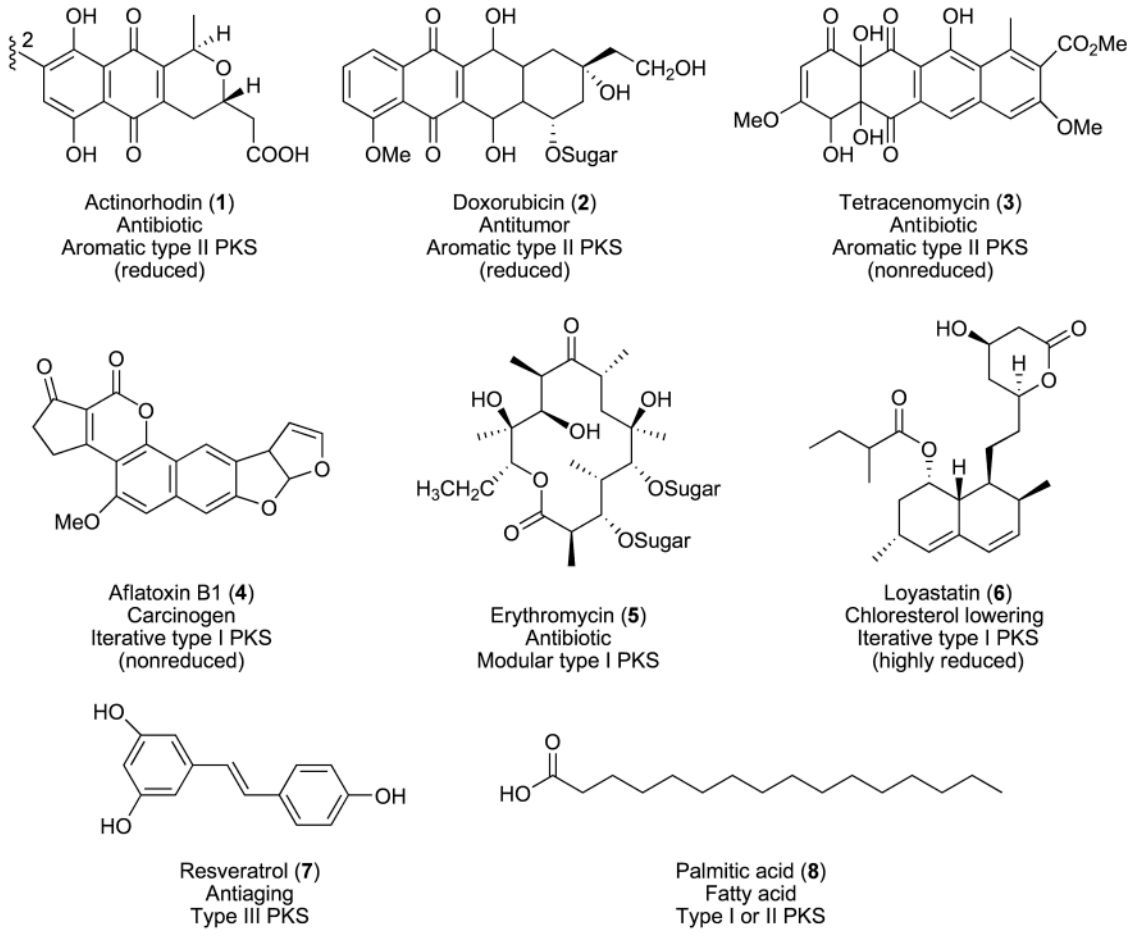


Figure 1 Structural diversity of polyketide natural products (1–7) versus palmitic acid (8).

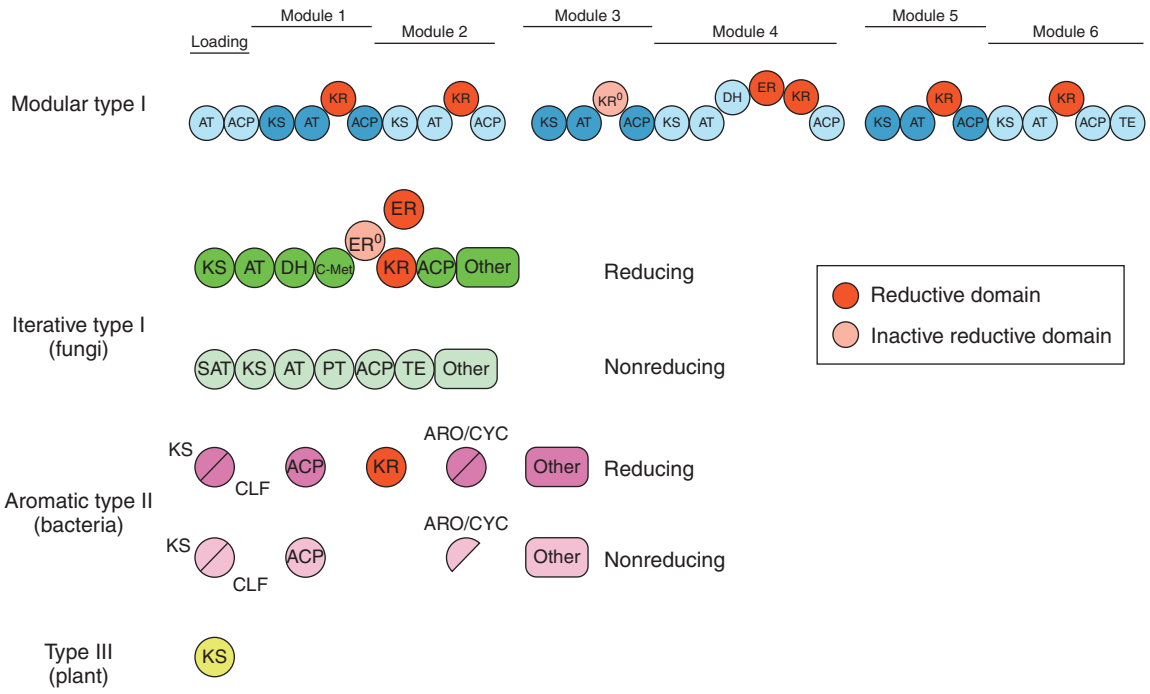


Figure 2 Different domain organization for modular type I, iterative type I, aromatic type II, and type III PKSs. The modular type I PKS is a large multimodular complex where the β-carbonyl processing can be read directly from the linear order of domains (linear logic). In contrast, iterative type I and aromatic type II PKS use an iterative process (i.e., each active site used multiple times) resulting in polyketides containing cryptic β processing profiles that can only be inferred from the resulting polyketide. Both the iterative type I and aromatic type II PKSs can further be broken down into reducing and nonreducing types, depending on the presence or absence of KR domains.

elongation, reduction, and cyclization, in combination with *in vitro* and *in vivo* functional studies, have allowed researchers to correlate the three-dimensional arrangement of PKS active sites with the observed substrate specificity and, in some cases, to elucidate the sequence–function–structure relationship. Because of the availability of type I and III PKS reviews elsewhere,[17,18] and the relative rarity of iterative type I PKS structural studies, a major part of this chapter will focus on reviewing the structural enzymology of type II PKS done in the past decade. However, due to the similarities between PKS and FAS and the availability of a high-resolution crystal for the entire mammalian FAS complex,[19] a brief description of the FAS structure will be presented first.

1.08.2 Fatty Acid Biosynthesis and Fatty Acid Synthase

Fatty acids are fundamental building blocks of life and are required for the survival of all living organisms. The fatty acid pathway has been validated as an excellent drug target against different diseases due to its importance in obesity, diabetes,[20] cancer,[21] as well as bacterial survival and pathogenesis.[22,23] Furthermore, fatty acid biosynthesis is catalyzed by a multidomain protein complex called FAS that is structurally and functionally related to PKS. FAS biosynthesizes fatty acids via the decarboxylative condensation between two-carbon building blocks (coined 'ketide' units) to produce a long-chain aliphatic fatty acid that is typically 16–18 carbons in length.[24]

The FAS consists of seven conserved proteins/domains that catalyze at least 50 reactions en route to the final product (**Figure 3**). First, the malonyl-CoA:ACP transacylase (MAT) primes the acyl carrier protein (ACP) by transferring malonate from malonyl-CoA to the phosphopantetheine (PPT) group of ACP. Next, the ketosynthase (KS) catalyzes chain elongation via iterative rounds of decarboxylative condensation between

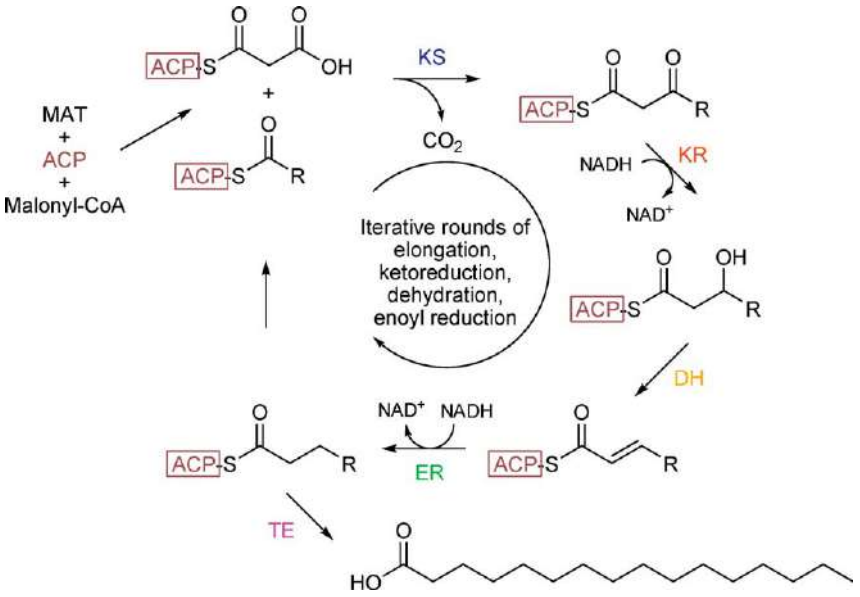


Figure 3 Fatty acid biosynthesis is an iterative process that proceeds via decarboxylative condensations between malonyl-ACP and acyl-ACP catalyzed by the KS domain. Reductions and dehydration at each cycle produce an aliphatic fatty acid released by the TE domain.

the KS-bound acyl group and malonyl-ACP. Following each elongation step, the ACP-bound chain is further modified by the nicotinamide adenine dinucleotide phosphate (NADP)H-dependent ketoreductase (KR), the dehydratase (DH), and the NADPH-dependent enoyl reductase (ER) to produce a fully reduced aliphatic chain. The final step to release the fatty acid product is catalyzed by the thioesterase (TE), which hydrolyzes the thioester bond between the fatty acyl chain and ACP. Many of these steps, specifically those catalyzed by MAT, KS, and KR, are chemically equivalent in FAS and PKS and are catalyzed by homologous enzymes.

Due to the importance of fatty acids, the FAS proteins/domains are highly conserved structurally and functionally from mammals to fungi to bacteria. Based on protein architecture, the FAS, similar to PKS, is classified as type I or type II (**Figure 4(a)**). In the type I FAS found in mammals, fungi, and some bacteria, each FAS catalytic domain is covalently linked as a single polypeptide with each active site used iteratively during fatty acid biosynthesis. In contrast, in the type II FAS found mainly in plants and bacteria, each FAS catalytic domain is a stand-alone enzyme. FAS crystal structures of isolated proteins or domains have been solved, including the TE domain cloned from type I FAS,[25] as well as the *Escherichia coli* type II KS, MAT, KR, and DH proteins.[26–29] One recent review that compares type I FAS and PKS offers a summary of the FAS structural work.[17] These isolated FAS protein structures shed light on the molecular details of individual steps and provide a framework for understanding the steps in PKS. However, in the absence of a larger PKS complex structure, mechanistic details about protein–protein interactions are unavailable. Fortunately, the full-length mammalian-type I FAS solved to 3.2 Å represents the first step toward elucidating the complex issues involved in fatty acid biosynthesis[24] with direct implications for understanding polyketide biosynthesis.

1.08.2.1 The Porcine Type I FAS Structure

The mammalian type I FAS is a 540 kDa homodimer that contains two fully active monomers.[24] Each 270 kDa monomer consists of KS, acyltransferase (AT), DH, ER, KR, ACP, and TE domains (**Figure 4(a)**). Although the resolution of the initial full-length structure was too low (5 Å) to allow the modeling of side chains into the electron density, a more recent 3.2 Å structure allowed five individual catalytic domains KS, AT, KR, DH, and ER to be placed, identified two nonenzymatic domains (ψME (methyltransferase) and ψKR), provided structural information about interdomain protein–protein interactions, and developed a molecular basis for substrate shuttling between domains by ACP.[19] The structures confirm electron microscopy[30] and biochemical

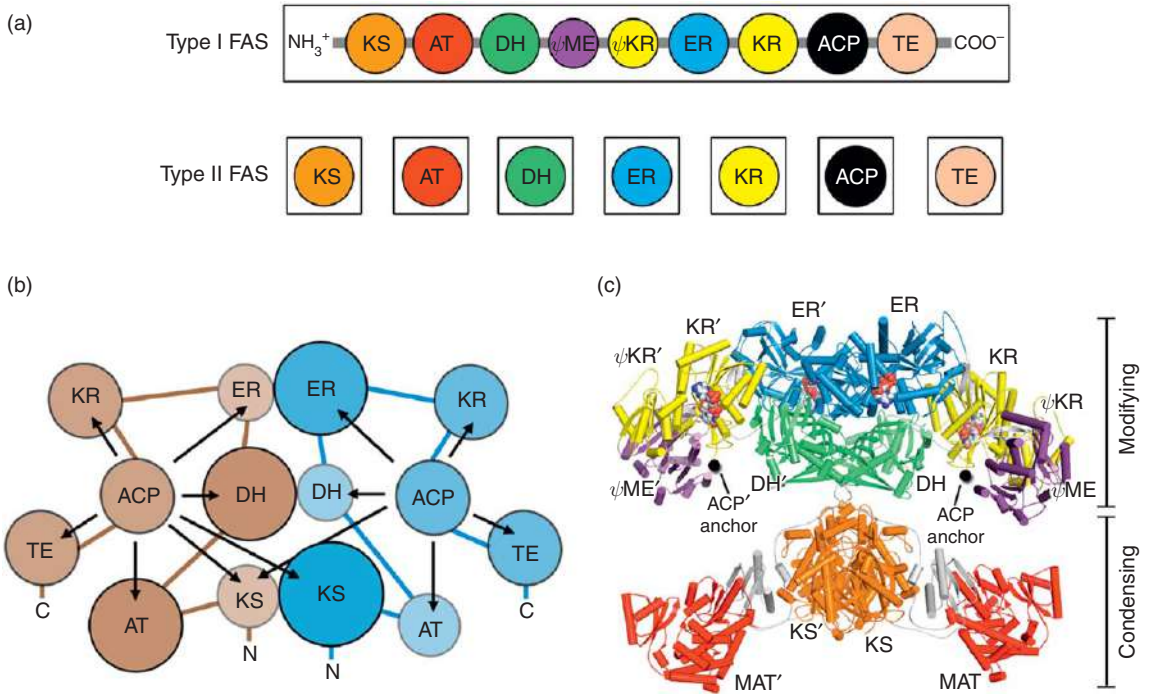


Figure 4 Structural overview of FAS. (a) Domain organization of type I and type II FAS. (b) The head-to-head model for type I FAS complex formation. Arrows highlight the ability of ACP to interact, not only with all domains from its own monomer but also with the KS from the other monomer. (c) The structure of mammalian type I FAS solved to 3.2 Å ($NADP^+$ are in spheres).

results[31,32] that suggest that the type I FAS adopts an 'X-like' conformation in a head-to-head orientation where the N-terminal KS domains interact to form part of the core (**Figures 4(b)** and **4(c)**). This unique 'X-like head-to-head' organization of the two FAS monomers provides a rationale for the formation of two distinct groupings of active sites where the condensing domains (MAT, KS) are located on the lower half, while the modifying domains (KR, DH, ER) are located in the upper half, separating the two events. Analysis of the structure shows that the central core is composed of the KS, DH, and ER domains, and the dimer is held together by KS–KS, DH–DH, and KS–DH interactions. Interestingly, it is apparent in the high-resolution FAS structure that less than 10% of the total sequence is devoted to linkers between domains, showing that interactions between the catalytic domains themselves are responsible for a major part of the complex. This is important if we are to assume that the individual proteins of the type II FAS form a complex similar to the type I FAS. Furthermore, this modest 'head-to-head' arrangement also provides flexibility that allows the ACP to interact with the KS from both monomers, consistent with previous biochemical studies (**Figure 4(b)**).[32,33] The DH–KS interaction, which links the top and the bottom of the complex, is mediated by a short (~20 amino acid) portion of the linker between KS and MAT, which is also conserved in the modular type I PKS KS–AT structure. The actual interacting surface between KS and DH is a modest 230 $Å^2$, and the authors suggest that the relatively weak interaction between KS and DH may allow the top portion to rotate and tilt, enabling ACP to access both reaction chambers and more efficient catalysis. The MAT and KR domains branch out as 'legs' and 'arms' from the bottom and top of the protein, respectively, to form two asymmetric reaction chambers on either side of the central core. In the high-resolution structure, the KR domain (which forms a heterodimer with an inactive pseudo-KR that mimics half of the KR) is responsible for bridging interactions between the DH and the ER domains. The flexible regions in the full-length protein, particularly the central connection region between KS and DH, allow each reaction chamber to pulse open and close during the reaction cycle, permitting the ACP-bound acyl group access to each domain. Significantly, it has been shown that the mammalian FAS shares more sequence and structural similarities with bacterial FAS and PKSs than with fungal FASs, suggesting that bacterial FASs and PKSs (both types I and II) will adopt similar architectures. Although invisible in the current

structure, ACP and TE, which are attached at the C-terminus following the KR domain, should presumably be located near the center of each reaction chamber (**Figures 4(b)** and **4(c)**). A central location for ACP is essential, because it must move around in order to present each domain with the acyl substrate. Although there is no concrete proof, it is hypothesized that the type II FAS will adopt a similar complex as the type I FAS due to sequence conservation of individual domains.[17] An excellent review of the type II FAS structural enzymology work can be found elsewhere.[34]

1.08.3 The Type II PKS

Similar to the type II FAS, the type II PKS carries out a series of reactions catalyzed by individual soluble enzymes, each of which are encoded by a discrete gene and show considerable sequence conservation (**Figures 5** and **6**). The type II PKS is found predominantly in *Streptomyces* and produces aromatic

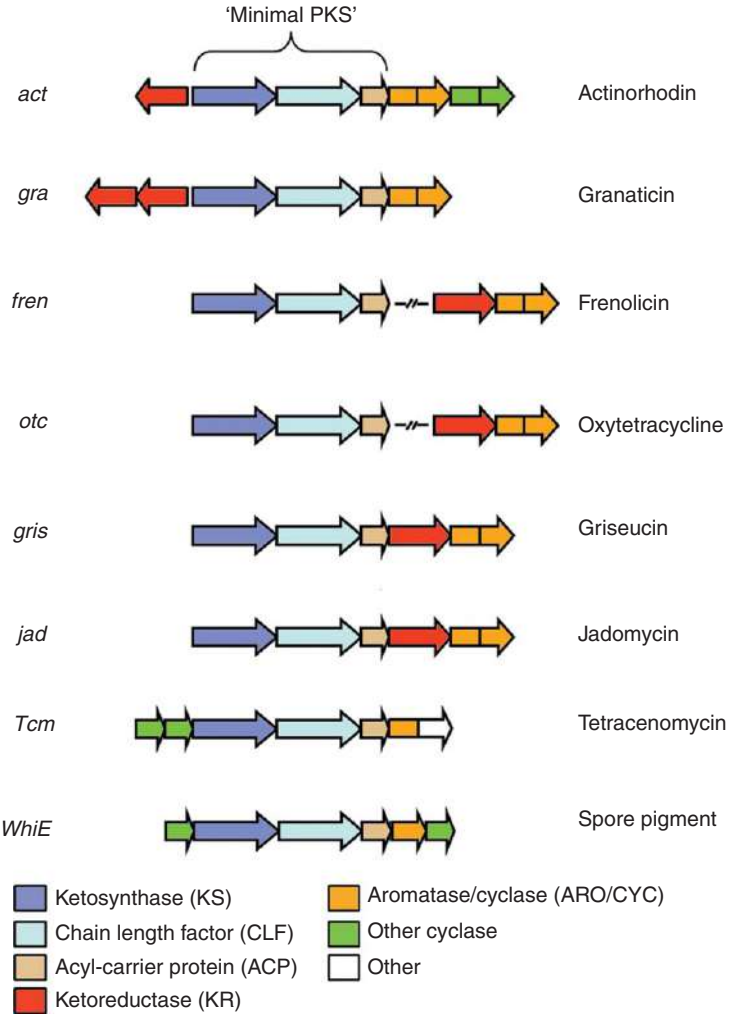


Figure 5 Organization of gene clusters in type II PKSs. The 'minimal PKS' corresponds to the minimal protein machinery required for assembling the polyketide chain, including KS/CLF and ACP. Like the iterative type I PKS, the type II PKS can be grouped into reductive and nonreductive classes depending on the presence (red arrow, *act-jad*) or absence (*Tcm* and *WhiE*) of a KR. The names and structures of the final product are shown on the right. The product for the WhiE PKS is a spore pigment of unknown structure.

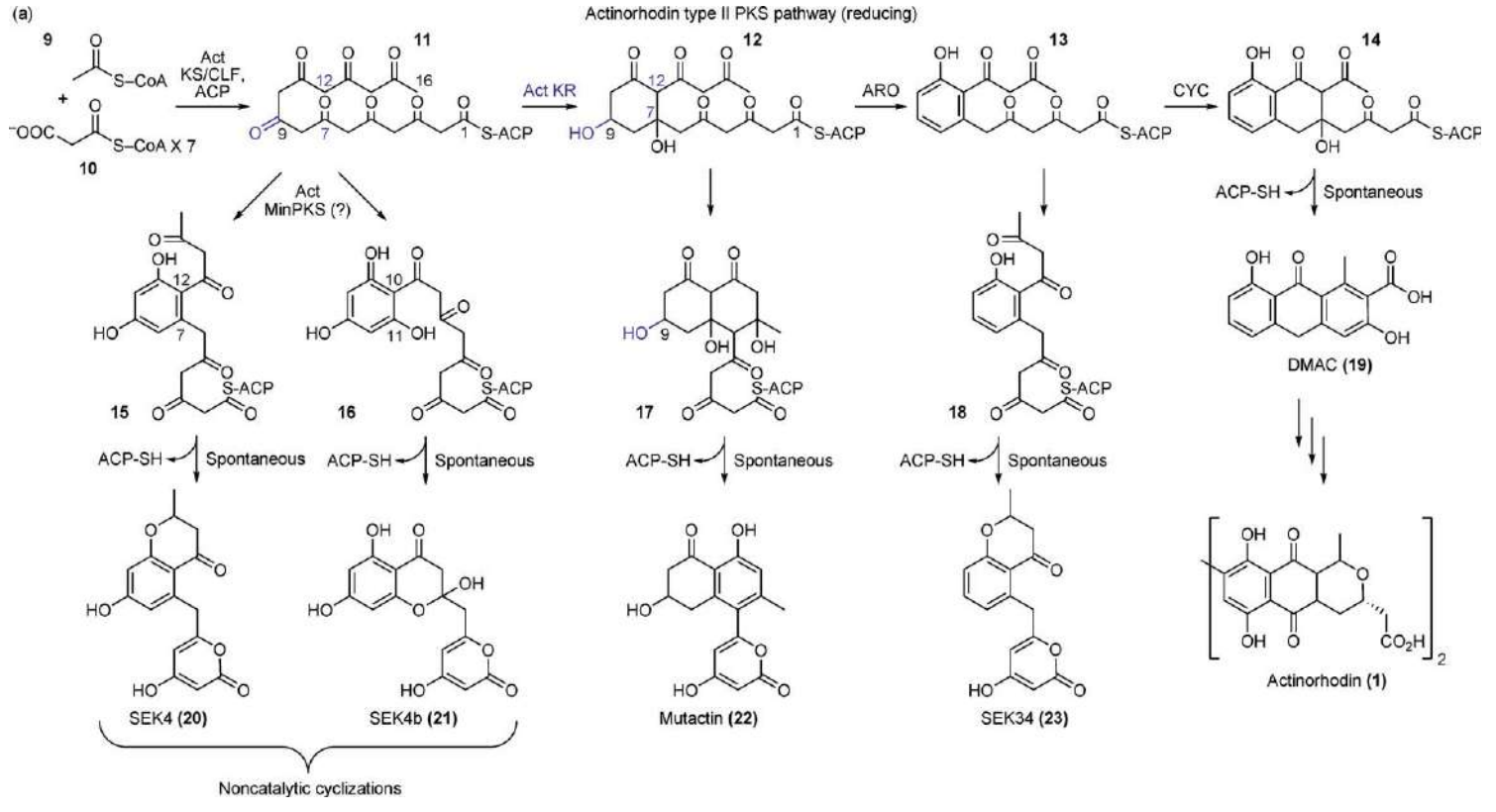


Figure 6 (Continued)

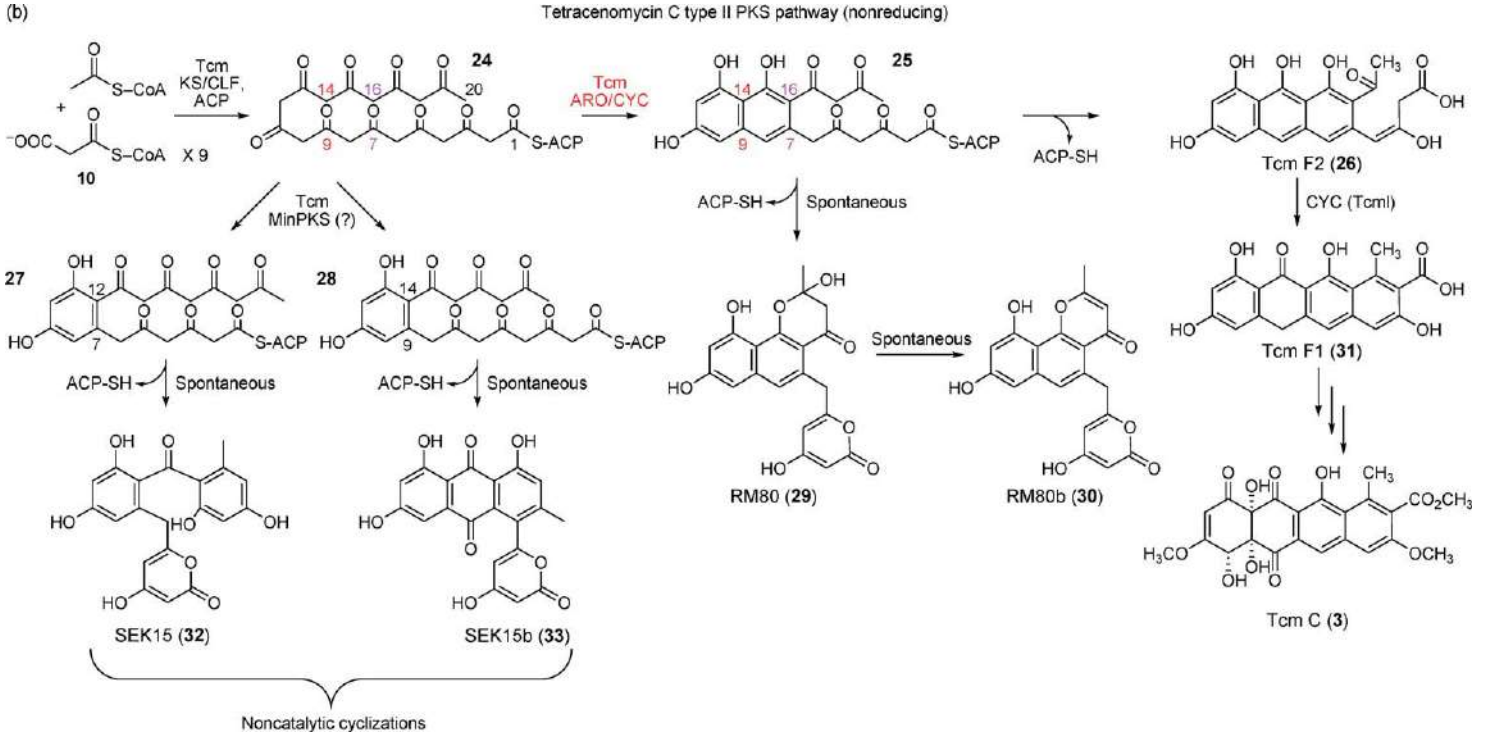


Figure 6 A comparison of the biosynthetic pathways for the actinorhodin (act) and tetracenomycin (Tcm) PKSs. (a) The act PKS produces a C16 product and utilizes a KR to specifically reduce the C9 position (blue) of a C7–C12 cyclized intermediate. (b) The Tcm PKS lacks a KR and produces a C20 intermediate whose cyclization specificity is controlled by ARO/CYC (red) to be C9–C14 for the first ring and C7–C16 for the second.

polyketides[35] (**Figure 1**), which are characterized by the presence of multiple aromatic rings fused in a linear arrangement. These aromatic polyketides include commonly used antibiotics (such as act and tetracycline) and anticancer agents (such as doxorubicin) that target DNA/RNA and protein biosyntheses in bacteria and cancer cells.[36–38]

Over the past decade, there has been extensive advances in *Streptomyces* genetics, which has allowed the mass production of large quantities of type II polyketides rapidly, affording a more economic and technically simpler protocol to produce potentially bioactive polyketides than traditional organic synthesis.[39] Past molecular genetic studies in the groups of Hopwood and coworkers,[35] Hutchinson,[40] Floss,[41] Khosla and coworkers,[42,43] Robinson,[44] Moore and Piel,[45] and Mendez and Salas[46] have established the mechanism for a typical aromatic PKS (**Figure 6**), where the polyketide chain is assembled by a 'minimal PKS' comprised of the ketosynthase/chain length factor (heterodimeric protein, also known as KS_α–KS_β) (KS/CLF) heterodimer and ACP. Chain elongation can then be followed by first ring cyclization, chain reduction by KR, and subsequent ring aromatization/cyclization by aromatase/cyclase (ARO/CYC). The chain transfer among different enzymes is mediated by the phosphopantetheinyl (PPT) group, which is covalently attached to a conserved serine of ACP. The reducing type II PKS (**Figure 6(a)**, act PKS) contains a KR enzyme and typically reduces a carbonyl group to hydroxyl group at the C9 position with a first ring C7–C12 cyclization (such as **12**). On the other hand, the non-reducing type II PKS (**Figure 6(b)**, the Tcm PKS), in the absence of KR, can produce a C9–C14 cyclized first ring in the presence of ARO/CYC (such as **25**). Similar to the 'mix and match' experiments with modular type I PKS, separate aromatic PKS proteins from different type II PKSs have been combined to create over 100 novel polyketides.[14,47–56] These results led to the conclusion that the large diversity of naturally occurring aromatic polyketides is a result of the controlled variation in chain length, choice of building blocks, and differential chain reductions and cyclizations mediated by type II PKSs. In this chapter, we will discuss the sequence–structure–function relationship of type II PKSs in the following areas:

1. Polyketide chain length control by both the priming KS and the elongation KS. Sections 1.08.4.2 and 1.08.4.3 will discuss this topic in detail.
2. Regiospecific ketoreduction by the KR (**Figure 6(a)**). Ketoreduction is chemically identical to the corresponding fatty acid ketoreduction; however, the regiospecificity is drastically different. Whereas fatty acid KR reduces every carbonyl group, the aromatic polyketide KR has a high specificity for the C9 carbonyl group.[57] Section 1.08.4.4 discusses the sequence–structure–function relationship that leads to the highly regiospecific behavior of KR.
3. Formation of aromatic rings that are essential for the antibiotic and anticancer activity of polyketides.[37] The cyclization and aromatization of the polyketide rings are key steps that control polyketide diversity. For example, doxorubicin and mithramycin are both biosynthesized from a 20-carbon linear polyketide.[7,9] However, doxorubicin and mithramycin are produced by cyclization at C7–C12 and C9–C14, respectively.[58,59] Studies by the Khosla, Moore, and Salas groups have demonstrated that KR inclusion results in products that are cyclized at the C7–C12 positions (**19**, **22**, **23**),[50,56,59] whereas the inclusion of a monodomain ARO/CYC (in the absence of KR) results in products that are cyclized at the C9–C14 positions (**29**, **30**, **26**).[49,59] Sections 1.08.4.4–1.08.4.6 discuss how KR and ARO/CYCs may accomplish the observed cyclization specificities.
4. Protein–protein interactions affecting polyketide biosynthesis. Protein–protein interactions are important in the following aspects: (a) Polyketide chain transfer between proteins is often abolished when shuffling enzymes due to the loss of protein–protein interactions that are important for polyketide chain transfer.[60,61] (b) The timing of chain elongation versus chain modification is closely related with protein–protein interactions.[50] As a result, the polyketide chain length, reduction, and cyclization patterns can be dramatically changed by the inclusion of certain PKS enzymes, such as KR and ARO/CYC.[50] (c) The regiospecificities of ketoreduction and ring cyclization are closely related to protein–protein interactions among KS/CLF, ACP, KR, and ARO/CYCs.[62] Sections 1.08.4.7, 1.08.5, and 1.08.6 discuss how protein–protein interactions and the mediation of ACP may affect the product outcome in type II PKS.

1.08.3.1 Similarities and Differences between FAS and PKS

FAS and PKS share many structural and functional similarities:

1. Both FAS and PKS can be grouped into type I and type II, depending on whether the complex is a single polypeptide with multiple domains, or numerous stand-alone proteins.
2. In both FAS and PKS, the KS catalyzes chain elongation using small, two to four carbon building blocks attached to ACP.
3. Both use KR, DH, and ER to modify the growing chain after each elongation cycle.
4. Structural conservation seen in high-resolution structures of isolated type I and type II PKS and FAS domains suggest that both FAS and PKS may form similar mega complexes.

However, subtle differences in reduction patterns between PKS and FAS lead to the staggering diversity of polyketide natural products compared to the simplicity of fatty acids (**Figure 7**) due to the following features: (1) Chain modification. FAS uses KR, ER, and DH after each elongation step to fully reduce every carbon. In comparison, the chain-modification events are much more complicated in PKS. In some type II and iterative type I PKSs, the polyketide is left completely unreduced due to the lack of KR, as in the case of Tcm C (**3**)[63] and norsolorinic acid[64] biosynthesis. In other cases (such as act (**1**) and doxorubicin (**2**)), the PKS KR catalyzes a regio- and stereospecific reduction at a single position,[1] as opposed to reducing every carbonyl group in FAS KR. (2) An altered regiospecificity: in PKS, the action of the KR is not necessarily followed by DH or ER. Therefore, from a product standpoint, polyketides often contain multiple hydroxyl groups that undergo further reactions such as cyclization, aromatization, glycosylation, and methylation, leading to the huge chemical diversity observed in nature (**Figures 1** and **7**). How ketoreduction and cyclization are controlled, especially in type II PKSs remains a mystery, and the lack of information has so far prevented rational control of reduction and cyclization during polyketide biosynthesis. Therefore, understanding the substrate specificities and molecular mechanisms governing the ketoreduction and cyclization events in type II PKS systems will be scientifically significant by filling in the knowledge gaps and providing a structural guide for manipulating these proteins to produce polyketides with new reduction and cyclization patterns.

1.08.4 Structural Enzymology of Individual Type II PKS Enzymes

Random domain shuffling often results in complete inactivation of PKSs. For example, Burson *et al.*[61] demonstrated that shuffling heterologous KS/CLF domains resulted in at least 90% inactive complexes. Furthermore, because of an incomplete understanding of the mechanisms and specificities of individual enzymes, as well as the influence of protein–protein interactions in PKS synthetic potential, many possible manipulations such as the rational control of cyclization patterns are currently not available. This lack of understanding about the structure–function relationship between PKS enzymes has recently been highlighted as a problem by Staunton and Weissman in the millennium PKS review.[65] Below, we discuss the available structural details about type II PKS enzymes as an important step toward remedying the current engineering limitations caused by a lack of understanding about sequence–structure–function in PKSs.

1.08.4.1 Malonyl-Coenzyme A:ACP transacetylase

The enzyme MAT has been proposed to charge the PKS ACP protein with a malonyl group, the building block for polyketide biosynthesis. However, the actual role of MAT has long been debated, mainly because of the ability of ACP to self-malonylate.[66] Summarized below is the research progress during the past 15 years on the ACP self-malonylation issue.

In 1995, the Hutchinson group cloned the ACP from the Tcm C (**3**) PKS gene cluster (Tcm),[67] and found an upstream gene that encodes a protein highly homologous to the *E. coli* MAT, which charges the FAS ACP with the malonyl group. The 1995 study further showed that the newly found MAT was capable of transferring radioactive malonate from malonyl-CoA to the *E. coli* AcpP and *Streptomyces glaucescens* FabC ACPs, as well as to the TcmM ACP protein of the Tcm type II PKS. On the basis of its *in vitro* activity, Shen and Hutchinson

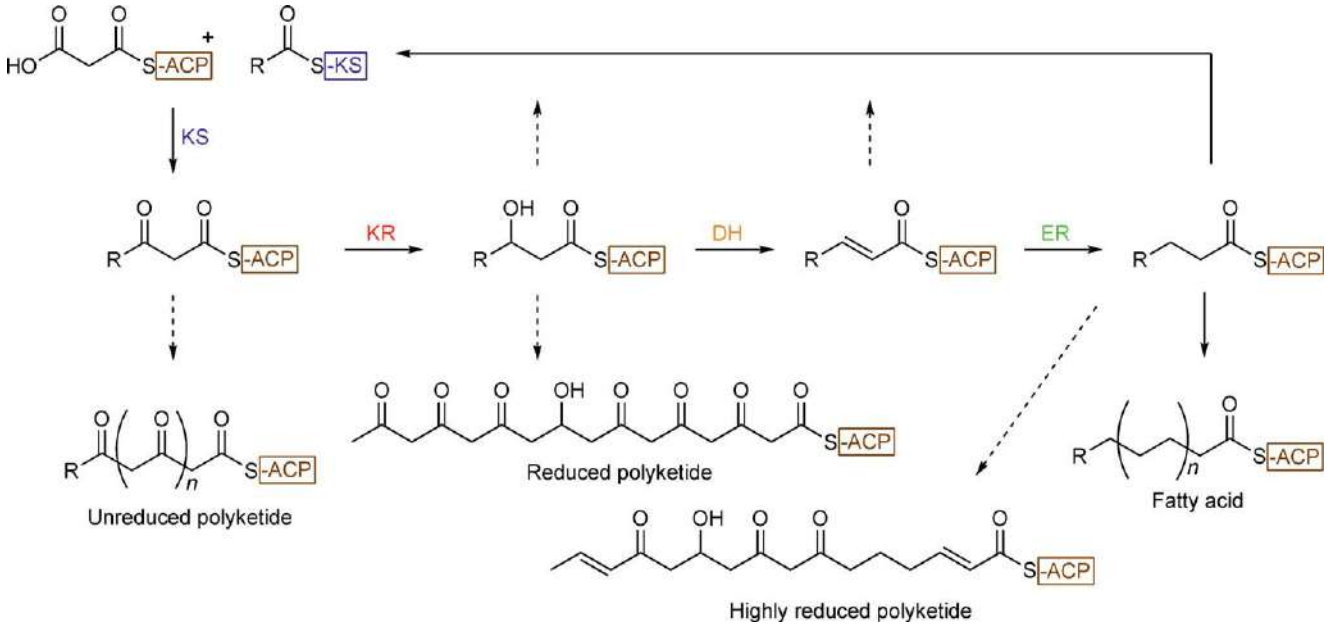


Figure 7 A general scheme for the biosynthesis of polyketides and fatty acids. Fatty acid biosynthesis is primed with an acetyl starter unit, and the chain extended via iterative condensations with malonyl-ACP. In contrast, polyketide biosynthesis can be initiated with acetyl, propionyl, or other acyl groups that can be coupled with malonyl-ACP or methylmalonyl-ACP. In the FAS, the β-carbonyl is fully reduced after each chain elongation round to produce an aliphatic fatty acid. The large diversity of polyketide structures is a result of partial reduction after each chain elongation round.

proposed that the newly found MAT may be responsible for charging the TcmM ACP with malonate *in vivo*, and that MAT may provide a functional connection between fatty acid and polyketide metabolism in *S. glaucescens*.

The above hypothesis was supported in 1998 by results from the Khosla lab based on the malonylation of frenolicin (fren) ACP by partially purified *Streptomyces coelicolor* MAT.[68] The Hutchinson group also showed that with purified components, the inclusion of MAT will greatly increase polyketide production.[69] However, in the absence of MAT, self-malonylation does occur albeit with a 10-fold slower rate.[69] A similar conclusion was also drawn for the R1128 MAT.[53] The above results agree with careful studies of the act PKS conducted by the Simpson group at Bristol University, where they showed that MAT dependence is correlated with ACP concentration; that is, when ACP is limiting, MAT is necessary; however, when ACP is in excess, self-malonylation can take place without MAT.[66] The Simpson group further showed that ACP is capable of self-malonylation and, along with the Leadlay group, showed that decarboxylation of malonyl-ACP can happen in the CLF or KS enzymes.[70] Therefore, up to 1998, the MAT studies from the above three groups were consistent.

In 1999, the Khosla group reported three findings about the MAT-mediated malonylation of fren ACP: (1) the K_d of MAT is very low (nmol l^{-1} range), (2) the S97A mutant of MAT still supports polyketide production, and (3) S97A MAT alone (in the absence of ACP and KS/CLF) can still be malonylated, leading to the proposal that there is a second 'active site' in MAT that catalyzes malonylation.[71] The 'two catalytic sites' hypothesis was proposed again in 2001 by Khosla and coworkers[72] based on mutation experiments that double mutations (H96A–S97A) of both 'sites' are necessary to completely abolish malonylation activity. For this unusual proposal to work, H96 needs to be a nucleophile that attacks malonyl-CoA to form a transient malonyl-histidine adduct intermediate (as opposed to the common serine-*O*-malonyl intermediate), a mechanism not previously observed for any AT. A detailed analysis in 2002 from the Bristol researchers discussed several experimental issues about the 'two-sites' proposal.[73] Furthermore, in 2003, it was reported that the biosynthesis of branched-chained fatty acid in *Bacillus insolitus*[74] also does not require MAT, with self-malonylation proposed as a possible cause. Also, throughout the studies, the Stanford group used the fren ACP in labeling and *N*-acetylcysteamine (SNAC) in kinetic assays,[72] while the Bristol group used the FAS ACP or 'natural' C17S act ACP mutant (which alleviates complications from a surface Cys).[73] The successful development of a MAT assay using the 'natural' ACP as its substrate is an important accomplishment in itself and for the continuation of any MAT-related studies.[73] In 2005, the Bristol group chemically synthesized apo act ACP (thus eliminating the possibility of *E. coli* ACP contamination during protein purification), and showed that the resultant holo-ACP underwent self-malonylation in the presence of malonyl-CoA.[75,76] In 2006, the Bristol group then showed that the act PKS ACP can malonylate *S. coelicolor* FAS ACP and other type II PKS ACPs *in vitro*, while mutations showed that the self-malonylation and malonylate transfer are two uncoupled catalytic events. This seminal work with act ACP was followed in 2007 with a detailed study dissecting chain initiation and elongation of the minimal PKS in the absence and presence of MAT.[77] The study confirmed that the rate of loading increases in the presence of MAT, and the decarboxylation of malonyl-ACP to form acetyl-ACP (initiation) is the rate-limiting step. When an excess of acetyl ACP is supplied, either chain extension, cyclization, or release steps become rate limiting. Although it is true that MAT is not absolutely necessary, because ACP can self-malonylate, in a living cell of *S. coelicolor*, it is still likely that the presence of MAT may facilitate the malonylation of a PKS ACP. For this reason, the MAT crystal structure and enzymological studies are included and discussed in this chapter.

1.08.4.1.1 Crystal structure of MAT

The 2.0 Å crystal structure of the *S. coelicolor* MAT reveals that it is a structural homolog of the *E. coli* fatty acid MAT (fabD).[78] The 32 kDa monomeric MAT structure consists of two subdomains (**Figure 8(a)**). The large subdomain (residues 1–132 and 198–316) has a central parallel β-sheet surrounded by 12 helices, and has a α,β-hydrolase fold. The small subdomain (residues 133–197) has a ferredoxin-like fold, with a four-stranded antiparallel β-sheet and two helices. There are two acetates cocrystallized in the MAT active site. The catalytic dyad H201-S97 is identified near one acetate, with the catalytic S97 located at the 'nucleophilic elbow', a signature of α,β-hydrolase fold, in a highly conserved GHSXG motif. Furthermore, one of the acetates is bound

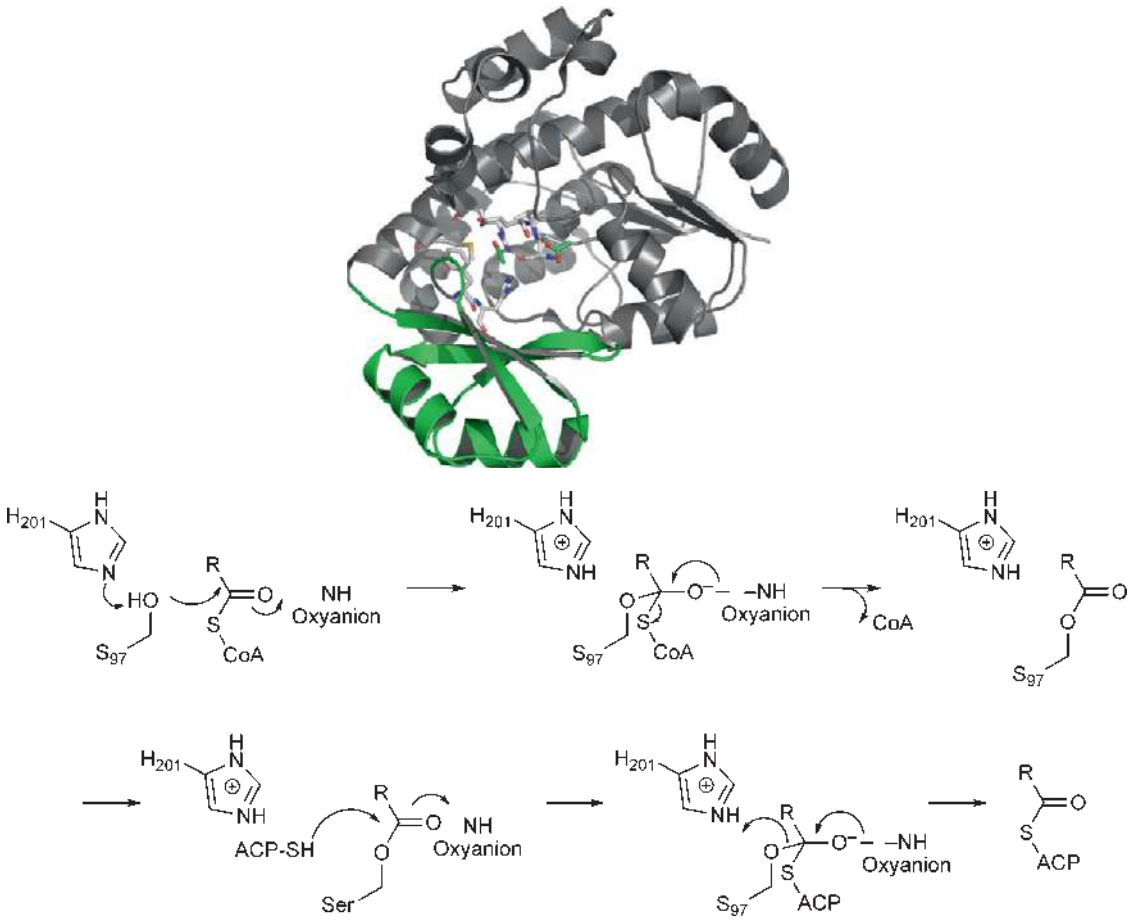


Figure 8 (a) The crystal structure of /S. coelicolor/ MAT showed that it belongs to the a,b-hydrolase superfamily with the large domain (in grey) and a ferredoxin-like small domain (in green). Residues mentioned in text are in sticks. (b) Proposed catalytic mechanism.

to Q9 and R122 in the active site, mimicking the carboxyl end of a bound malonyl group, and the backbone amides of Q9 and V98 may create the oxyanion hole (**Figure 8(a)**).

1.08.4.1.2 Proposed mechanism of MAT

MAT catalyzes malonyl transfer via a ping-pong bi–bi mechanism (**Figure 8(b)**). In the first half-reaction, S97 attacks malonyl-CoA (**10**) (mediated by the active site base H201), resulting in the Ser-*O*-malonyl intermediate. Assuming the malonyl carboxylate adopts a similar orientation to the bound acetate, the thioester would then be located near the oxyanion hole (Q9 and V98), and the tetrahedral intermediate can be stabilized through favorable charge–dipole interaction in the oxyanion hole. H201 subsequently protonates CoA and liberates it from MAT. The side chain of R122 is then proposed to interact with the malonyl carboxylate, similar to the *E. coli* MAT. In the second half reaction, the PPT group of ACP enters the MAT active site, and the PPT attacks the Ser-*O*-malonyl group, possibly via the formation of the second tetrahedral intermediate that is stabilized by the oxyanion hole. H201 then reprotonates the catalytic serine and releases malonyl-ACP.

1.08.4.1.3 Proposed molecular basis of substrate specificity of MAT

The *S. coelicolor* MAT is highly specific for malonyl-CoA, but not methylmalonyl-CoA. Malonyl-specific ATs from both type I and type II PKSs have active site residues that are consistently different than those from methylmalonyl-specific ATs. Based on the MAT crystal structure and sequence alignment, M126 and F200 were proposed to impose steric hindrance that prevents methylmalonyl-CoA from binding MAT. These two

residues become Ile/Leu and Ser in methylmalonyl-specific ATs. To probe if M126 and F200 are the sole determining residues for the MAT substrate specificity, a series of single mutants were constructed and assayed with malonyl- and methylmalonyl-CoAs (Q. Li and C. Khosla, unpublished data). The mutations of these proposed residues did not change the MAT substrate specificity. This suggests that the MAT does not discriminate against R-substituted acyl-CoA thioesters solely on the basis of substrate size. Rather, similar to the type I PKS MATs, several structural components, including long-range effects, must be taken into consideration for the substrate specificity of MATs.

1.08.4.2 The ZhuH Priming Ketosynthase

ZhuH is a priming KS that primes a unique 'priming' ACP with acetyl, propionyl, butyryl, and isobutyryl groups during the biosynthesis of the R1128 family of polyketides (**Figure 9(a)**). The R1128 substances are anthraquinone natural products that were previously reported as nonsteroidal estrogen receptor antagonists.[79–81] In 2000, the gene cluster for R1128 biosynthesis was cloned into *Streptomyces lividans*, and its sequence revealed genes encoding two KSs (ZhuA and ZhuH), a chain length factor (ZhuB), other chain-modification genes and, remarkably, two ACPs (ZhuN and ZhuG).[82] Each protein was expressed and purified from *E. coli*, and it was found by kinetic analysis that ZhuH promotes the formation of acyl-ACPs (acetyl-, propionyl-, butyryl-, and isobutyryl-), where propionyl-CoA and isobutyryl-CoA were the two most preferred substrates. The specificity of ZhuH is consistent with the chemical structures of products R1128B and R1128C, as well as the ZhuH substrate pocket deduced from its crystal structure (see Section 1.08.4.2.1).[83] It was further shown in 2003 that ZhuH is highly specific for the ACP ZhuG (the ACP of the 'priming module' that teams up with a priming KS such as ZhuH), but not ZhuN (the ACP of the 'minimal PKS' that teams up with KS/CLF).[84] The Khosla group showed that in the presence of a KS/CLF and MAT, all ACPs from minimal PKSs supported polyketide synthesis at comparable rates, whereas PKS activity was largely attenuated in the presence of an ACP from an initiation module. Conversely, the opposite specificity pattern was observed with priming KSs, for which the ACPs from initiation modules afford higher PKS rate, while ACPs from minimal PKSs resulted in significantly lower PKS rate. This marks an important discovery that in type II PKS, an 'initiation module' can exist to prime the PKS with different starter units.

In 2004, with the R1128 initiation module isolated and characterized (consisting of the priming KS ZhuH, the priming ACP ZhuG, and MAT), it was further demonstrated that this 'priming module' can interact with a downstream 'elongation module' (consisting of the minimal PKS: KS/CLF and ACP), resulting in new polyketide products with unnatural starter units such as propionyl and isobutyryl groups.[48] The same study also concluded that the minimal PKS controls polyketide chain length by counting the number of polyketide backbone carbons, rather than the number of elongation cycles. In comparison, downstream polyketide-modifying enzymes such as KR and ARO/CYC need to recognize the specific chemical moiety in the polyketide chain, rather than the carbon chain length. In 2005, this bimodular system was expanded to biosynthesize analogs of the decaketide aklanonic acid.[85] The above studies show that the engineered 'bimodular type II PKS', containing the 'initiation' and 'elongation' modules, provides a versatile tool to biosynthesize new polyketides with different starter units and chain lengths in a combinatorial fashion. In the future, if one can extend the substrate tolerance of ZhuH (such as via structure-based mutagenesis), more starter units can be primed to a bimodular type II PKS system further increasing the diversity of engineered type II polyketides.

1.08.4.2.1 Crystal structure of ZhuH

The ZhuH crystal structure shows that ZhuH is a dimer with a thiolase fold that has similar dimensions (45 Å × 60 Å × 75 Å) to the type II KSs, including the priming KS FabH and the elongation KS, β-ketoacyl synthase (KAS) I, and KAS II (**Figure 9(b)**).[83] The ZhuH monomer has a five-layered core ($2\alpha-5\beta-2\alpha-5\beta-2\alpha$): three layers of α-helices interspersed by two layers of β-sheet, with extensive connecting loops. The core is divided into N- and C-domains with a similar fold, where a pseudo-twofold axis lies between Nα3 and Cα3 that is parallel to the twofold dimer axis. The major differences between ZhuH and *E. coli* FabH are that ZhuH has a longer N-terminus and two insertions in L9 and L13α1. Since the L9 insertion is close to the dimer interface, while L13α1 is at the entrance of the substrate-binding pocket, the different structural features likely contribute to the different substrate specificity between ZhuH and FabH.

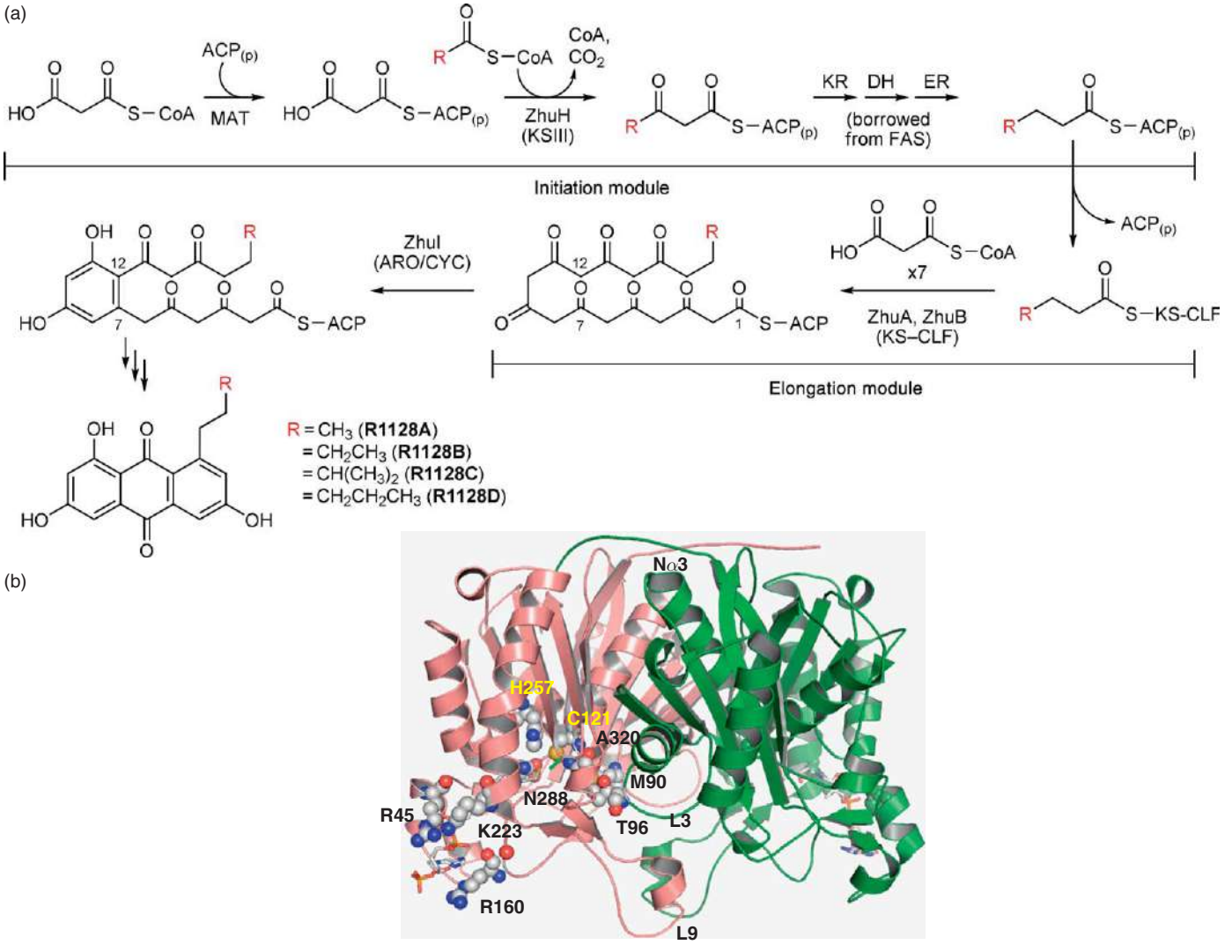


Figure 9 (a) The R1128 PKS. (b) The crystal structure of ZhuH with the thiolase fold as a homodimer (monomer A and B in red and green, respectively). The bound CoA helps define the substrate binding pocket (in sticks). Residues discussed are in spheres.

The extensive dimer interface has approximately 3000 Å^2 of surface area near Nβ3, Nα3, L9, and L3, and a 10-stranded β-sheet traverses the entire ZhuH dimer. The L9 and L3 loops interact extensively with L9′ and L3′ from the other monomer (**Figure 9(b)**), and this dimer interface defines the bottom of the substrate channel. Many residues that define the active site are also close to the dimer interface, such as the catalytic C121, and M90 and T96, which define the shape of the acyl group binding pocket, are located on L3 and L3′, respectively. Therefore, the extensive ZhuH dimer interface is proposed to be important for maintaining the fold and positioning the catalytic site of ZhuH.

1.08.4.2.2 Proposed mechanism of ZhuH

The ZhuH structure was solved with CoA bound (a degradation product from acetyl-CoA, **Figure 9(b)**), helping to define the 20 Å long substrate pocket in which the CoA extends from the pocket entrance to the catalytic triad. An electropositive groove (comprising R45, R160, and K223) is located near the ZhuH entrance (**Figure 9(b)**, spheres) and is conserved between ZhuH and FabH; this region may possibly act as the binding site for the incoming malonyl-ACP. The extensive CoA–ZhuH interaction is consistent with the observation that CoA binding is important for maintaining protein conformation for substrate channel shape.

In the proposed mechanism, an incoming acyl-CoA binds its acyl group at the end of the substrate-binding pocket, where the catalytic triad, C121–H257–N288, form a hydrogen-bonded network. H257 is the proposed active site base that deprotonates C121 to increase its nucleophilicity. The amide NHs of C121 and A320 define the proposed oxyanion hole. When the ZhuH structure is overlayed with other FAS KS structures, the ZhuH triad (H257–N288–C121) superimposes well with that of the *E. coli* priming KS FabH (H244–N274–C112), as well as the N343–H375–C125 triad of thiolase I. In contrast, the elongating KSs, FabB, and FabF have a His-His-Cys triad, where N288 of ZhuH is replaced by H340 of FabF. The Asn versus His difference between the priming and elongation KSs help explain why inhibitors such as cerulenin and thiolactomycin (TLM) preferentially bind the elongation KSs.

Based on the crystal structure and mutational studies, a catalytic mechanism for ZhuH was proposed: acyl-CoA substrates bind to the ZhuH substrate pocket, where the extensive CoA–ZhuH interactions help orient the substrate near the active site C121. The thiol group of C121, possibly deprotonated by H257, attacks the acyl carbonyl, while the NHs of C121 and A320 stabilize the oxyanion. CoA release is followed by malonyl-ACP binding, and the decarboxylation of malonate produces the enolate, stabilized by H257 and N288. The enolate collapses and then undergoes nucleophilic attack at the acyl-thioester, resulting in C–C bond formation and the product β-(ketoacyl)-ACP (**Figure 9(c)**).

1.08.4.2.3 Proposed molecular basis of substrate specificity of ZhuH

ZhuH and other FAS priming KSs are highly homologous. However, their corresponding substrate specificities are highly diverse. For example, similar to ZhuH, FabH from *E. coli*, *Streptomyces pneumoniae* and *S. glaucescens* can be primed with C2-C4 acyl-CoAs, while the *Bacillus subtilis* FabH prefers branched substrates, and the *Mycobacterium tuberculosis* FabH prefers much longer chains, such as myristoyl-CoA. It is of great interest to understand the molecular basis for the diverse substrate specificities of these priming KSs.

The acyl group-binding pocket of ZhuH is lined by highly conserved, hydrophobic residues. Therefore, although these conserved residues define the shape of the acyl group-binding pocket, they do not appear to be responsible for the variable substrate specificity observed among the priming KSs. Based on the ZhuH crystal structure, it was proposed that the size and shape of the substrate pocket may be defined by structural features that form the bottom of the substrate pocket. In the case of ZhuH, the bottom of the substrate pocket is defined by M90 and T96, the only two nonconserved residues at the end of ZhuH substrate pocket. When different acyl-CoAs were docked *in silico* into ZhuH, the substrate pocket consistently accommodated C2–C4 acyl-CoAs, but not longer acyl-CoAs. This is consistent with *in vivo* and *in vitro* studies showing that ZhuH is preferentially primed by isobutyryl or propionyl groups, but can also tolerate acetyl and butyryl primer units. It was proposed that the size of the 'gate-keeping residues' at the bottom of the substrate pocket, namely M90 and T96 in ZhuH (**Figure 9(b)**), may be important for substrate specificity. While this hypothesis is attractive in its potential application in engineering new priming KSs that accept different primer units leading to the biosynthesis of new polyketides, it was found that single or double mutations of M90 and T96 did not change the ZhuH

substrate specificity. Clearly, because the L3 loop (where M90 and T96 are located) interact extensively with other structural components, long-range effects may also be important in defining the size and shape of the priming KS substrate pocket.

1.08.4.3 The Elongating Ketosynthase/Chain Length Factor Complex

KS/CLF assembles multiple copies of malonyl-CoA via decarboxylative Claisen condensation to produce the polyketide backbone. KS/CLF and ACP form the 'minimal PKS' (**Figure 6**, **9**/**10** to **11** or **24**). During the process, ACP can either accept a malonyl group from MAT or self-malonylate to become malonyl-ACP, via the PPT group that is covalently attached to an ACP serine residue. Decarboxylation of malonyl-CoA produces acetyl-CoA that is transferred from MAT to KS/CLF via ACP (using an 18-Å-long PPT arm) and subsequently added to the growing chain by KS/CLF. In the case of act biosynthesis, this catalytic cycle is repeated 8 times to form a 16C polyketide. Without tailoring enzymes, the act minimal PKS produces the shunt products, SEK4 and SEK4b (**Figure 6(a)**, **20**, **21**).

The Claisen condensation has been well studied in FAS KSs. However, different functions have been proposed for the CLF protein, whose active site cysteine is mutated to a Gln. Based on heterologous expression of the act minimal PKS and its product analysis, it was proposed that CLF may determine the chain length.[49,52] However, later studies showed that heterologous CLF may not alter the product chain length in the predicted fashion.[63] Therefore, it was proposed that CLF may exert major influence in the presence of its cognate KS, and that KR and ARO/CYC may also have important effects on the length of the product. The important enzymological issues of KS/CLF can be summarized below:

1. *Chain initiation.* Although it is agreed that malonyl-CoA but not acetyl-CoA is the priming unit for act and Tcm PKSs, it was unclear whether KS or CLF carries out the initiation reaction that turns the malonyl group into the acetyl group via decarboxylation, and whether KS has AT activity that transfers the resulting acetyl group from CLF to the KS active site. In 1999, researchers from Cambridge and Bristol universities showed by mutagenesis that when the act KS active site Cys is mutated to Ala, acetyl-ACP was still produced as detected by mass spectroscopy.[70] In contrast, the double mutations of KS-Cys to Gln and CLF-Gln to Ala resulted in the abolishment of acetyl-ACP production. Furthermore, double mutations of the KS-Cys to Gln and CLF-Gln to Ala, as well as single mutations of CLF-Gln to Ala, still result in malonyl decarboxylation. The above results support that both KS and CLF can initiate polyketide biosynthesis by providing acetyl-ACP via decarboxylation of malonyl-CoA. A detailed mechanism proposing residues involved in chain initiation (besides KS-Cys and CLF-Gln) require a high-resolution structure of KS/CLF.
2. *The molecular basis of chain length control.* Although KS/CLF was proposed and proven to be important for defining the polyketide chain length, how KS/CLF controls the polyketide chain length such that a dominant product is observed in different PKSs was not clear. For example, act KS/CLF produces a 16-carbon polyketide (**Figure 6(a)**), whereas Tcm KS/CLF produces a 20-carbon polyketide (**Figure 6**). How KS/CLF controls the chain length is not well understood before a high-resolution structure became available.
3. *Cyclization specificity.* It was proposed that KS/CLF may be involved in subsequent cyclization, since in the absence of downstream tailoring enzymes such as KR or ARO/CYC, some minimal PKSs produce one or two dominant polyketide products, mainly with C7–C12 first ring cyclization, while others, such as WhiE minimal PKS, produce a mixture of products that are highly diverse in the polyketide chain lengths and cyclization patterns. The different product outcomes of act, Tcm versus WhiE minimal PKS have not been explained.
4. *Timing for downstream KR and ARO/CYC.* KR catalyzes a highly specific C9 reduction, while the inclusion of certain ARO/CYC such as Tcm ARO/CYC produces a dominant C9–C14 cyclized polyketide product. To precisely control downstream reduction and cyclization/aromatization, KS/CLF must display precise timing control between elongation and reduction/cyclization. How KS/CLF achieves the precise timing is not well understood.

1.08.4.3.1 Crystal structure of the actinorhodin KS/CLF

In 2004, the crystal structure of act KS/CLF shed light on the above four enzymological issues.[86] The act KS/CLF crystal structure was solved to 2.0 Å. Similar to ZhuH, the act KS/CLF has a thiolase fold, with an alternating α-helix and β-sheet architecture ($\alpha\beta\alpha\beta\alpha$, **Figure 10(a)**). Both act KS and CLF are highly homologous in sequence and structure to the *E. coli* fatty acid elongation KS, FabF, with 37 and 28% of sequence identity to FabF, respectively. KS and CLF have an extensive hetero-dimer interface, with approximately 6400 Å^2 of buried, highly hydrophobic surface area. The highly conserved protein fold and dimer interface of KS/CLF versus other fatty acid KSs allow detailed comparison of its active site and substrate pocket.

1.08.4.3.2 Proposed mechanism of KS/CLF

The KS active site Cys is located in the middle of KS (**Figures 10(a)** and **10(b)**), with an amphipathic tunnel that extends 17 Å from Cys169 to the KS/CLF dimer interface. The Claisen condensation mechanism is proposed to be similar to that of ZhuH and FAS KS domains.[86] The structure resolves many of the questions about the roles of KS and CLF. Previously, CLF was proposed to have an active site similar to KS, capable of carrying out the first decarboxylation centered at the reactive Q161 of CLF (**Figure 10(a)**). However, conformation seen in the crystal structure showed Q161 in an orientation that cannot form the oxyanion hole. Furthermore, the observed CLF active site is filled by large side chains that leave little room for a malonyl substrate. Therefore, CLF must undergo extensive conformational change near the active site upon malonyl-CoA binding if it is to decarboxylate the malonyl group, and the crystal structure may just be one snapshot along the catalytic cycle.[86]

Because the minimal PKS can produce polyketides from malonyl-CoA, KS was thought to possess AT activity. Furthermore, the AT motif, 'GHSXG', is present in the KS sequence. However, the putative AT nucleophile in the KS enzyme, S347, is buried and not near the substrate channel (**Figure 10(a)**).[86] Whether the AT activity is predominantly carried out by ACP (via self-malonylation) or other residues in KS requires further studies.

Using electrospray mass spectrometry of KS and CLF, it was found that the intermediates such as monoketide and diketide are bound to KS, not CLF.[86] The study supports the hypothesis that, under normal situations (when KS has the active site Cys168 intact), CLF may also regulate chain length (see next paragraph for details), but KS can catalyze both chain initiation and elongation with a similar enzyme mechanism as other elongation KSs. It was further proposed that KS and CLF had coevolved to complement each other, where KS has been optimized for catalysis and CLF for regulating chain length.

The act KS/CLF structure shows that it has a long, narrow tunnel that can accommodate a linear, 16-carbon (octaketide) intermediate (**11**) (**Figures 6(a)** and **10(b)**). Following chain elongation to 16 carbons, cyclization must occur, and a C7–C12 dominant product, SEK4 (**20**), was observed (**Figure 6(a)**). Where does the

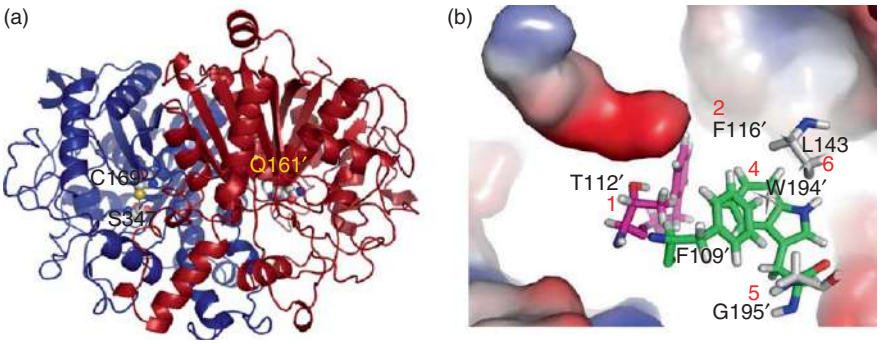


Figure 10 (a) The crystal structure of act KS/CLF shows a heterodimer with a 20 Å long channel. The KS active site Cys169, the CLF Q161, and the KS S347 are shown in spheres. (b) The KS/CLF substrate pocket (rendered as electrostatic surface) can accommodate a 16-carbon polyketide chain, and residues discussed in the text are in sticks. The three gates are shown in purple, green, and white atoms, respectively.

cyclization take place? To cyclize the polyketide in the linear chamber will require extensive protein conformational change, and may not be favorable. Instead, if the cyclization indeed takes place inside KS/CLF, an alternative cyclization chamber has been suggested that lies between the pocket opening and the active site Cys, where a crystalline water was observed that may act as a general acid during cyclization. This is clearly the case in the plant type III PKS such as the chalcone synthase, in which the tetraketide cyclizes inside the same tunnel in which it was polymerized. However, superimposition of KS/CLF with chalcone synthase suggests that chalcone synthase has a much wider cyclization chamber (for a tetraketide) than KS/CLF (for a much larger octaketide).[18] This issue will be further discussed in Sections 1.08.4.4 and 1.08.4.5. More biochemical and structural work needs to be done to elucidate the exact location of first ring cyclization.

1.08.4.3.3 Proposed molecular basis of substrate specificity of KS/CLF

The major difference between FAS KS and PKS KS/CLF structures is that FAS KS has a hydrophobic substrate channel,[19] while KS/CLF has an amphipathic substrate channel. This is consistent with the different nature of FAS KS and PKS KS/CLF. For the FAS KS, a growing fatty acid chain must leave the KS after each condensation reaction so that downstream enzymes (such as KR, DH, and ER) can reduce/dehydrate the β-carbonyl to a methylene group. In comparison, the amphipathic substrate tunnel in KS/CLF may interact via hydrogen bonding with carbonyl groups on the growing polyketide chain, thus preventing the growing polyketide from leaving KS/CLF – note however that it has been proposed that a truncated polyketide may leave KS/CLF, cyclize, and then re-enter act KS/CLF to complete elongation.[87] Although the nature of immature polyketide dissociation from KS/CLF remains to be determined, the KS/CLF substrate channel can fit a 16-carbon chain (the natural product of act KS/CLF) perfectly, with CLF defining the bottom of the channel (**Figure 10(b)**). If the channel size controls the polyketide chain length, it should be able to change the substrate specificity of KS/CLF (hence the polyketide chain length) by mutating CLF residues, thus changing the KS/CLF substrate channel size.

Sequence analyses of CLF from different type II PKSs and the KS/CLF structure predicted that there are potentially three gates at the bottom of the substrate channel that may control the chain length. All gates are closed in the 16-carbon (octaketide) KS/CLF. Smaller residues replace T112′ and F116′ in 20-carbon (decaketide) CLFs such as the Tcm KS/CLF, thus enlarging the substrate channel, relative to act KS/CLF. In 24-carbon (dodecaketide) CLFs, W194′ and F112′ are replaced by smaller residues, whereas G1953′ and L143 (of KS) are replaced by larger residues to form the end of the tunnel. In tridecaketide synthases, only L143′ is replaced by a larger residue. Therefore, it was hypothesized that these residues at the KS/CLF dimer interface modulate the length of the putative polyketide channel and determine the chain length. Indeed, the triple alanine mutation of F109′, T112′, and F116′ is sufficient to convert an octaketide synthase (16C, the act KS/CLF) into a decaketide synthase (20C, the Tcm KS/CLF), and vice versa.[55] The study successfully identified residues important in act and Tcm KS/CLF that determine the chain length, and also show that in act and Tcm KS/CLF, mutation of CLF is sufficient to change the product chain length.

1.08.4.4 The Ketoreductase

The actinorhodin KR (act KR) from *S. coelicolor* reduces the C9 carbonyl group of a 16 carbon (octaketide) polyketide chain (**Figure 6(a)**, **11**), which folds into the C7–C12 first ring cyclized product mutactin (**Figure 6(a)**, **22**) (**Figure 11(d)** provides details of the act KR-catalyzed reaction). The ketoreduction catalyzed by act KR, as well as by other type II PKS KR enzymes, is chemically similar to the FAS KR (**Figure 3**), yet with very different regiospecificities.[57] Despite conservation of the catalytic tetrad, the molecular details influencing the stereo- and regiospecificity of reduction catalyzed by type II KRs such as act KR have remained a mystery. For example, the high specificity of type II KRs for only the C9 carbonyl group (such as the KRs biosynthesizing act,[5] doxorubicin,[58] and enterocin)[56,87] suggest that type II PKS KR substrate intermediate is quite different than the linear FAS KR substrate, though the molecular basis for the C9 regio- and stereospecificity has not been determined.[88] Results showing that the inclusion of KR often results in products that are cyclized between C7 and C12[50,59] followed by aromatization of the first ring and cyclization of the second ring by the action of a didomain ARO/CYC hint that the actual KR substrate may be already cyclized although the timing of first and second ring formation (whether before or after KR reduction) is

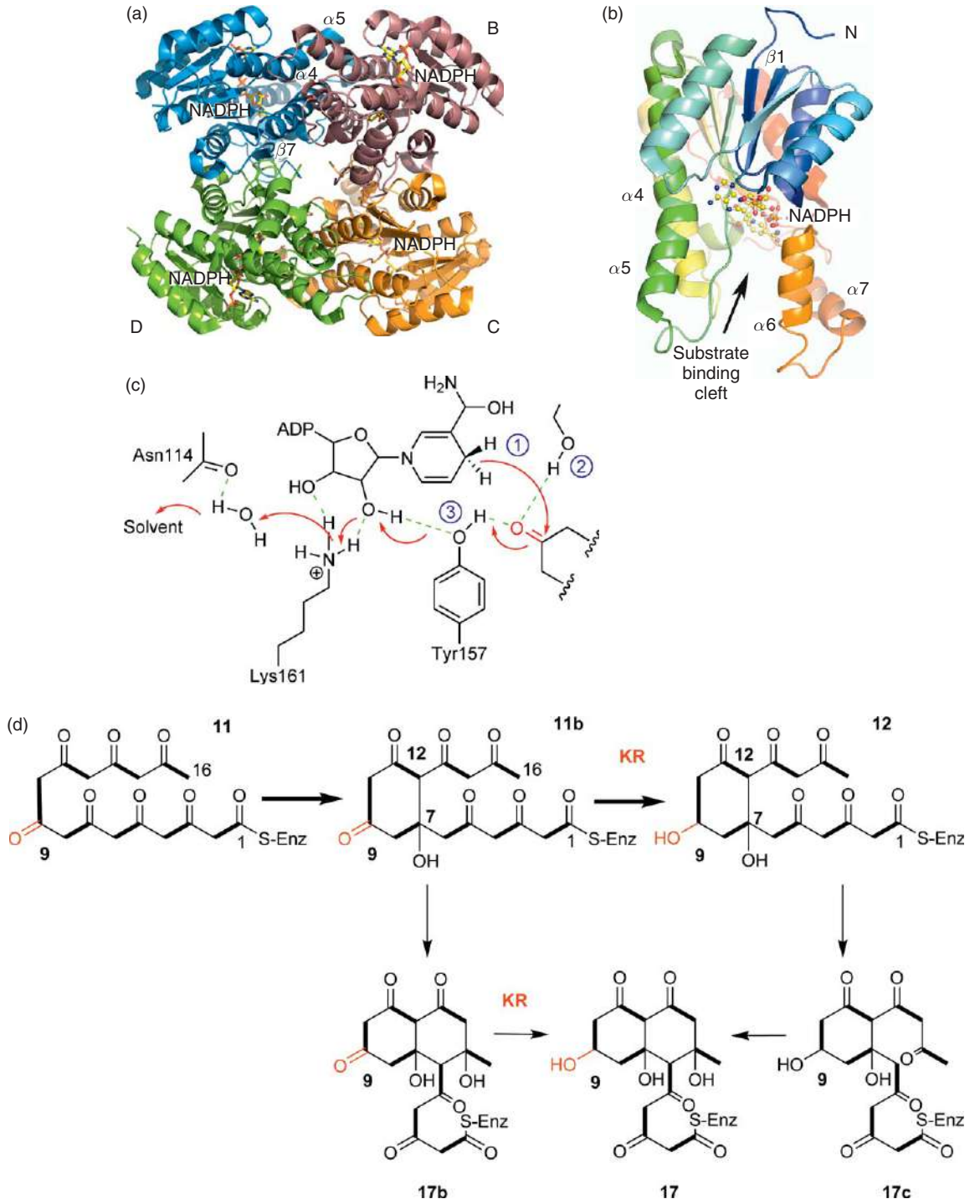


Figure 11 (a) Act KR is a homotetramer, each monomer consists of a short-chain dehydrogenase/reductase (SDR) fold. (b) The substrate binding pockets are defined by the α4–α5 and α6–α7 regions. (c) Proposed proton relay mechanism of act KR. (d) Proposed natural substrates of act KR.

unknown. Finally, whether the action of the didomain ARO/CYC is dependent on the stereochemistry of the C9 position is also not determined.

1.08.4.4.1 Crystal structure of the actinorhodin ketoreductase

The cocrystal structures of act KR bound with cofactors NADPH or $NADP^+$ help elucidate some of the above enzymological questions. Act KR exists as a tetramer (**Figure 11(a)**).[89,90] Act KR has a highly conserved Rossmann fold, which consists of two right-handed α–β–α–β–α motifs that are connected by $\alpha 3$, and the core region consists of a seven-stranded β-sheet blanked by α-helices (**Figure 11(b)**). The cofactor NADPH is bound at the junction of two α–β–α–β–α motifs in a highly conserved groove, while the polyketide substrate binding pocket consists of a large cleft (15 Å in width, 19 Å in length, 17 Å in depth) formed by helices $\alpha 6$–$\alpha 7$ and the loops between $\alpha 4$ and $\alpha 5$ (**Figure 11(b)**). The act KR structure shows that all sequence motifs important for the short-chain dehydrogenase/reductase (SDR) fold are conserved, such as the TGxxxGxG motif (12–19), D63, and the NNAG motif (89–92). The active site tetrad N114, S144, Y157, and K161 (**Figure 11(c)**), the active site N183, the PG motif (187–188), and T192 are also highly conserved. The biggest difference between the type II PKS KRs and other SDRs (such as FabG[91] and tropinone reductase[92]) is a 10-residue insertion (residues 199–209) between helices 6 and 7 (**Figure 11(b)**), and this may account in part for the different substrate specificity between type II PKS KR and other SDRs. There are extensive communications between monomers A–D in the *E. coli* FabG enzyme. In comparison, the twofold symmetric axis between monomers A and B in act KR is tilted approximately 3° from the axis in *E. coli* FabG.[28] The differences at the quaternary structural level between fatty acid and polyketide KRs may also influence the substrate specificity and the allosteric behavior of type II PKS KRs.

1.08.4.4.2 Proposed mechanism of ketoreductase

The act KR active site tetrad, N114–S144–Y157–K161, is near the nicotinamide ring of NADPH, where Y157 and K161 form hydrogen bonds with the NADPH ribose and nicotinamide ring, and four crystalline water molecules form extensive hydrogen bonds with N114 and K161 (**Figure 11(c)**). This extensive proton relay water network is very similar to the one observed in FabG–$NADP^+$.[91] Due to the similarity between the active sites of act KR and *E. coli* FabG, the proposed water relay mechanism for FabG is presumed to be applicable act KR, in which the ketone substrate, possibly from an enolized phenol intermediate, is hydrogen bonded to both S144 and Y157 that form the oxyanion hole. Following hydride transfer from NADPH to the ketone substrate, the resulting alkoxide is stabilized by the oxyanion hole, while the tyrosyl proton is transferred to the alkoxide. An extensive proton relay then takes place to replenish the proton extracted from the tyrosyl–OH, sequentially including 2-OH of NADPH ribose, lysine-NH followed by the four water molecules (**Figure 11(c)**). A search through literature, Protein Data Bank and sequence comparison, proposed that the proton relay network may be a conserved feature in both FAS and PKS KRs.

Genetic experiments in *Streptomyces* hosts have implicated that act KR may promote the polyketide first ring cyclization (**Figure 11(d)**).[59,93,94] A wide range of potential substrates were screened, including linear, monocyclic, and bicyclic polyketide mimics. It was found that act KR has a clear, strong preference for bicyclic compounds, such as *trans*-1-decalone (**35**), *trans*-2-decalone (**36**), and the aromatic α-tetralone (**37**) (**Figure 12(c)**). This result shows that the type II PKS KRs have different substrate specificity from that of FAS[95,96] and type I PKS KRs,[97] which can reduce linear and monocyclic ketones. It was further found that bicyclic aromatic polyketides such as emodin (**38**) (**Figure 12(c)**) are strong inhibitors of act KR. Therefore, the strong preference of act KR for bicyclic polyketides support that its natural substrate is likely a cyclic polyketide or a tautomer dehydrated at C7 (**Figure 11(d)**, **11a**, **17a**). Using the bicyclic substrate *trans*-1-decalone and the bicyclic inhibitor emodin, detailed steady-state and inhibition kinetics further confirmed that KR proceeds through an ordered Bi Bi mechanism, in which the cofactor $NADP^+$ binds KR prior to the substrate 1-decalone.[98] Furthermore, product release is the most likely rate-limiting step, as shown by viscosity dependence of k_{cat} and k_{cat}/K_m.

An important question for aromatic polyketide biosynthesis is the timing of the first ring cyclization relative to ketoreduction. Extensive docking simulations showed that act KR can dock with every carbonyl group pointed toward the cofactor NADPH, if the substrate is a linear polyketide (**Figure 6(a)**, **11**). However, this is

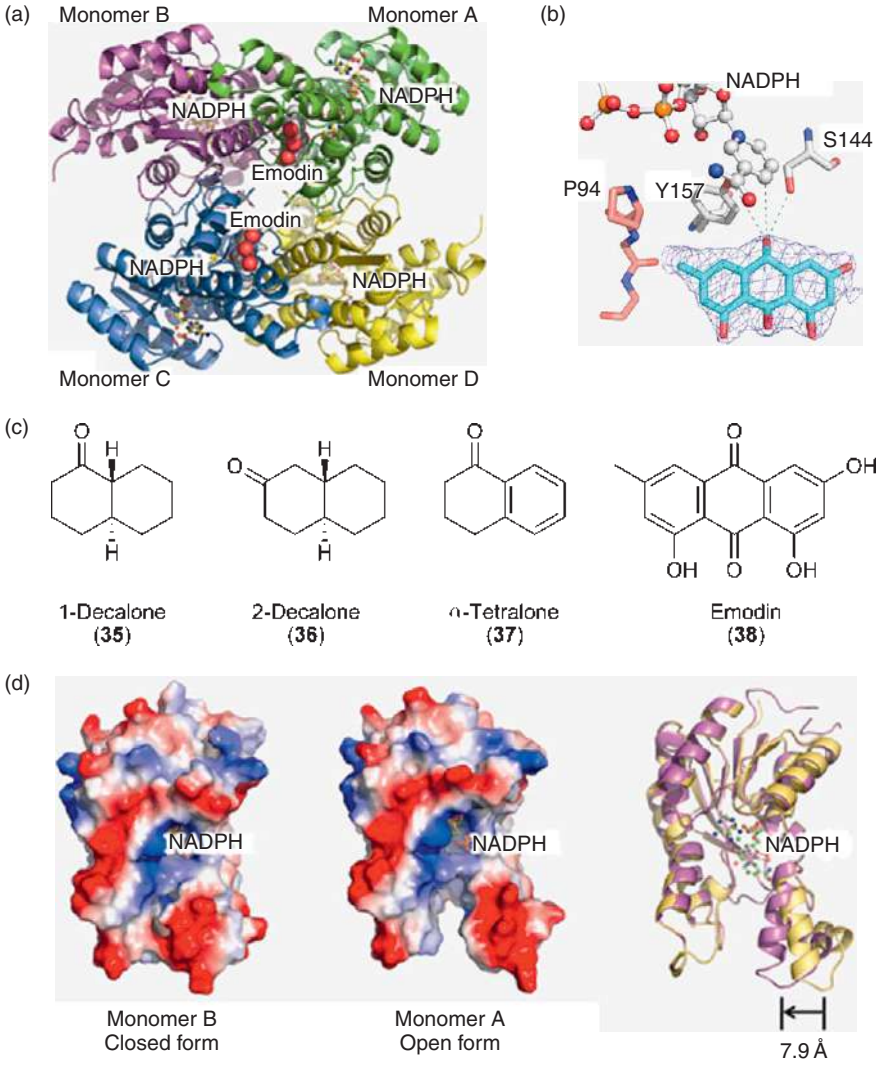


Figure 12 (a) The tetramer (in different colors) cocrystal structure of wild-type act KR–NADPH–emodin; NADPH (in ball and sticks) is bound to all four monomers; emodin (in spheres) can be found in monomers A and C. (b) The carbonyl group of emodin is hydrogen-bonded to the active sites S144 and Y157, and within hydride transfer distance of the cofactor, supporting the proposed three-point docking mechanism. (c) *In vitro* substrate of act KR. (d) The asymmetric unit contains monomers A (middle panel) and B (left panel) in open and close conformations, respectively. Surface potentials were colored from negative (red) to positive (blue). When the open and closed conformations are overlapped (right panel, open in yellow, close in purple), the major conformational change is in the flexible α6–α7 region, with a 7.9 Å difference between the two monomers.

not consistent with the observed C9 reduction specificity. On the other hand, if a bicyclic polyketide is docked into act KR (**Figure 11(d)**, **17b**), the dual constraints imposed by the ring structure and three-point docking would position the C9 carbonyl group optimally for reduction. Therefore, the docking simulation result implies that cyclization precedes reduction in act KR.

The proposed mechanism is supported by mutations, inhibitor-bound structures, and docking simulations. The alanine mutations of the active site tetrad abolished enzyme activity. Furthermore, with the exception of Y157F, all other active site mutants cannot bind the cofactor NADPH, as detected by the loss of fluorescence increase from NADPH–enzyme interactions.[99] This result confirmed the importance of the active site tetrad (Y157–S144–K161–N114) for enzyme activity and cofactor binding. The inhibitor emodin-bound act KR structure showed that the carbonyl group of emodin is hydrogen bonded to the active site S144 and Y157,

and within hydride transfer distance of the cofactor, supporting the proposed three-point docking mechanism (**Figures 12(a)** and **12(b)**).

The emodin-bound act KR structure also revealed that act KR can exist with at least two different conformations: the open and closed forms (**Figure 12(d)**). Within the act KR tetramers, only monomers A and C (**Figure 12(a)**) are bound with emodin, and adopt an open conformation. In comparison, monomers B and D have a closed conformation that precludes the binding of substrates or inhibitors. The major conformational difference between the open and the closed form is the 10-residue loop insert (residues 199–209) between helices 6 and 7 (**Figure 12(d)**). This insert is a unique feature in type II PKS KR (not observed in type I PKS or FAS KRs). This loop region also defines half of the substrate-binding pocket, and the observed conformational flexibility in the 10-residue insertion loop is proposed to influence the KR substrate specificity. The observed opening and closing of the act KR pocket is further proposed to reflect substrate- and product-binding conformations.

1.08.4.4.3 Proposed molecular basis of ketoreductase substrate specificity

Act KR (and by inference based on sequence alignment, other type II PKS KRs) strongly prefers bicyclic substrates *in vitro* (**Figure 12(c)**). Previously, it was also observed that two other SDR proteins, the fungal 1,3,8-trihydroxynaphthalene reductase (T_3HNR) and 1,3,6,8-tetrahydroxynaphthalene reductase (T_4HNR), also strongly prefer sterically constrained substrates, and they share 30 and 25% sequence identity with act KR, respectively.[100–102] The substrates of T_3HNR and T_4HNR are structurally similar to the first ring cyclized product in act KR biosynthesis (**Figure 6(a)**, **11**).[100] Therefore, in the absence of downstream ARO and CYC enzymes, act KR may reduce an intermediate with one or two rings cyclized. Based on the observation that the emodin-bound act KR structure displays a bent emodin (the bending was presumably due to the time averaging of different emodin tautomers), the actual substrate for act KR may be a tautomerized form of the monocyclic or the bicyclic intermediate.

Among the bicyclic substrates, act KR shows a distinct preference for *trans*-1-decalone over *trans*-2-decalone, while the aromatic α-tetralone is a much poorer substrate than either decalone. This is a surprise finding, because the carbonyl group of 2-decalone mimics the natural C9 ketone (**Figure 11(d)**). If it is assumed that the first ring (C7–C12) cyclization occurs before reduction of the C9 carbonyl of **17b**-1-decalone. A possible explanation why the opposite trend was observed is due to different binding motifs of 1- and 2-decalone, which is without the PPT tether of the natural substrate **11b** or **17b** (**Figure 11(d)**). Docking simulation showed that 1-decalone is consistently docked in a productive motif, while 2-decalone has several binding motifs that are nonproductive. The importance of substrate adaptation in the act KR substrate pocket is also supported by the fact that the more rigid α-tetralone is a much poorer substrate than decalones. Furthermore, besides lacking the PPT tether, both decalone and tetralone also do not have multiple carbonyl groups (as in **11b** and **17b**) and lack resonance that would allow the substrate to form hydrogen bonds with the KR substrate pocket residues. Indeed, when PPT-tethered **11b** and **17b** were docked into the act KR substrate pocket, the PPT group consistently docked into a surface positively charged patch containing three arginines, with multiple carbonyl groups forming hydrogen bonds with the active site residues. The above analyses support that the act KR substrate specificity is defined by a combination of enzyme conformation, specific molecular interactions between the substrate, and active site residues, as well as substrate and protein flexibility. Due to the dynamic nature of the binding cleft, it should be possible for KR to be altered in a way to accept substrates with variable chain lengths or cyclization patterns.

1.08.4.4.4 The stereochemistry of ketoreductase

For the type I PKS KRs, the first motif analysis was described by Celmer in 1965 by the analysis of the reduced polyketide stereochemistry.[103] Type I KRs are designated A-type and B-type KRs depending on the presence of either a Trp at position 149 or LDD residues at positions 94–96 (act KR numbering) that lead to the '*S*' or '*R*' reduced product, respectively.[104] The major determinant of stereochemistry at the resulting β-hydroxyl has been attributed to the second aspartic acid in 'LDD', which is invariant in KRs that produce the *R* isomer and absent in those that produce the *S* isomer. The role of the LDD motif also extends to other homologous SDR proteins. In comparison, nearly all type II KRs contain a unique motif corresponding to 94-XGG-96 (**Figure 13(a)**). In the case of act KR, this motif corresponds to 94-PGG-06. To complicate the issue, act KR

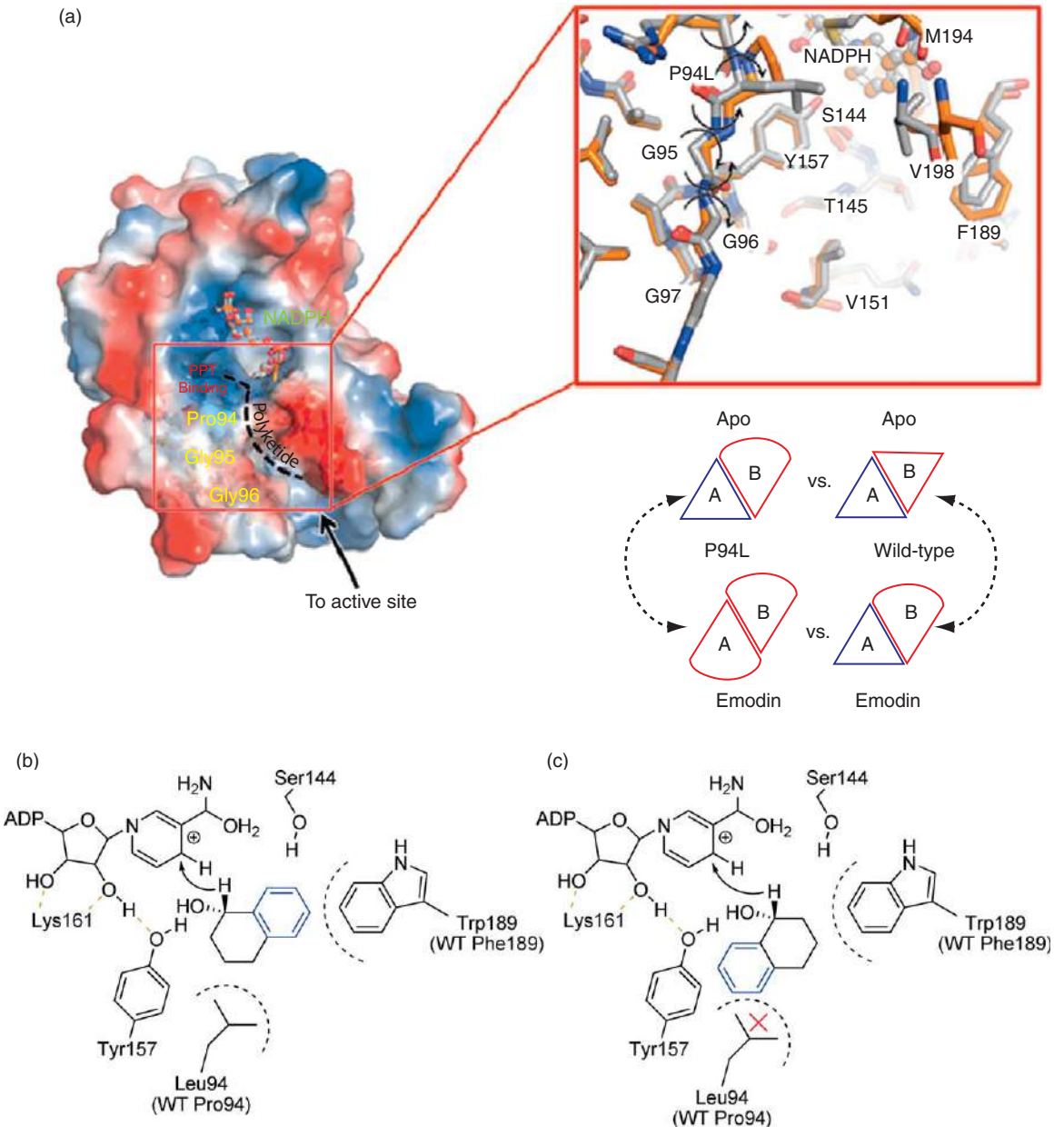


Figure 13 (a) Overlay of the act KR wild-type (WT) and P94L mutant structures. Left panel: surface representation of the P94L mutant protein and the location of the PGG motif in relation to NADPH (orange ball and stick), the predicted PPT binding site (red letters), and the direction of polyketide binding into the active site (dashed line). Right panel: close-up view of the overlaid WT (orange) and P94L (gray) structures. Curved arrows suggest that the P to L mutation allows the LGG loop more rotational freedom. (b and c) Docking of *S*-(+)- or *R*-(–)-tetralol to the P94L mutant.

may be one of the few SDR members that do not display stereospecificity. To analyze if the 'PGG' sequence in act KR can influence its stereospecificity, a series of act KR mutants were generated that systematically changes P94, G95, and G96 to L94, D95, and D96 via single, double, and triple mutations. For the *in vitro* enzyme assay, the oxidation of *S*- and *R*-tetralol to tetralone was conducted. Wild-type (WT) act KR was found to have a threefold preference for *S*-tetralol over *R*-tetralol. Significantly, the single P94L mutant resulted in the complete abolishment of *R*-tetralol oxidation activity, while retaining full catalytic capacity for *S*-tetralol.

Heterologous expression of WT or P94L act KR, followed by a purification of the product mutactin over chiral high-performance liquid chromatography (HPLC) column, also consistently showed that WT act KR has a threefold preference of one stereomer versus the other, while the P94L mutant predominantly produces only one stereomer.

A structural overlay of the P94L structure with solved SDR structures shows that (1) the position and orientation of the loop containing the PGG (or LDD) motif is similar in all of these structures and (2) the main difference is the location of helix $\alpha 6$ that forms the lid region of the substrate pocket (**Figure 12(d)**). For the type I KR1, it was hypothesized that the presence of LDD in B-type KRs influences the orientation of the polyketide substrate by interacting with the hydrophobic portion of the ACP-bound PPT group. However, the PPT group also binds to the lid region, which is an approximately 11 residue insertion present in all SDR proteins except FabG. The absence of this insertion may contribute to the FabG's lack of specificity, and FabG reduces every β carbonyl group of the growing fatty acid chain. In the WT enzyme, without larger and bulkier leucine, the WT act KR can accommodate both *S*-(+)- and *R*-(−)-tetralol. When the P94 is changed to L94, steric interactions preclude oxidation of the *R*-isomer (**Figure 13(c)**). This implies that the stereospecific reduction of the ACP-bound polyketide is influenced at least partially by steric constraints imposed by pocket residues.

Previously, Kalaitzis and Moore[87] proposed that act KR accepts two possible substrates, the full-length octaketide (16 carbons) or a truncated pentaketide (10 carbons), that undergo subsequent rounds of elongations to produce a reduced hexaketide. However, docking of putative act KR substrates with either WT or P94L mutant structures suggests that a minimum of 12 carbons is necessary to position the C_5═O in the oxyanion hole, assuming the PPT group is anchored in the arginine patch (R38, R65, and R93), as previously proposed.[98] This region has also been identified in the homologous SCO1815 KR from *S. coelicolor*[105] and would restrict the attached polyketide to a single side of the pocket. This suggests that similar to KS/CLF, the act KR physically counts the carbon chain length, starting from the end of the PPT group (which is anchored to the same arginine patch), and the chain length is determined by the physical distance between the PPT group and the KR active site. This hypothesis helps explain how act KR can reduce the C9 position, even when presented with longer substrates such as C18 or C20. In the future, more work needs to be conducted to determine the timing of first ring cyclization versus chain elongation, as well as the true physical location of polyketide cyclization.

1.08.4.5 The Aromatase/Cyclase

The aromatic rings are essential for the antibiotic and anticancer activity of polyketides.[37] The cyclization and aromatization of the polyketide rings are key steps that control polyketide diversity. For example, doxorubicin and mithramycin are both biosynthesized from a 20-carbon linear polyketide;[7,9] however, doxorubicin and mithramycin are produced by cyclization at C7–C12 versus C9–C14, respectively.[58,59] Studies in the Khosla, Moore, and Salas groups have demonstrated that KR inclusion results in products that are cyclized at the C7–C12 positions,[50,56,59] whereas the inclusion of a monodomain ARO/CYC (in the absence of KR) results in products that are cyclized at the C9–C14 positions.[49,59] Based on previous studies of first ring cyclization, there are at least three different classes of ARO/CYCs (**Figure 14(a)**): (1) C9–C14 cyclization is associated with monodomain ARO/CYCs such as Tcm ARO/CYC or WhiE ARO/CYC.[6,42,45,106] (2) C7–C12 cyclization in the absence of KR by monodomain ARO/CYC such as ZhuI in the R1128 biosynthetic pathway.[78] (3) Didomain ARO/CYCs, which are associated with KR-containing type II PKSs that produce a C7–C12 first ring cyclized polyketide, such as the act[49] and griseusin (gris)[42] ARO/CYCs. Although the cyclization specificity is different, these three classes of ARO/CYCs have a sequence identity of 35–85%. Furthermore, ARO/CYC was proposed to be a DH that catalyses the dehydration of the polyketide hydroxyl groups following aldol cyclization.[42] If this is correct, ARO/CYC is a novel DH whose sequence and structure are not homologous to any known DHs.

The best studied monodomain ARO/CYC is the Tcm ARO/CYC of *S. glaucescens*, which consists of the N-terminal 160 residues of the bifunctional protein TcmN. The inclusion of either TcmN or Tcm ARO/CYC *in vivo* or *in vitro* (in the absence of KR) strongly promotes C9–C14 first ring (and C7–C16 second ring) cyclization[107,108] (**Figure 14(b)**). Following the production of the linear decaketide intermediate, Tcm ARO/CYC is proposed to fold, cyclize, and aromatize the ACP-tethered polyketide via aldol condensation and

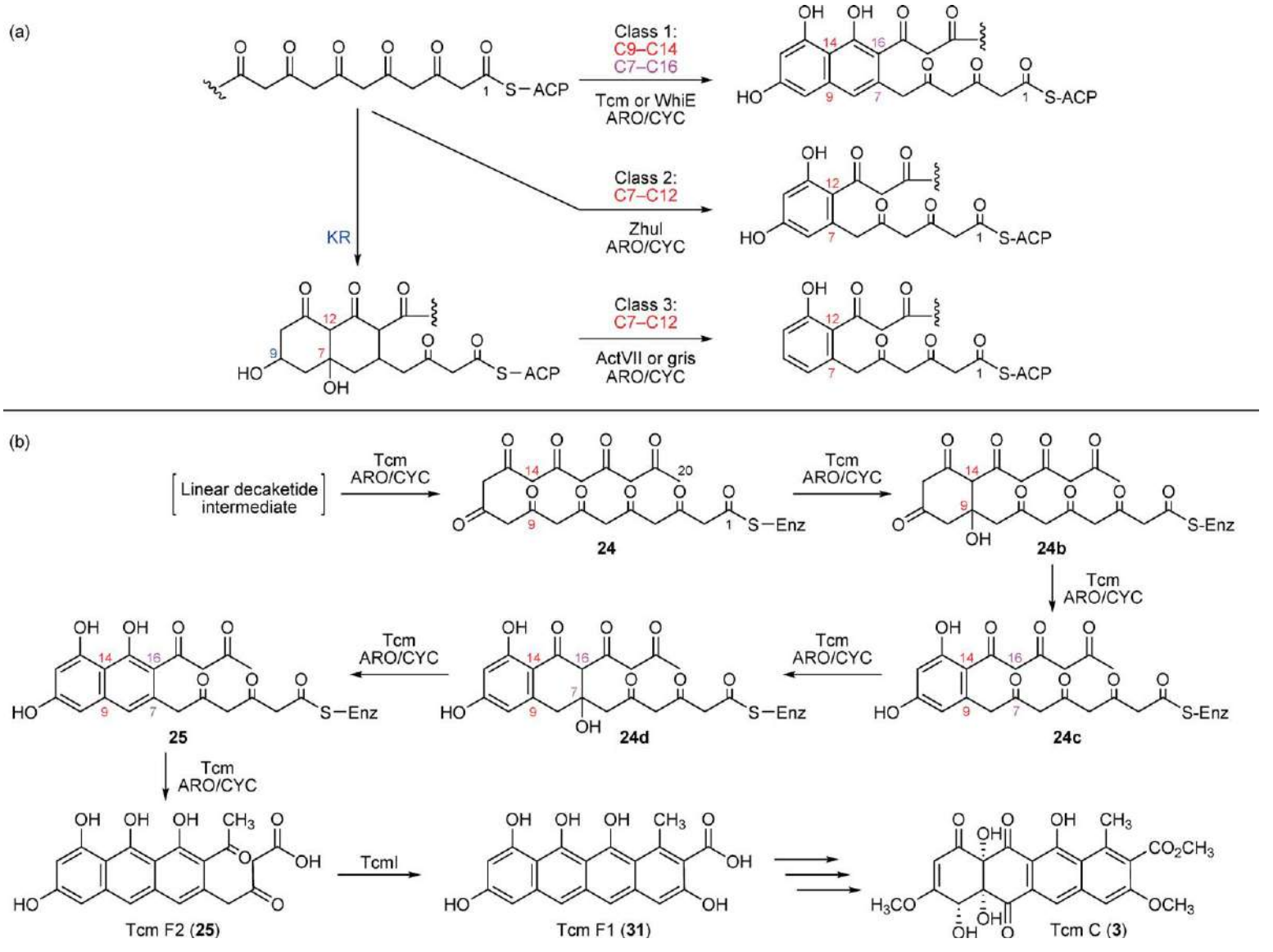


Figure 14 (a) Three types of ARO/CYCs. (b) Tandem reactions proposed to be catalyzed by Tcm ARO/CYC en route to the final natural product Tcm C.

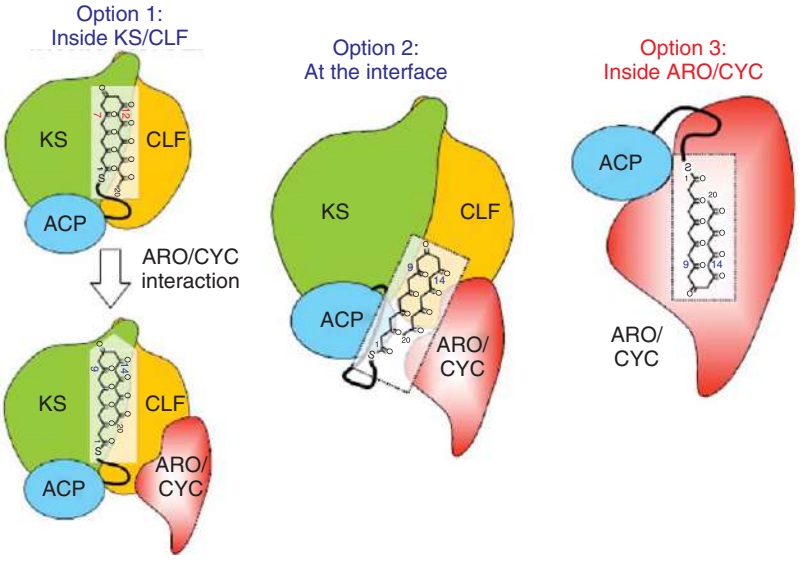


Figure 15 Different scenarios that explain how Tcm ARO/CYC may promote C9–C14 first ring cyclization.

dehydration reactions.[109] Tcm ARO/CYC is proposed to promote C9–C14 first (**24b**) and C7–C16 second ring cyclizations (**24d**), with the third ring forming spontaneously in its natural host to produce Tcm F2 (**26**), which is then cyclized by TcmI (a cyclase that is structurally and functionally distinct from Tcm ARO/CYC, see the next section) at the fourth ring to produce Tcm F1 (**31**).[110] How Tcm ARO/CYC promotes so many tandem reactions and achieves the C9–C14 cyclization specificity is not well understood, and it is unclear if ARO/CYC ever actually binds the polyketide substrate, or if it instead stabilizes protein–protein interactions between KS/CLF and ACP for cyclization to occur at the interface.

The timing of the first cyclization event is a key issue of polyketide biosynthesis. Based on genetic analyses, it had been argued that the first cyclization of a linear polyketide chain may occur either in the active site of KS/CLF,[86] in the solution (without enzyme catalysis),[56] or in the binding pocket of KR or ARO/CYC.[62] However, the presence of the act KR results in the dominant formation of C7–C12 product.[49,59] Similar observations in more than 10 aromatic PKSs (except in enterocin discussed below[56]) have led to the general conclusion that KR promotes C7–C12 cyclization,[35] whereas the monodomain ARO/CYCs in the nonreducing type II systems promote C9–C14 cyclization. There are *three possible scenarios* to interpret how ARO/CYC promotes C9–C14 cyclization (**Figure 15**).

Scenarios 1 and 2. The cyclization occurs in the KS/CLF active site, or at the junction of KS/CLF-ACP-KR,[86] and the inclusion of ARO/CYC causes a conformational change in the substrate pocket of KS/CLF to promote C9–C14 cyclization. The substrate pocket of KS/CLF can be separated into two halves: (a) the polyketide binding half is formed by KS and CLF. Inspection of the act KS/CLF binding half reveals that this narrow region can only accommodate a linear polyketide chain (**Figure 10(b)**); (b) the PPT binding half binds the PPT group of ACP and extends from the KS surface to active site Cys.[86] If the cyclization occurs in the polyketide binding half of KS/CLF, it would require an extensive protein conformational change and a high-energy penalty. Such a conformational change has not been observed in any fatty acid or polyketide KS enzymes.[111,112] Furthermore, sequence alignment of KSs indicate that regardless of the cyclization patterns (C7–C12 or C9–C14), the pocket residues in these KSs are highly conserved in both the polyketide and the PPT binding halves, except at the protein surface. Therefore, if KS/CLF is involved in the first ring cyclization, this location needs to be close to the entrance of the substrate pocket, at the protein–protein interface between KS/CLF, ACP, and ARO/CYC.

Scenario 3. The first cyclization occurs in the ARO/CYC active site. The linear polyketide chain is transferred from the active site of KS/CLF to that of ARO/CYC via the PPT arm of ACP, which serves as the 'Swiss Knife

Shear' that holds the active polyketide intermediate in its inner cleft, as proposed in the 'Switchblade' hypothesis proposed by Ban and coworkers based on the fungal FAS crystal structure. This is followed by a fast transfer of the polyketide chain into the ARO/CYC active site. This scenario requires (a) a polyketide binding site in ACP, (b) a fast chain transfer from KS/CLF to ARO/CYC, via ACP as the 'Switch-Blade', and (c) a substrate binding pocket in the ARO/CYC.

As a related extension to Scenario 3, it may be that during chain elongation the growing intermediate can be transferred from KS/CLF to ARO/CYC, which cyclizes the polyketide intermediate during a certain elongation cycle that favors C9–C14 cyclization (rather than acting on the full-length chain), followed by subsequent completion of chain elongation in KS/CLF. The C9–C14 cyclization results from the polyketide transfer event between ARO/CYC and KS. This mechanism has been proposed for enterocin and act KR based on the observation of truncated products.[56,65,87] This mechanism may apply to enterocin, whose KS/CLF must have a larger pocket to accommodate the phenyl moiety[56] and cyclization can take place in between chain elongation. However, for other aromatic PKSs, the narrow polyketide binding half region of KS/CLF cannot accommodate the ring intermediate that is presumably cyclized in between chain elongation.

Note that the above possibilities can also apply to the C7–C12 specific cyclization associated with act KR. The crystal structure and mutagenesis of Tcm ARO/CYC described below offer support that Scenario 3 may be the likely cyclization event. However, further validation is necessary to rule out any of the above four possibilities.

1.08.4.5.1 Crystal structure of Tcm ARO/CYC

The 1.9 Å crystal structure of Tcm ARO/CYC[113] reveals that this enzyme possesses a helix-grip fold,[114,115] which consists of a seven-stranded antiparallel β-sheet that surrounds a long C-terminal α-helix. Two small helices between β1 and β2 form a helix–loop–helix motif that seals one end of the β-sandwich (**Figure 16(a)**). Despite low sequence homology, the topology of Tcm ARO/CYC is remarkably similar to members of the Betv1-like superfamily (11–14% identity).[116] Members of the Betv1-like superfamily bind small molecules such as phytosteroid hormones, lipids, enediyne, and cholesterol,[116–119] and include the birch pollen allergen Betv1[117] and the MLN64-START domain.[116] Similar to other Betv1-like superfamily members, Tcm ARO/CYC has a deep interior pocket formed between the β-sheet interior and the C-terminal α-helix that is ideally suited to bind multicyclic compounds (**Figure 16(c)**).

Due to the close relationship of FAS and PKS, one may wonder if there is an enzyme in type II PKS, equivalent to the DH domain in type II FAS. A detailed structural comparison of Tcm ARO/CYC versus the *E. coli* DH enzyme FabA shows that whereas Tcm ARO/CYC contains a helix-grip fold (**Figure 16(a)**), FabA contains a hotdog fold[26] (**Figure 16(b)**). In the FabA hotdog fold, the β-sheets have strand order 1–2–3–5–6–7–4, whereas the strand order is 1–7–6–5–4–3–2 for the helix-grip fold. Also, the central helix is tucked against the β-sheets in FabA, thus precluding the formation of a deep interior pocket. Instead, the active site of FabA is formed at the dimer interface. In comparison, the central helix of Tcm ARO/CYC is offset from strand 1, allowing formation of an interior pocket. The presence of the interior pocket, which is amphipathic and is defined by several highly conserved residues, supports a role for ARO/CYC in polyketide binding and direct involvement in cyclization (Scenario 3). This is further supported by mutagenesis results (below). Although topologically different, the helix-grip fold still shares a striking resemblance to the 'hotdog-in-a-bun' fold (**Figures 16(a)** and **16(b)**). For instance, both contain seven β-sheets that are wrapped around a central helix. That both ARO/CYC and DH catalyze dehydration reactions in evolutionarily related complexes further suggests that they may play similar biological roles implied by their similar topologies.

Consistent with previous studies,[108] the monodomain Tcm ARO/CYC is isolated as a monomer in solution. However, the addition of 0.1 mol l^{-1} NaI results in the formation of an iodide-bridged homodimer (**Figure 17(a)**). The NaI dimer is also of interest, because the full-length protein TcmN is a homodimer in solution.[109] For the type II FAS, two DH enzymes, FabA[26] and FabZ,[120] have been solved and analyzed in detail. Both FabA and FabZ contain single hotdog folds that come together to form an active dimer (**Figure 17(b)**). Each dimer is formed by the continuous β-sheet (seven strands from each monomer), plus helix–helix interactions from the two core helices. Similarly, the mammalian FAS contains a 'double hotdog' fold where two identical DH domains are fused together (**Figure 17(b)**). It is proposed that the didomain

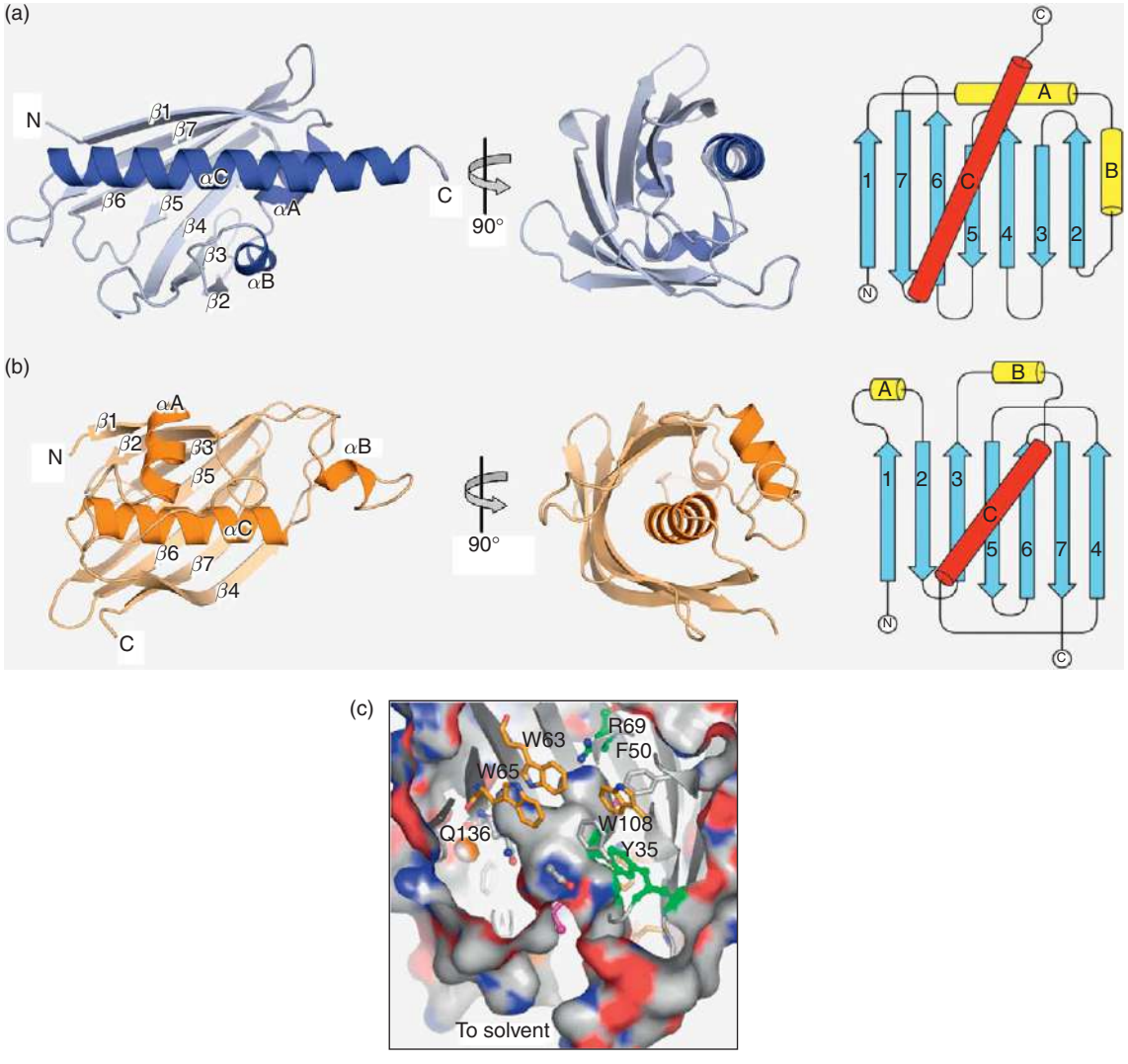


Figure 16 Structural comparison of Tcm ARO/CYC and *Escherichia coli* FabA. (a) Tcm ARO/CYC helix-grip fold. (b) The FabA hotdog fold. (c) The interior pocket in Tcm ARO/CYC.

ARO/CYCs (such as act ARO/CYC) contain two repeats of the ARO/CYC enzyme that presumably form a pseudodimer imitating the 'double hotdog' fold observed in the mammalian DH.[24,121] Interestingly, the overall architecture of the Tcm ARO/CYC–iodide dimer resembles the homodimeric 'double hotdog' fold from FabA/Z or the mammalian FAS DH.[122]

The DH domains of type I and type II FASs are both 'hotdog-in-a-bun' folds.[24,34] The recently solved mammalian FAS shows that the DH domain also plays a major role in megasynthase formation through an intimate association with the KS domain[24] (**Figure 17(c)**). The DH–KS association is facilitated by the fact that both proteins are tethered together on the same polypeptide, which also renders the type I FAS more efficient than the type II FAS. In the *E. coli* type II FAS, this 'loose' association allows the bacteria to produce a wide array of fatty acid products that are important for maintaining the biophysical properties of the membrane.[34] In contrast, both type II PKS and the mammalian FAS produce only a single product when all of the domains are present. As a result, it is possible that the megasynthase architecture of type II PKS may be more related to the type I FAS.

Although the type II PKS lacks structural linker regions as in type I FAS, the fidelity of aromatic polyketide biosynthesis suggests the formation of a megasynthase during polyketide biosynthesis. In the previous section,

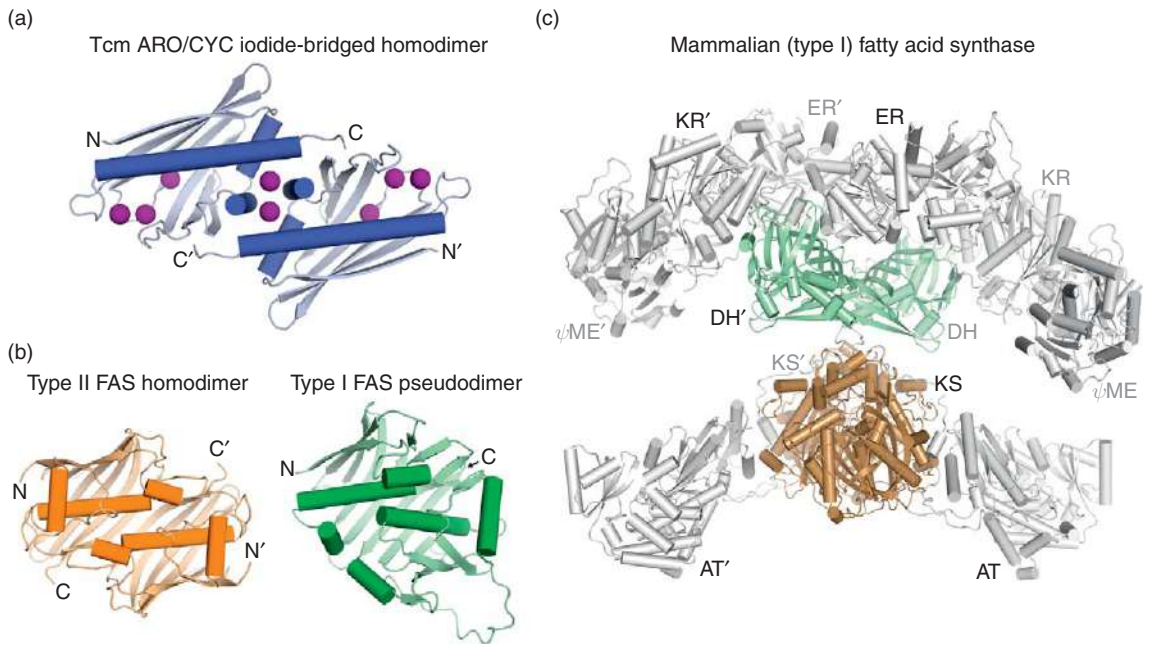


Figure 17 A comparison of the Tcm ARO/CYC dimer with the FAS DH domain. (a) Tcm ARO/CYC dimerizes across the α1–loop–α2 region in the presence of NaI. (b) The double hotdog fold of the FAS DHs: left, FabZ (*Escherichia coli* type II FAS) is dimeric; right, the DH domain of porcine FAS is psuedodimeric (mammalian type I FAS). (c) The 3.2 Å mammalian FAS crystal structure. The DH double hotdog forms a dimer that links the top (DH, KR, and ER) to the bottom (KS and MAT).

the similarities between the hotdog and helix-grip folds were compared. The structural similarity between DH and ARO/CYC suggests that ARO/CYC may interact with the KS/CLF enzyme analogous to the FAS DH–KS interaction. However, in the case of the mammalian FAS, DH needs to form a dimer (namely, four hotdog folds). This implies that four Tcm ARO/CYC monomers (a dimer of dimers) would need to come together to form a similar megasynthase architecture. This may be plausible, because in solution and crystal structure, act KR also exists as a tetramer. Furthermore, KS/CLF is a tetramer in the solution, as confirmed by protein sizing experiments.[1] Clearly, much more work is needed to elucidate the megasynthase architecture in type II PKS.

1.08.4.5.2 Proposed mechanism of Tcm ARO/CYC

Based on site-directed mutagenesis, computer-simulated docking, and the crystal structures of Tcm ARO/CYC, the following mechanism was proposed (**Figure 18**):[113]

(A) *First ring cyclization and aromatization.* (1) The polyketide C13 carbonyl oxygen, anchored in close proximity to R69, deprotonates the guanidinium proton. (2) The protonated C13 carbonyl is now more electrophilic, and the Y35 tyrosinate near the α carbon-14 facilitates enol formation. (3) The C13–C14 enol then collapses and undergoes nucleophilic attack on the C9 carbonyl carbon and cyclize the first ring (C9–C14), during which the C9 alkoxide is stabilized by the Y35 side chain. (4) Proton transfer of Y35–OH regenerates the tyrosinate. (5) The tryrosinate again facilitates enol formation at C14. (6) The C9 hydroxyl group is then protonated by Y35–OH. (7) The C13 enol collapses to eliminate the C9 water. (8 and 9) The conserved S67–OH can now deprotonate at C10, which stabilizes a delocalized cation. (10) In the final step of first ring aromatization, R69 can deprotonate C12, and the resultant C11 oxide can be quenched by the hydroxonium cation developed on S67 in steps 8 and 9.

(B) *Second ring cyclization and aromatization.* (11) The second ring cyclization begins with the Y35 tyrosinate deprotonating at C16 to facilitate enolate formation, which can be stabilized as a pseudo-six-membered ring. (12) The enolate collapses, and the C16 undergoes nucleophilic attack at C7 to close the second ring (C7–C16). (13) The resultant C7 hydroxyl is protonated by the proximal R82 guanidinium and becomes a

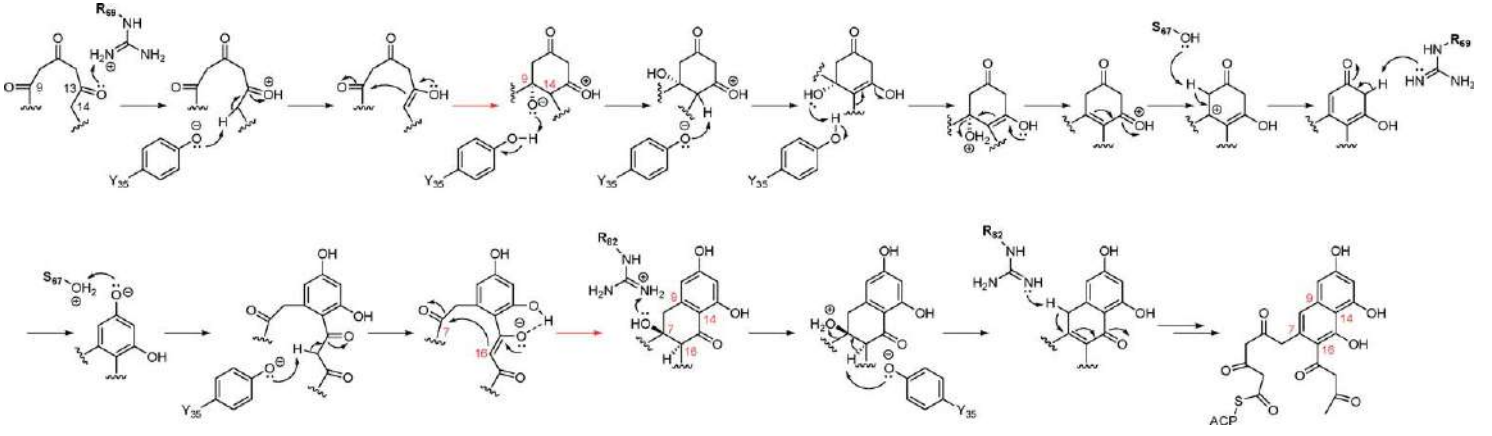


Figure 18 Proposed ARO/CYC mechanism. R82, R69, Y35, and the active site water of Tcm ARO/CYC pocket may serve as anchoring points to fold the polyketide chain and as general acids/bases for C9–C14 first ring and C7–C16 second ring cyclization/aromatization.

good leaving group. (14) A regenerated tyrosinate deprotonates at C16 to eliminate water at C7. (15) The neutral R82 guanidine deprotonates at C8, thus aromatizing the second ring.

(C) *Subsequent cyclizations*. Docking shows that the Tcm ARO/CYC pocket can accommodate up to three linearly fused rings, but not the fourth, because the intramolecular C19–C2 aldol condensation results in clashes with pocket residues of helix 3. Therefore, after third ring cyclization (likely spontaneous), the polyketide must be transferred to TcmI,[110] where fourth ring cyclization occurs to produce Tcm F1 (**31**).

In the case of didomain ARO/CYCs, the first ring cyclization has been proposed to occur prior to ARO/CYC action, either in upstream KS/CLF or KR.[123] Therefore, in the didomain ARO/CYCs, the pocket residues likely orient the intermediate for first ring aromatization but do not direct the first ring cyclization.

1.08.4.5.3 The cyclization specificity of ARO/CYC

To probe for the C9–C14 first ring cyclization specificity when Tcm ARO/CYC is included in the PKS complex, point mutations were generated for pocket residues in Tcm ARO/CYC (**Figure 16(c)**).[113] The inclusion of Tcm MinPKS + WT Tcm ARO/CYC in the heterologous host *S. lividans* produces the 20-carbon C9–C14 cyclized products RM80 (**29**) and RM80b (**30**) (**Figure 6(b)**). In comparison, with only the Tcm MinPKS, the dominant product is the 20-carbon C7–C12 cyclized product SEK15 (**32**) (**Figure 6(b)**). It was found that the Tcm ARO/CYC pocket residue mutants R69A, R69D, and Y35A completely abolished the production of the C9–C14 cyclized polyketide RM80 and supports Scenario 3 (above) where ARO/CYC plays an active role in cyclization. Docking simulations show that the highly conserved residues R69 and Y35, located at the bottom and neck of the pocket, anchor polyketides in the correct orientation for C9–C14 cyclization. When the polyketide forms hydrogen bonds with R69 and Y35, C7–C12 first ring cyclization will cause steric clashes between polyketides and side chains of pocket residues, whereas C9–C14 first ring cyclization does not cause such steric clash (**Figure 19**). The results of the mutational analysis for the Tcm ARO/CYC pocket residues show that changes in pocket can significantly alter polyketide product distribution, and strongly suggest that the polyketide chain binds to the interior pocket (Scenario 3), where R69 and Y35 anchor the polyketide chain. Other pocket-defining residues such as W108 and Q110 help orient the polyketide intermediate so that C9–C14 first ring and C7–C16 second ring cyclizations can occur. As a control experiment, the R69A crystal structure showed no conformational change to the Tcm ARO/CYC backbone or interior pocket side chains. Therefore, in Tcm ARO R69A, the protein–protein interactions between the ARO/CYC and the Tcm MinPKS should be maintained (reflecting cyclization Scenarios 1 and 2) and there is no mutation-induced rearrangement of pocket side chain packing that would affect product formation. Because R69A is located deep inside the pocket, the above result suggests that the deleterious affect of the R69A mutation on RM80 production (C9–C14 first ring cyclization) is residue-specific and may be due to a loss of anchoring potential, acid–base chemistry potential, or the loss of positive charge when this Arg residue is mutated to Ala. Taken together these data indicate that cyclization Scenarios 1 and 2 (protein–protein interactions) are not directly responsible for the observed influence of Tcm ARO/CYC on cyclization specificity but rather cyclization Scenario 3, in which the polyketide enters the interior pocket of ARO/CYC for cyclization, is the likely route for the observed influence of ARO/CYC to control cyclization specificity.

1.08.4.6 The Fourth Ring Cyclases

Following first and second ring cyclizations promoted by the ARO/CYC, presumably the third ring is formed spontaneously. This is then followed by fourth ring cyclization catalyzed by the fourth ring cyclase. For example, the TcmI cyclase promotes the formation of the fourth ring that transforms Tcm F2 to Tcm F1 (**Figure 6(b)**). The TcmI crystal structure showed that it has a dimeric β–α–β ferredoxin-like fold (four antiparallel β-strands, **Figure 20**), in which the dimer is assembled by swapping β4 at the dimer interface.[110] Similar to Tcm ARO/CYC, TcmI also forms a cavity between the antiparallel β-sheet and the α-helices. The tertiary structure and substrate cavity of TcmI is very similar to the act monooxygenase from *S. coelicolor*,

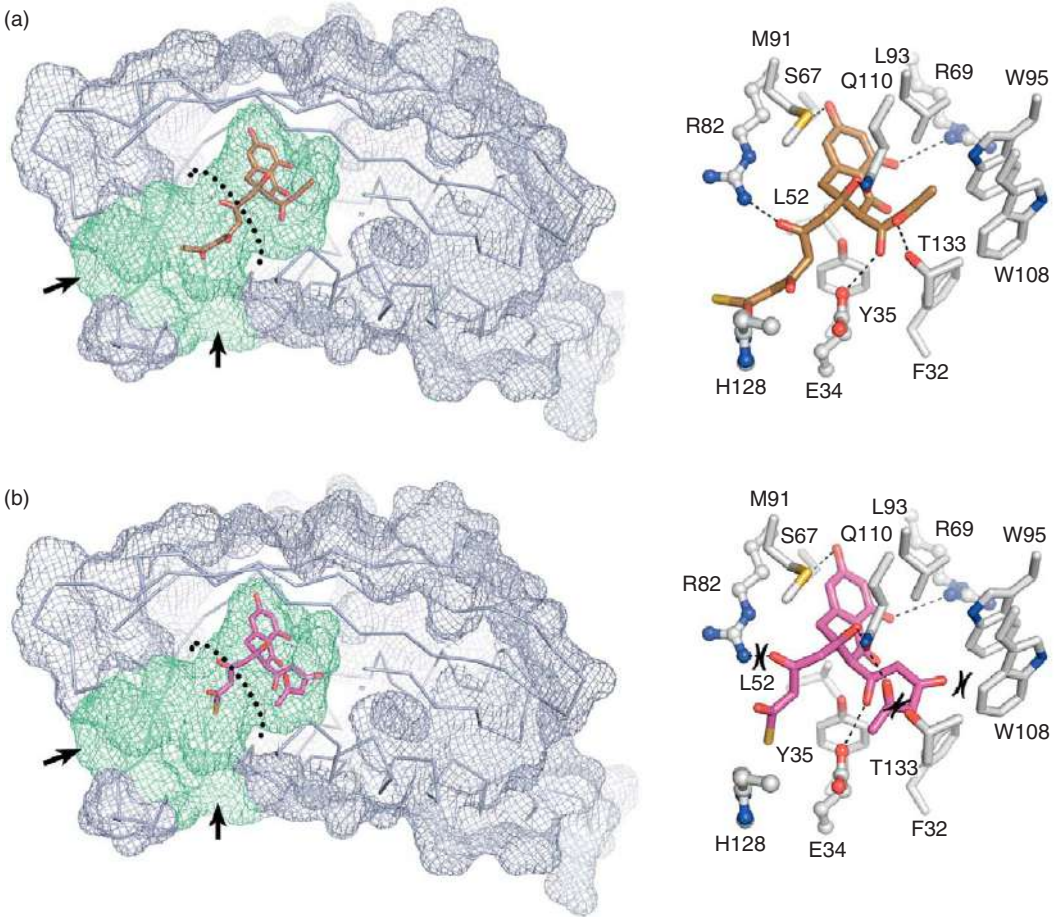


Figure 19 (a) Docking simulation of C9–C14 cyclized polyketide intermediate anchored at R69 and Y35 side chains. (b) In comparison, C7–C12 cyclized polyketide intermediate clashes with pocket residue side chains, if the polyketide is anchored at R69.

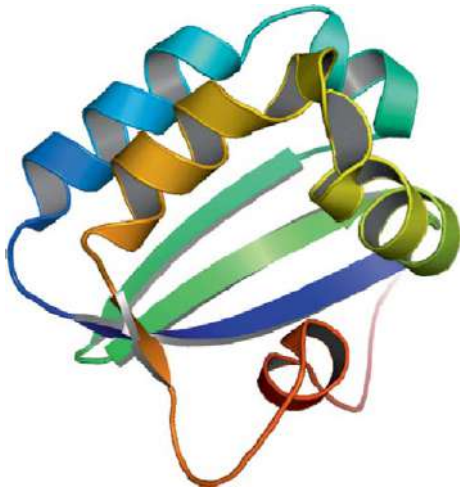


Figure 20 The fourth ring cyclase TcmI has a ferredoxin-like fold.

indicating that these late-stage PKS enzymes may be evolutionarily related. However, unlike Tcm ARO/CYC, mutations of pocket residues in TcmI do not abolish the fourth ring cyclization activity. This suggests that TcmI may serve as a 'cyclization mold' that brings the polyketide into its pocket, where the aldol cyclization may take place spontaneously, or the oxyanion can be stabilized by pocket water. Therefore, the pocket residues may help steer the fourth ring cyclization rather than providing a real active site. A similar fold to TcmI is also observed for other fourth ring cyclases such as SnoaL[124] and AknH[125] in *Streptomyces galilaeus*, but they differ from TcmI as a true enzyme, and the mutation of an active site Asp abolished cyclization activity. It was concluded that similar to the case in Tcm ARO/CYC, residues 15 and 51 in these fourth ring cyclases orient the C9 oxygen of the substrate via hydrogen bonds, which then help determining the product stereochemistry.

1.08.4.7 Acyl Carrier Protein

ACPs from FAS and PKS are notoriously difficult to crystallize. Except the report of butyryl-ACP from *E. coli*, commercial crystallization screening of apo-, holo-, or acyl-ACPs often resulted in phase separations. The small size of ACP (100 residues), combined with the difficulty in crystallization, rendered nuclear magnetic resonance spectroscopy (NMR) the choice for observing ACP structures. In fact, the NMR solution structure of act ACP (*act* apo-ACP) from *S. coelicolor* in 1997 represents the first PKS component for which detailed structural information was obtained.[35] The act ACP structure is composed of four helices, with helix 3 shorter than the other three, and a large loop region between helices 1 and 2. The act ACP protein fold is very similar to the *E. coli* FAS ACP (**Figures 21(a) and 21(b)**). Later in 2003, both the fren[126] and otc[127] ACP solution structures also displayed a similar fold. When comparing the FAS and PKS ACPs, the *act* apo-ACP has a disordered N-terminus and a longer flexible loop than the FAS ACP. Furthermore, act ACP contains

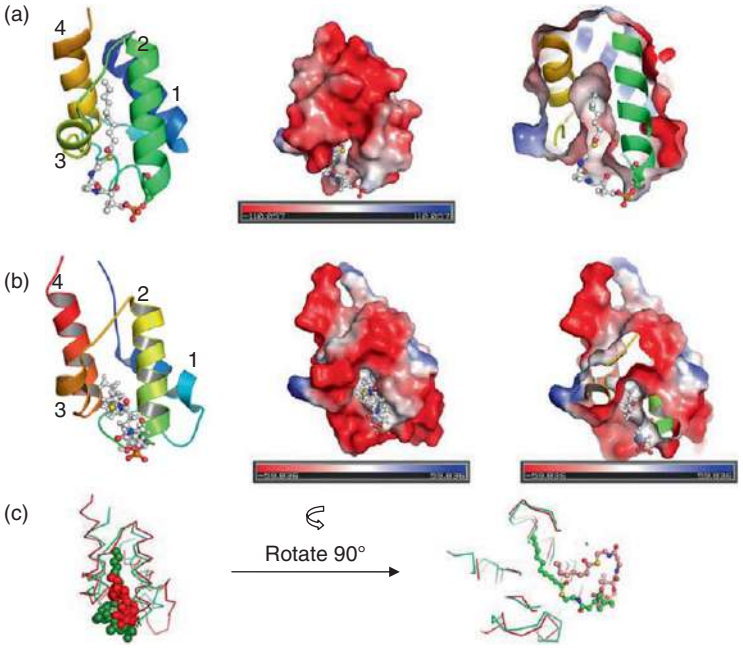


Figure 21 (a) The FAS decanoyl-ACP crystal structure. Left, cartoon. Middle, the negatively charged groove between helices 2 and 3. Right, cutaway surface view revealed that decanoyl-PPT binds ACP in a linear conformation. (b) The NMR structure of act ACP bound with octanoyl-PPT. Left, cartoon. Middle, the negatively charged groove between helices 2 and 3. Right, cut-away surface view revealed that octanoyl-PPT binds act ACP in a bent conformation. (c) Structural overlap between FAS decanoyl-ACP and the actinorhodin octanoyl-ACP showed that the main difference is the linear versus bent PPT. Red, the FAS decanoyl-ACP; green, the actinorhodin octanoyl-ACP.

hydrophilic R72 and N79 that are highly conserved in PKS ACPs but not in the FAS ACPs, and these hydrophilic residues may help stabilize the growing polyketide chain.

In 2003, the first holo-ACP solution structure was solved for the fren ACP,[126] where the PPT prosthetic group is observed. Similar to the apo act ACP structure, the averaged fren ACP structure showed three long α-helices packed in the core bundle, while three short helices are observed in intervening loops with one slowly exchanging between two conformations. A comparison of the fren ACP NMR structure with the crystal structure of butyryl-ACP from *E. coli* suggests that the aforementioned conformational change may be triggered upon polyketide binding to the binding groove proposed in the act ACP work. The solution structure of oxytetracycline (otc) ACP further revealed that the long loop between helices I and II is flexible.[127] Furthermore, helix II is proposed to be the 'hot spot' for protein–protein interactions between ACP and downstream PKS enzymes (**Figure 21(b)**). The hydrophilic groove is distinctly different between PKS and FAS ACPs: the former generally has both positively and negatively charged residues, as well as hydrophobic residues, thus the helical groove is capable of interacting with an incoming polyketide chain with both hydrophobic interactions and hydrogen bonds. Finally, the NMR experiments also support that while the three major helices in type II ACP form a rigid scaffold for the protein, the flexible linker regions may trigger conformational changes that help harbor the highly active polyketide intermediates in the hydrophilic groove region formed between the helices. This hypothesis is validated in two recent publications: (1) The crystal structure of *E. coli* FAS decanoyl-ACP structure shows a deep, interior pocket in ACP between helices 2 and 3, and the pocket extends from the surface Ser (in the signature 'Asp-Ser-Leu' motif, the Ser is covalently attached to the PPT moiety) to the center of ACP (**Figure 21(a)**).[128] Decanoyl-PPT binds the interior pocket in a linear conformation (**Figure 21(a)**), and the decanoyl group extends almost to the top of ACP (**Figure 21(a)**, right panel). (2) In comparison, the recently reported NMR structure of octanoyl-act ACP showed that PPT has a bent conformation (**Figure 21(b)**),[129] and the octanoyl group binds ACP on a surface groove (**Figure 21(b)**). Although both FAS and PKS ACPs have negatively charged surface near the pocket entrance (**Figures 21(a)** and **21(b)**, middle panel), the difference is especially striking when the two structures are overlapped (**Figure 21(c)**). The difference may be caused by bulkier residues in the act ACP pocket, such as L45 (a Thr in the FAS ACP) or L74 (an Ala in the FAS ACP) that prevent octanoyl-PPT from extending into the interior of act ACP.

1.08.5 Protein–Protein Interactions and Transport of Polyketide Intermediates between Enzymes

The studies on isolated type II PKS proteins outlined above clearly indicate that there is an intricate balance between different type II PKS enzymes, such that chain elongation, reduction, and cyclization happen in a precisely timed fashion. Such accurate timing requires extensive protein–protein interactions, as well as precise control in transporting the polyketide intermediates between KS/CLF, KR, and ARO/CYC. Presumably, such extensive protein–protein interactions should suggest that a tight protein complex forms, especially between ACP and other enzyme proteins. Furthermore, KS/CLF, KR, ARO/CYC, and ACP exist as heterotetramer, homotetramer, monomer, and monomer, respectively; any complex formation should be detectable via protein-sizing experiments. However, past studies by native gel and size-exclusion chromatography showed that no stable ACP–enzyme complex can be detected between apo or holo ACP and KS/CLF, KR, or ARO/CYC. Similarly, no stable enzyme complex was observed between KS/CLF and KR or ARO/CYC. The only exception is when the KS active site Cys is labeled with a 16-carbon acyl chain, resulting in the detection of a tight KS/CLF–ACP complex in the protein-sizing experiments. The above results imply that the protein surface–surface interactions may not be sufficient to warrant a stable complex formation between type II PKS proteins. Rather, the bound polyketide substrate is at least as important a factor in fostering complex formation between the type II PKSs. Of course, the type II PKS may form a highly transient complex that evades the detection of protein-sizing experiments. Therefore, in type II PKS, protein–protein interactions likely result from a

combination of protein surface interacting 'hot spots', as well as the polyketide intermediate that is attached to ACP via the PPT anchor. This also helps explain the perfect timing observed in chain transport – only when the polyketide is 'matured' would the next protein–protein interactions be promoted, resulting in the formation of a dynamic protein complex that may dissolve after subsequent reactions.

With the above argument, two key questions remain: (1) Is there a 'hot spot' on ACP, KS/CLF, KR, and ARO/CYC? and (2) Can one trap the protein complex with polyketide bound? The former question can be visualized via protein–protein docking simulation between ACP and other proteins. Indeed, in every case, a Arg/Lys-rich, positively charged groove is identified near the substrate pocket entrance of KS/CLF, KR, and ARO/CYC, and the 'arginine patch' surface generally complements well with the negatively charged surface of the ACP helical groove, where the polyketide is proposed to be stored in the 'Switch Blade' theory, as observed in the fungal FAS complex structure. This also helps explain how the highly reactive polyketide intermediate may be transported from one enzyme to the next: in essence, ACP serves as the 'time capsule' that alternatively shuttles the polyketide intermediate between one enzyme and its downstream partner, using its helical groove as the sheath to protect the polyketide intermediate. Only when the polyketide chain is 'ripe' can a stable, albeit semitransient, complex be formed between the upstream protein, ACP, and the downstream enzyme (e.g., KS/CLF–ACP–KR or KS/CLF–ACP–ARO/CYC), and the polyketide chain is transferred to the next enzyme as soon as the semitransient complex is formed. While this hypothesis may be attractive, it still requires experimental proof, preferentially in the form of a polyketide-bound complex of KS/CLF–ACP–KR or KS/CLF–ACP–ARO/CYC.

1.08.6 Summary and Future Prospects

Since Collie and Raper first discussed the origin of fatty acids and polyketides a century ago, many details have been learned about both biosynthetic pathways. Gene clusters were cloned and sequenced, their biosynthetic mechanism elucidated, and during the last decade, protein structures of most type II FAS and PKS proteins have been solved. This overall accomplishment, a result of the hard work of many labs, offers a clearer understanding of how each PKS enzyme catalyzes their corresponding reaction and achieves precise regiospecificity. Despite high sequence homology and similarities between type II FASs and PKSs at the level of the individual reaction mechanisms and the structures of individual enzymes, key differences still exist. In type I FASs, each domain is used iteratively, and these domains rigidly carry out each round of chain elongation and modification to ensure the faithful production of a saturated long-chain fatty acid. On the other hand, PKSs provide a remarkable example of how nature diversifies its produces via a controlled variation of chain length, reduction pattern, cyclization diversity, and subsequent chain modifications. The structure–function studies of type II PKS enzymes have provided a detailed physical picture of how the chain length, reduction, and cyclization specificities are controlled. However, key questions remain to be elucidated, such as the timing and precise location of the first ring and second ring cyclizations, how ACP transfer the active polyketide intermediates between KS/CLF, KR, and ARO/CYC, as well as the detailed mechanisms of KS-catalyzed malonate decarboxylation. Finally, to probe for protein–protein interactions, the next goal is to obtain high-resolution multiprotein complex structures of type II PKS enzymes. The solutions to the above mysteries will shed light on the intricate balancing acts of type II PKSs that produce the huge varieties of aromatic natural products in nature.

Acknowledgments

This article is dedicated to Ms. Agnes E. Farris (1924–2008). Sheryl Tsai is supported by a grant GM076330 from the NIH.

Glossary

acyl carrier protein A small (<10 kDa) protein that transports the growing polyketide chain from one domain to the next.
aromatase/cyclase The enzyme domain responsible for cyclization/aromatization of a type II polyketide. Three types of aromatase/cyclase are currently identified in nature: C9–C14 monodomain in a nonreducing polyketide synthase, C7–C12 monodomain in a nonreducing polyketide synthase, and C7–C12 didomain in a reducing polyketide synthase.
fourth ring cyclase The cyclase that acts to close the final ring of a type II polyketide, often the fourth ring.
ketoreductase The type II PKS reductase domain that often specifically reduces only the C9 carbonyl group of a C7–C12 cyclized polyketide.
ketosynthase/chain length factor The heterodimeric complex that catalyze the chain elongation of type II polyketides.
polyketide synthase The multienzyme complex that biosynthesizes polyketide natural products.
polyketide A family of natural products that include many pharmaceutically important compounds.
self-malonylation The ability of type II PKS ACP to prime itself with the malonyl group.
type II polyketide synthase A subclass of polyketide synthase whose catalytic domains are stand-alone enzymes. Depending on the presence of ketoreductase domain, the type II polyketide synthase can be further divided into reducing and nonreducing type II polyketide synthase.
type II polyketide A subclass of polyketides that are produced by type II polyketide synthase, often containing a series of conjugated aromatic rings.

Abbreviations

ACP	acyl carrier protein
act	actinorhodin
ARO/CYC	aromatase/cyclase
AT	acyltransferase
CoA	coenzyme A
DH	dehydratase
ER	enoyl reductase
FAS	fatty acid synthase
fren	frenolicin
gra	granaticin
gris	griseusin
jad	jadomycin
KR	ketoreductase
KS	ketosynthase
KS/CLF	ketosynthase/chain length factor (heterodimeric protein, also known as KS_α–KS_β)
MAT	malonyl-CoA:ACP transacylase
ME	methyltransferase
NADP(H)	nicotinamide adenine dinucleotide phosphate
NMR	nuclear magnetic resonance spectroscopy
otc	oxytetracycline
PKS	polyketide synthase
PPT	phosphopantetheine
SDR	short-chain dehydrogenase/reductase
T_3HNR	trihydroxynaphthalene reductase
T_4HNR	tetrahydroxynaphthalene reductase

Tcm	tetracenomycin
TE	thioesterase
TLM	thiolactomycin
WT	wild type

Nomenclature

Å	ångström
k_{cat}	catalytic rate constant
K_d	dissociation constant
kDa	kilodalton
K_m	Michaelis constant
mol l^{-1}	molar concentration

References

1. D. A. Hopwood, *Chem. Rev.* **1997**, *97* (7), 2465–2498.
2. H. Seto; T. Sato; H. Yonehara, *J. Antibiot. (Tokyo)* **1973**, *26* (10), 609–614.
3. M. G. Watve; R. Tickoo; M. M. Jog; B. D. Bhole, *Arch. Microbiol.* **2001**, *176* (5), 386–390.
4. T. H. Haight; M. Finland, *Proc. Soc. Exp. Biol. Med.* **1952**, *81* (1), 175–183.
5. F. Malpartida; D. A. Hopwood, *Nature* **1984**, *309* (5967), 462–464.
6. H. Motamedi; C. R. Hutchinson, *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84* (13), 4445–4449.
7. S. L. Otten; K. J. Stutzman-Engwall; C. R. Hutchinson, *J. Bacteriol.* **1990**, *172* (6), 3427–3434.
8. K. Jakobi; C. Hertweck, *J. Am. Chem. Soc.* **2004**, *126* (8), 2298–2299.
9. F. Lombó; G. Blanco; E. Fernández; C. Méndez; J. Salas, *Gene* **1996**, *172* (1), 87–91.
10. M. Manzoni; M. Rollini, *Appl. Microbiol. Biotechnol.* **2002**, *58* (5), 555–564.
11. A. Schoppner; H. Kindl, *J. Biol. Chem.* **1984**, *259* (11), 6806–6811.
12. S. E. Baker; S. Kroken; P. Inderbitzin; T. Asvarak; B.-Y. Li; L. Shi; O. C. Yoder; B. G. Turgeon, *Mol. Plant Microbe Interact.* **2006**, *19* (2), 139–149.
13. J. Schumann; C. Hertweck, *J. Biotechnol.* **2006**, *124* (4), 690–703.
14. C. W. Carreras; G. W. Ashley, *EXS* **2000**, *89*, 89–108.
15. B. Shen, *Curr. Opin. Chem. Biol.* **2003**, *7* (2), 285–295.
16. P. Caffrey; D. J. Bevitt; J. Staunton; P. F. Leadlay, *FEBS Lett.* **1992**, *304* (2–3), 225–228.
17. S. Smith; S. C. Tsai, *Nat. Prod. Rep.* **2007**, *24* (5), 1041–1072.
18. M. B. Austin; J. P. Noel, *Nat. Prod. Rep.* **2003**, *20* (1), 79–110.
19. T. Maier; M. Leibundgut; N. Ban, *Science* **2008**, *321* (5894), 1315–1322.
20. J. C. Clapham; J. R. Arch, *Diabetes Obes. Metab.* **2007**, *9* (3), 259–275.
21. S. J. Kridel; W. T. Lowther; C. W. Pemble, *Expert Opin. Investig. Drugs* **2007**, *16* (11), 1817–1829.
22. H. T. Wright; K. A. Reynolds, *Curr. Opin. Microbiol.* **2007**, *10* (5), 447–453.
23. H. J. Purohit; S. Cheema; S. Lal; C. P. Raut; V. C. Kalia, *Infect Disord. Drug Targets* **2007**, *7* (3), 245–250.
24. T. Maier; S. Jenni; N. Ban, *Science* **2006**, *311* (5765), 1258–1262.
25. B. Chakravarty; Z. Gu; S. S. Chirala; S. J. Wakil; F. A. Quiocho, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (44), 15567–15572.
26. M. Leesong; B. S. Henderson; J. R. Gillig; J. M. Schwab; J. L. Smith, *Structure* **1996**, *4* (3), 253–264.
27. J. G. Olsen; A. Kadziola; P. von Wettstein-Knowles; M. Siggaard-Andersen; Y. Lindquist; S. Larsen, *FEBS Lett.* **1999**, *460* (1), 46–52.
28. A. C. Price; Y. M. Zhang; C. O. Rock; S. W. White, *Biochemistry* **2001**, *40* (43), 12772–12781.
29. L. Serre; E. C. Verbree; Z. Dauter; A. R. Stuitje; Z. S. Derewenda, *J. Biol. Chem.* **1995**, *270* (22), 12961–12964.
30. F. J. Asturias; J. Z. Chadick; I. K. Cheung; H. Stark; A. Witkowski; A. K. Joshi; S. Smith, *Nat. Struct. Mol. Biol.* **2005**, *12* (3), 225–232.
31. A. K. Joshi; V. S. Rangan; S. Smith, *J. Biol. Chem.* **1998**, *273* (9), 4937–4943.
32. A. K. Joshi; A. Witkowski; S. Smith, *Biochemistry* **1997**, *36* (8), 2316–2322.
33. A. Witkowski; A. Ghosal; A. K. Joshi; H. E. Witkowska; F. J. Asturias; S. Smith, *Chem. Biol.* **2004**, *11* (12), 1667–1676.
34. S. W. White; J. Zheng; Y. M. Zhang; C. O. Rock, *Annu. Rev. Biochem.* **2005**, *74*, 791–831.
35. M. P. Crump; J. Crosby; C. E. Dempsey; J. A. Parkinson; M. Murray; D. A. Hopwood; T. J. Simpson, *Biochemistry* **1997**, *36* (20), 6000–6008.

36. C. A. Frederick; L. D. Williams; G. Ughetto; G. A. van der Marel; J. H. van Boom, *Biochemistry* **1990**, *29* (10), 2538–2549.
37. A. Rabbani; R. M. Finn; J. Ausio, *Bioessays* **2005**, *27* (1), 50–56.
38. I. Chopra; M. Roberts, *Microbiol. Mol. Biol. Rev.* **2001**, *65* (2), 232–260.
39. B. A. Pfeifer; S. J. Admiraal; H. Gramajo; D. E. Cane; C. Khosla, *Science* **2001**, *291* (5509), 1790–1792.
40. C. R. Hutchinson, *Chem. Rev.* **1997**, *97* (7), 2525–2536.
41. H. G. Floss, *J. Ind. Microbiol. Biotechnol.* **2001**, *27* (3), 183–194.
42. R. J. Zawada; C. Khosla, *J. Biol. Chem.* **1997**, *272* (26), 16184–16188.
43. R. McDaniel; P. Licari; C. Khosla, *Adv. Biochem. Eng. Biotechnol.* **2001**, *73*, 31–52.
44. J. A. Robinson, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **1991**, *332* (1263), 107–114.
45. B. S. Moore; J. Piel, *Antonie Van Leeuwenhoek* **2000**, *78* (3–4), 391–398.
46. C. Mendez; J. A. Salas, *Comb. Chem. High Throughput Screen.* **2003**, *6* (6), 513–526.
47. B. Shen; C. R. Hutchinson, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93* (13), 6600–6604.
48. Y. Tang; T. S. Lee; C. Khosla, *PLoS Biol.* **2004**, *2* (2), E31.
49. R. McDaniel; S. Ebert-Khosla; D. A. Hopwood; C. Khosla, *Nature* **1995**, *375* (6532), 549–554.
50. R. McDaniel; S. Ebert-Khosla; H. Fu; D. A. Hopwood; C. Khosla, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91* (24), 11542–11546.
51. H. Fu; R. McDaniel; D. A. Hopwood; C. Khosla, *Biochemistry* **1994**, *33* (31), 9321–9326.
52. R. McDaniel; S. Ebert-Khosla; D. A. Hopwood; C. Khosla, *Science* **1993**, *262* (5139), 1546–1550.
53. E. S. Meadows; C. Khosla, *Biochemistry* **2001**, *40* (49), 14855–14861.
54. Y. Shen; P. Yoon; T. W. Yu; H. G. Floss; D. Hopwood; B. S. Moore, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (7), 3622–3627.
55. Y. Tang; S. C. Tsai; C. Khosla, *J. Am. Chem. Soc.* **2003**, *125* (42), 12708–12709.
56. C. Hertweck; L. Xiang; J. A. Kalaitzis; Q. Cheng; M. Palzer; B. S. Moore, *Chem. Biol.* **2004**, *11* (4), 461–468.
57. D. O'Hagan, *Nat. Prod. Rep.* **1993**, *10* (6), 593–624.
58. G. Meurer; M. Gerlitz; E. Wendt-Pienkowski; L. C. Vining; J. Rohr; C. R. Hutchinson, *Chem. Biol.* **1997**, *4* (6), 433–443.
59. J. Kantola; G. Blanco; A. Hautala; T. Kunnari; J. Hakala; C. Mendez; K. Ylihonko; P. Mantsala; J. Salas, *Chem. Biol.* **1997**, *4* (10), 751–755.
60. R. S. Gokhale; S. Y. Tsuji; D. E. Cane; C. Khosla, *Science* **1999**, *284* (5413), 482–485.
61. K. K. Burson; W. H. Huestis; C. Khosla, *Abstr. Pap. Am. Chem. Soc.* **1997**, *213*, 83 -Biot.
62. R. J. Zawada; C. Khosla, *Chem. Biol.* **1999**, *6* (9), 607–615.
63. B. Shen; C. R. Hutchinson, *Proc. Natl. Acad. Sci.* **1996**, *93*, 6600–6604.
64. C. M. H. Watanabe; C. A. Townsend, *Chem. Biol.* **2002**, *9* (9), 981–988.
65. J. Staunton; K. J. Weissman, *Nat. Prod. Rep.* **2001**, *18* (4), 380–416.
66. A. L. Matharu; R. J. Cox; J. Crosby; K. J. Byrom; T. J. Simpson, *Chem. Biol.* **1998**, *5* (12), 699–711.
67. R. G. Summers; A. Ali; B. Shen; W. A. Wessel; C. R. Hutchinson, *Biochemistry* **1995**, *34* (29), 9389–9402.
68. C. W. Carreras; C. Khosla, *Biochemistry* **1998**, *37* (8), 2084–2088.
69. W. Bao; E. Wendt-Pienkowski; C. R. Hutchinson, *Biochemistry* **1998**, *37* (22), 8132–8138.
70. C. Bisang; P. F. Long; J. Cortes; J. Westcott; J. Crosby; A. L. Matharu; R. J. Cox; T. J. Simpson; J. Staunton; P. F. Leadlay, *Nature* **1999**, *401* (6752), 502–505.
71. J. Dreier; A. N. Shah; C. Khosla, *J. Biol. Chem.* **1999**, *274* (35), 25108–25112.
72. J. Dreier; Q. Li; C. Khosla, *Biochemistry* **2001**, *40* (41), 12407–12411.
73. A. E. Szafranska; T. S. Hitchman; R. J. Cox; J. Crosby; T. J. Simpson, *Biochemistry* **2002**, *41* (5), 1421–1427.
74. H. Oku; N. Futamori; K. Masuda; Y. Shimabukuro; T. Omine; H. Iwasaki, *Biosci. Biotechnol. Biochem.* **2003**, *67* (10), 2106–2114.
75. C. J. Arthur; A. Szafranska; S. E. Evans; S. C. Findlow; S. G. Burston; P. Owen; I. Clark-Lewis; T. J. Simpson; J. Crosby; M. P. Crump, *Biochemistry* **2005**, *44* (46), 15414–15421.
76. C. J. Arthur; A. E. Szafranska; J. Long; J. Mills; R. J. Cox; S. C. Findlow; T. J. Simpson; M. P. Crump; J. Crosby, *Chem. Biol.* **2006**, *13* (6), 587–596.
77. P. Beltran-Alvarez; R. J. Cox; J. Crosby; T. J. Simpson, *Biochemistry* **2007**, *46* (50), 14672–14681.
78. A. T. Keatinge-Clay; A. A. Shelat; D. F. Savage; S. C. Tsai; L. J. Miercke; J. D. O'Connell, III; C. Khosla; R. M. Stroud, *Structure* **2003**, *11* (2), 147–154.
79. Y. Hori; Y. Abe; M. Nishimura; T. Goto; M. Okuhara; M. Kohsaka, *J. Antibiot. (Tokyo)* **1993**, *46* (7), 1069–1075.
80. Y. Hori; S. Takase; N. Shigematsu; T. Goto; M. Okuhara; M. Kohsaka, *J. Antibiot. (Tokyo)* **1993**, *46* (7), 1063–1068.
81. Y. Hori; Y. Abe; M. Ezaki; T. Goto; M. Okuhara; M. Kohsaka, *J. Antibiot. (Tokyo)* **1993**, *46* (7), 1055–1062.
82. T. Marti; Z. Hu; N. L. Pohl; A. N. Shah; C. Khosla, *J. Biol. Chem.* **2000**, *275* (43), 33443–33448.
83. H. Pan; S.-c. Tsai; E. S. Meadows; L. J. W. Miercke; A. T. Keatinge-Clay; J. O'Connell; C. Khosla; R. M. Stroud, *Structure* **2002**, *10* (11), 1559–1568.
84. Y. Tang; T. S. Lee; S. Kobayashi; C. Khosla, *Biochemistry* **2003**, *42* (21), 6588–6595.
85. T. S. Lee; C. Khosla; Y. Tang, *J. Am. Chem. Soc.* **2005**, *127* (35), 12254–12262.
86. A. T. Keatinge-Clay; D. A. Maltby; K. F. Medzihradszky; C. Khosla; R. M. Stroud, *Nat. Struct. Mol. Biol.* **2004**, *11*, 888–893.
87. J. A. Kalaitzis; B. S. Moore, *J. Nat. Prod.* **2004**, *67* (8), 1419–1422.
88. U. Rix; C. Fischer; L. L. Remsing; J. Rohr, *Nat. Prod. Rep.* **2002**, *19* (5), 542–580.
89. A. T. Hadfield; C. Limpkin; W. Teartasin; T. J. Simpson; J. Crosby; M. P. Crump, *Structure (Camb.)* **2004**, *12* (10), 1865–1875.
90. T. P. Korman; J. A. Hill; T. N. Vu; S. C. Tsai, *Biochemistry* **2004**, *43* (46), 14529–14538.
91. A. C. Price; Y. M. Zhang; C. O. Rock; S. W. White, *Structure (Camb.)* **2004**, *12* (3), 417–428.
92. K. Nakajima; A. Yamashita; H. Akama; T. Nakatsu; H. Kato; T. Hashimoto; J. Oda; Y. Yamada, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95* (9), 4876–4881.
93. R. J. Zawada; C. Khosla, *Chem. Biol.* **1999**, *6* (9), 607–615.
94. H. Fu; S. Ebertkhosla; D. A. Hopwood; C. Khosla, *J. Am. Chem. Soc.* **1994**, *116* (10), 4166–4170.
95. H. Dutler; A. Kull; R. Mislin, *Eur. J. Biochem.* **1971**, *22* (2), 213–217.
96. A. K. Joshi; S. Smith, *J. Biol. Chem.* **1993**, *268* (30), 22508–22513.

97. L. H. Ostergaard; L. Kellenberger; J. Cortes; M. P. Roddis; M. Deacon; J. Staunton; P. F. Leadlay, *Biochemistry* **2002**, *41* (8), 2719–2726.
98. T. P. Korman; Y. H. Tan; J. Wong; R. Luo; S. C. Tsai, *Biochemistry* **2008**, *47* (7), 1837–1847.
99. Y. M. Zhang; C. O. Rock, *J. Biol. Chem.* **2004**, *279* (30), 30994–31001.
100. T. J. Simpson; M. K. B. Weerasooriya, *J. Chem. Soc. Perkin Trans. 1* **2000**, (16), 2771–2775.
101. A. Andersson; D. Jordan; G. Schneider; Y. Lindqvist, *Structure* **1996**, *4* (10), 1161–1170.
102. A. Andersson; D. Jordan; G. Schneider; Y. Lindqvist, *FEBS Lett.* **1997**, *400* (2), 173–176.
103. I. E. Holzbaur; A. Ranganathan; I. P. Thomas; D. J. Kearney; J. A. Reather; B. A. Rudd; J. Staunton; P. F. Leadlay, *Chem. Biol.* **2001**, *8* (4), 329–340.
104. P. Caffrey, *ChemBioChem* **2003**, *4* (7), 654–657.
105. Y. Tang; H. Y. Lee; Y. Tang; C. Y. Kim; I. Mathews; C. Khosla, *Biochemistry* **2006**, *45* (47), 14085–14093.
106. M. A. Alvarez; H. Fu; C. Khosla; D. A. Hopwood; J. E. Bailey, *Nat. Biotechnol.* **1996**, *14* (3), 335–338.
107. R. McDaniel; S. Ebert-Khosla; D. A. Hopwood; C. Khosla, *Nature* **1995**, *373*, 549–554.
108. R. J. X. Zawada; C. Khosla, *Chem. Biol.* **1999**, *6*, 607–615.
109. B. Shen; C. R. Hutchinson, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6600–6604.
110. T. B. Thompson; K. Katayama; K. Watanabe; C. R. Hutchinson; I. Rayment, *J. Biol. Chem.* **2004**, *279* (36), 37956–37963.
111. A. C. Price; C. O. Rock; S. W. White, *J. Bacteriol.* **2003**, *185* (14), 4136–4143.
112. M. B. Austin; M. E. Bowman; J. L. Ferrer; J. Schroder; J. P. Noel, *Chem. Biol.* **2004**, *11* (9), 1179–1194.
113. B. D. Ames; T. P. Korman; W. Zhang; P. Smith; T. Vu; Y. Tang; S.-C. Tsai, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (14), 5349–5354.
114. L. M. Iyer; E. V. Koonin; L. Aravind, *Proteins* **2001**, *43*, 134–144.
115. M. Gajhede; P. Osmark; F. M. Poulsen; H. Ipsen; J. N. Larsen; R. J. Joost van Neerven; C. Schou; H. Lowenstein; M. D. Spangfort, *Nat. Struct. Biol.* **1996**, *3* (12), 1040–1045.
116. Y. Tsujishita; J. H. Hurley, *Nat. Struct. Biol.* **2000**, *7* (5), 408–414.
117. Z. Markovic-Housley; M. Degano; D. Lamba; E. Roepenack-Lahaye; S. Clemens; M. Susani; F. Ferreira; O. Scheiner; H. Breiteneder, *J. Mol. Biol.* **2003**, *325*, 123–133.
118. J. E. Mogensen; R. Wimmer; J. N. Larsen; M. D. Spangfort; D. E. Otzen, *J. Biol. Chem.* **2002**, *277* (26), 23684–23692.
119. O. Pasternak; J. Biesiadka; R. Dolot; L. Handschuh; G. Bujacz; M. M. Sikorski; M. Jaskolski, *Acta Crystallogr. D* **2005**, *61*, 99–107.
120. M. S. Kimber; F. Martin; Y. Lu; S. Houston; M. Vedadi; A. Dharamsi; K. Fiebig; M. Schmid; C. O. Rock, *J. Biol. Chem.* **2004**, *279* (50), 52593–52602.
121. R. J. X. Zawada; C. Khosla, *J. Biol. Chem.* **1997**, *272* (26), 16184–16188.
122. S. C. Dillon; A. Bateman, *BMC Bioinf.* **2004**, *5*, 109–123.
123. T. P. Korman; J. A. Hill; T. N. Vu; S. C. Tsai, *Biochemistry* **2004**, *43* (46), 14529–14538.
124. A. Sultana; P. Kallio; A. Jansson; J. Wang; J. Niemi; P. Mantsala; G. Schneider, *EMBO J.* **2004**, *23*, 1911–1921.
125. P. Kallio; A. Sultana; J. Niemi; P. Mantsala; G. Schneider, *J. Mol. Biol.* **2006**, *357*, 210–220.
126. Q. Li; C. Khosla; J. D. Puglisi; C. W. Liu, *Biochemistry* **2003**, *42* (16), 4648–4657.
127. S. C. Findlow; C. Winsor; T. J. Simpson; J. Crosby; M. P. Crump, *Biochemistry* **2003**, *42* (28), 8423–8433.
128. A. Roujeinikova; W. J. Simon; J. Gilroy; D. W. Rice; J. B. Rafferty; A. R. Slabas, *J. Mol. Biol.* **2007**, *365* (1), 135–145.
129. S. E. Evans; C. Williams; C. J. Arthur; E. Ploskon; P. Wattana-Amorn; R. J. Cox; C. L. Willis; J. Crosby; T. J. Simpson; M. P. Crump, *J. Mol. Biol.* **2009**, *389* (3), 511–528.

Biographical Sketches

Tyler Korman received his BS and MS in Chemistry from the University of California, San Diego where he studied the effects of mutation on the co-operactivity of the F_1 portion of the F_0–F_1 ATP Synthase from the thermophilic *Bacillus* PS3 with Prof. William S. Allison. Tyler then received his PhD in Molecular Biology and Biochemistry from the Tsai Lab at the University of California, Irvine in 2008 where he studied the structure-function relationship

of key polyketide modification enzymes involved in the reduction and cyclization of bacterial and fungal natural products with the ultimate goal of rationally designing new polyketides. In 2008, Tyler joined the lab of Dr. James U. Bowie at UCLA as a postdoctoral fellow where he is developing a novel biosynthetic pathway for biodiesel production in *E. coli*.

Brian Ames attended the Brigham Young University, Hawaii, and graduated *summa cum laude* with a BS in biochemistry (2004). He is currently a Ph.D. candidate in molecular biology and biochemistry at the University of California, Irvine, working in the laboratory of Dr. Sheryl Tsai. His doctoral research focuses on understanding the structure and function of the bacterial aromatic and fungal highly reducing polyketide synthases, with the aim of utilizing a structure-based approach to engineer the biosynthesis of novel polyketides. Following completion of his Ph.D. degree in June 2009, Brian will begin postdoctoral research in the laboratory of Christopher T. Walsh at the Harvard Medical School, where he will study anthranilate-based nonribosomal peptide synthetase assembly lines.

Sheryl Tsai received her BS and MS in chemistry from the National Taiwan University, where she studied the syntheses and organometallic chemistry of tripodal ligands with Prof. Shiuh-Tzung Liu. She then received her Ph.D. in chemistry from University of California, Berkeley, in 1999, where she examined the hydrogen tunneling phenomena in enzymes with Prof. Judith Klinman. As a postdoctoral fellow, Sheryl investigated the structural enzymology of polyketide synthase in the laboratories of Prof. Chaitan Khosla of Stanford University and Prof. Robert Stroud in University of California, San Francisco. In 2003, Sheryl joined the University of California, Irvine, and is currently an associate professor of molecular biology and biochemistry, chemistry, and pharmaceutical sciences. She is a 2006 Pew Scholar and the recipient of the 2007 Golden Apple Teaching Excellence Award. Her research interest includes enzyme structures and mechanisms of multidomain enzyme complexes, such as acyl-CoA carboxylase and polyketide synthase.

1.09 Fungal Type I Polyketides

Russell J. Cox and Thomas J. Simpson, University of Bristol, Bristol, UK

© 2010 Elsevier Ltd. All rights reserved.

1.09.1	**Introduction**	347
1.09.2	**Diversity of Fungal Polyketides**	348
1.09.2.1	The Chemical Reactions of Fungal Polyketide Biosynthesis	348
1.09.2.2	Structural Types – Relationship between Compound, Protein, and Gene Structures	353
1.09.2.3	Biological Activities	353
1.09.3	**Biosynthesis of Fungal Polyketides**	354
1.09.3.1	Highly Reduced Polyketides	354
1.09.3.1.1	The lovastatin polyketide synthases	355
1.09.3.1.2	The squalestatin S1 polyketide synthases	356
1.09.3.1.3	The fumonisin and T-toxin polyketide synthases	357
1.09.3.1.4	Highly reduced polyketide synthase from *Alternaria solani*	357
1.09.3.2	Partially Reduced Polyketides	359
1.09.3.3	Nonreduced Polyketides	362
1.09.3.3.1	Nonreduced polyketide synthase loading component	363
1.09.3.3.2	Nonreduced polyketide synthase chain extension component	365
1.09.3.3.3	Nonreduced polyketide synthase processing component	367
1.09.3.3.4	Claisen cyclase/thiolesterase domains	367
1.09.3.3.5	Chain-shortening reactions	367
1.09.3.3.6	Acyl carrier protein domains	369
1.09.3.3.7	*C*-MeT domains	369
1.09.3.3.8	R domains	370
1.09.3.3.9	Interaction of isolated NR PKS domains with components of bacterial type II PKS	370
1.09.4	**Mixed Polyketide/Nonribosomal Peptides**	372
1.09.5	**Meroterpenoids**	374
1.09.6	**Post-PKS Reactions in Fungi**	376
1.09.7	**Conclusions**	378
References		380

1.09.1 Introduction

It is conservatively estimated that there are some one and a half million species of fungi on planet Earth.[1] They occupy a diverse range of ecological niches and impact human existence in many ways. They play roles in human disease and nutrition, in agriculture where they make up the largest body of plant pathogenic organisms, in animal husbandry, and in the construction industry where their wood-rotting abilities are problematic. Fungi have long been known as sources of biologically active compounds – toxins, mutagens, psychoactives, medicines, and dyes. Although relatively few fungal species have been investigated chemically (certainly less than 1%), the diversity of compounds isolated and characterized from this limited sample is very wide. This diversity in structure (and thus biological properties) reflects the extremely wide range of chemistry – often oxidative in nature – which fungi are able to perform. Fungal secondary metabolites include alkaloids, terpenoids, and nonribosomal peptides, but polyketides and related compounds form the largest class of natural products derived from fungi.

Historically, the fungi, and fungal polyketides in particular, were one of the first classes of organisms to be investigated systematically for their biosynthetic abilities. For example, Birch showed in 1955 that the simple polyketide 6-methylsalicylic acid (6-MSA) (**1**) is biosynthesized from the head-to-tail linkage of four acetates, using radioisotopic feeding experiments.[2] Since then, more than 50 years of effort have revealed increasingly more detail about the biosynthesis of polyketides in general, but the role of the fungi as host to some of the most

fascinating biosynthetic chemistry is still true today. This chapter sets out to explore the diversity of chemistry utilized during the polyketide biosynthesis in fungi. Much is now understood about fungal biosynthesis at the genetic level and as this forms the conceptual bedrock for the subject area, it is convenient to link genes, proteins, and chemistry throughout this work.

1.09.2 Diversity of Fungal Polyketides

Among the simplest fungal polyketides are the benzoic acids including 6-MSA (**1**), orsellinic acid (**2**), and its C-methylated derivatives (**Figure 1**). These are tetraketides (i.e., derived from four units of acetate). Related pentaketides include compounds such as mellein (**3a**) and methylmellein (**3b**), citrinin (**4**), and ochratoxin A (**5**), where the polyketide is bonded to phenylalanine. The pentaketide tetrahydroxy-naphthalene (**6**) and the heptaketide Ywa1 (**7**) show how the simple aromatics can be extended to derivatives of naphthalene and indeed anthraquinones such as the octaketide norsolorinic acid (**8**). Simple compounds such as the orsellinic acids and phloroacetophenone (**9**) are commonly found as metabolites of the mycobiont symbiotic partners of lichen and can be oligeromized to form the usnic acids (**10**) and depsides such as gyrophoric acid (**11**).[3] Compounds such as orsellinic acid (**2**) are also metabolized to form more highly oxidized species such as penicillic acid (**12**), while 6-MSA (**1**) is converted to the mycotoxin patulin (**13**). Similarly, the simple polyketide 3-methylorcinaldehyde (**14**) appears to be a precursor of the structurally complex xenovulene (**15**) series and tropolones such as stipitatonic acid (**16**). Other compounds combine aromatic and more aliphatic components, such as dehydrocurvularin (**17**) and monocerin (**18**), while others such as lovastatin (**19a**), T-toxin (**20**), and fumonisin B1 (**21**) are generally regarded as being more reduced.

Polyketides can be halogenated – commonly chlorinated as in griseofulvin (**22**) and strobilurin B (**23**) (**Figure 2**). Epoxidation is common as in fusarin C (**24**), and hydroxylation at 'nonacetate C-1' positions is also common, as in barnol (**25**). Cyclizations, through nitrogen (e.g., fusarin C (**24**)), oxygen (e.g., terrestric acid, **26**), and carbon (e.g., lovastatin, **19a**, which has undergone a Diels–Alder cyclization) are also observed. Condensation with terpenes (forming the meroterpenoids) is often observed and condensation with amino acids is also common as exemplified by the 2-pyridones such as funiculosin (**27**), tetramic acids such as pretenellin A (**28**), and cytochalasins and chaetoglobosins such as chaetoglobosin A (**29**). Compounds such as cyclopiazonic acid (**30**) illustrate the ability of polyketides to fuse with both amino acids and terpenes in fungi. Yet further complexity comes in compounds such as squalestatin S1 (**31**), where two polyketides are linked and a succinate-derived moiety makes up the core 5-hydroxy-(1,3)dioxane-4,5,6-tricarboxylic acid skeleton.

Thus, it appears that the variety of chemistry used during fungal polyketide biosynthesis is extremely wide. Indeed, the range of chemical reactions available to the fungi rivals – and in many cases outperforms – that of the modern synthetic chemist, both in scope and ambition of application. Faced with a huge and a wide variety of polyketide-derived structures available from fungi, it would be tempting to abandon hope of classifying them logically. However, as we shall see, understanding of biosynthesis, and more latterly of the proteins and genes involved, allows a rational scheme to be developed.

1.09.2.1 The Chemical Reactions of Fungal Polyketide Biosynthesis

The chemical complexity of fungal polyketides arises as a result of the array of chemical reactions deployed during their biosynthesis. These can be understood to occur in two phases: those involved in the construction of the main carbon chain catalyzed by a multifunctional protein known as a polyketide synthase (PKS), and those reactions occurring after the PKS, catalyzed by numerous and diverse post-PKS processing enzymes (**Scheme 1**).[4]

Fungal PKS enzymes are large multifunctional proteins, possessing at least three recognizable catalytic domains, but often many more. They have been studied most often at the genetic level where they are encoded by large single open reading frames (ORF) of 6–8 kb. The core protein domains consist of an acyl transferase (AT), an acyl carrier protein (ACP), and a β-ketoacyl synthase (KS). An AT domain is responsible for loading the starter unit (usually acetate from acetyl-CoA (**32**)) onto the thiol active site of the KS domain. The same, or a different (see Section 1.09.3.3.1), AT also loads the extender unit (malonate, derived from malonyl-CoA (**33**)) onto the ACP – it is held as a thiolester bound to the terminal thiol of a prosthetic phosphopantetheine arm of

the ACP. The KS then catalyzes the decarboxylative condensation of acetate and malonate to form the first polyketide intermediate – acetoacetate as a thiolester bound to the ACP (**34**). These are the minimal reactions required for polyketide biosynthesis and are not only common with most PKS enzymes in plants and bacteria, but also common with fatty acid biosynthesis in all organisms. *C*-Methyl transferase (*C*-MeT) domains can transfer a methyl group at this stage to give a methylated intermediate such as (**35**). Reduction by a *β*-keto

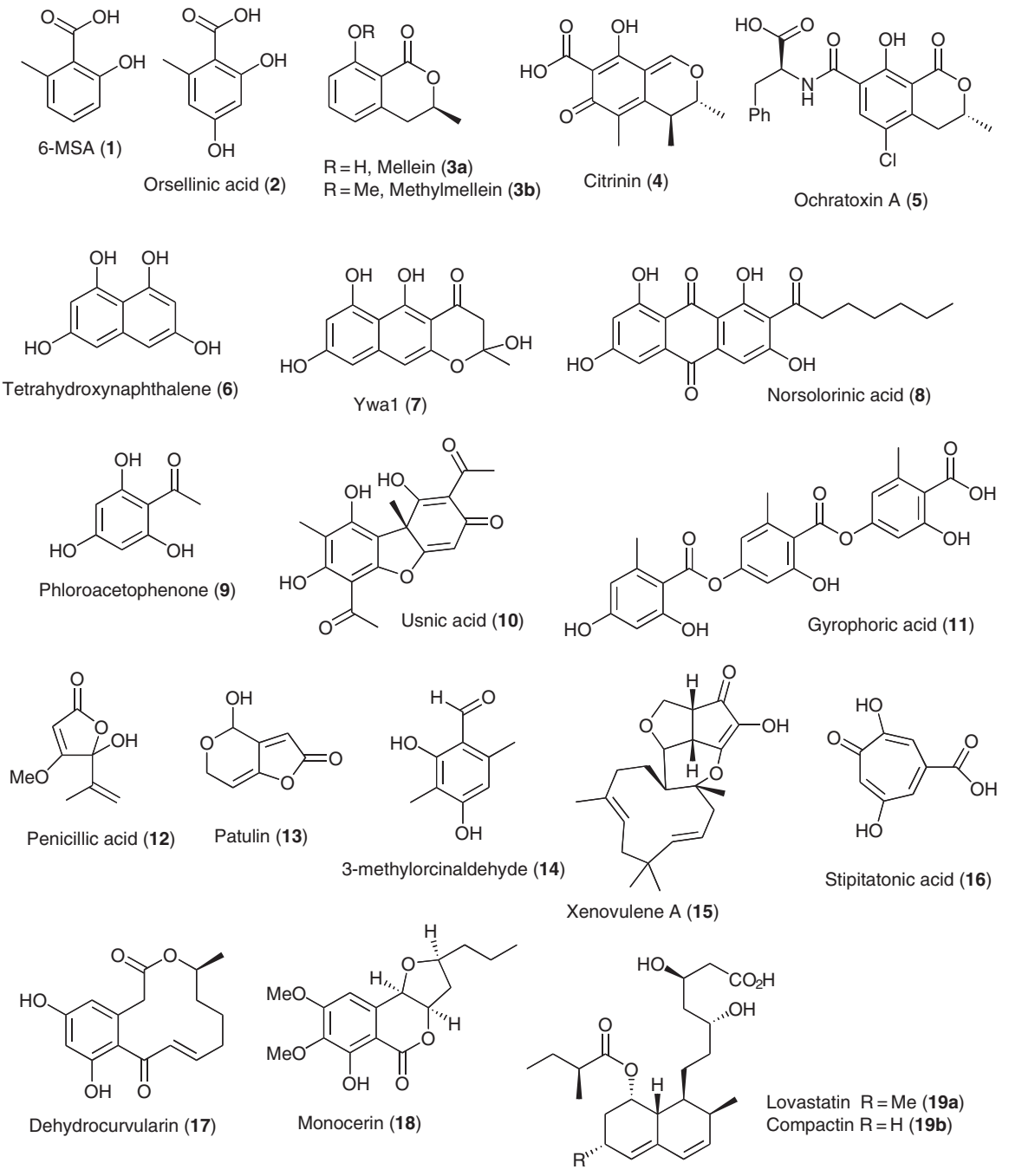
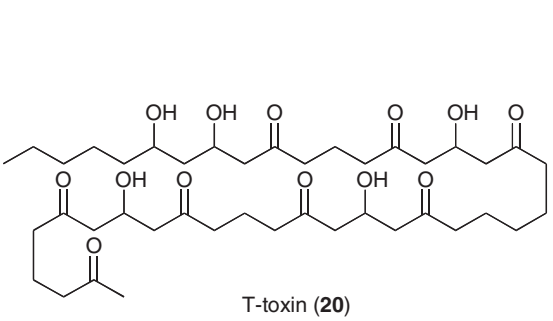
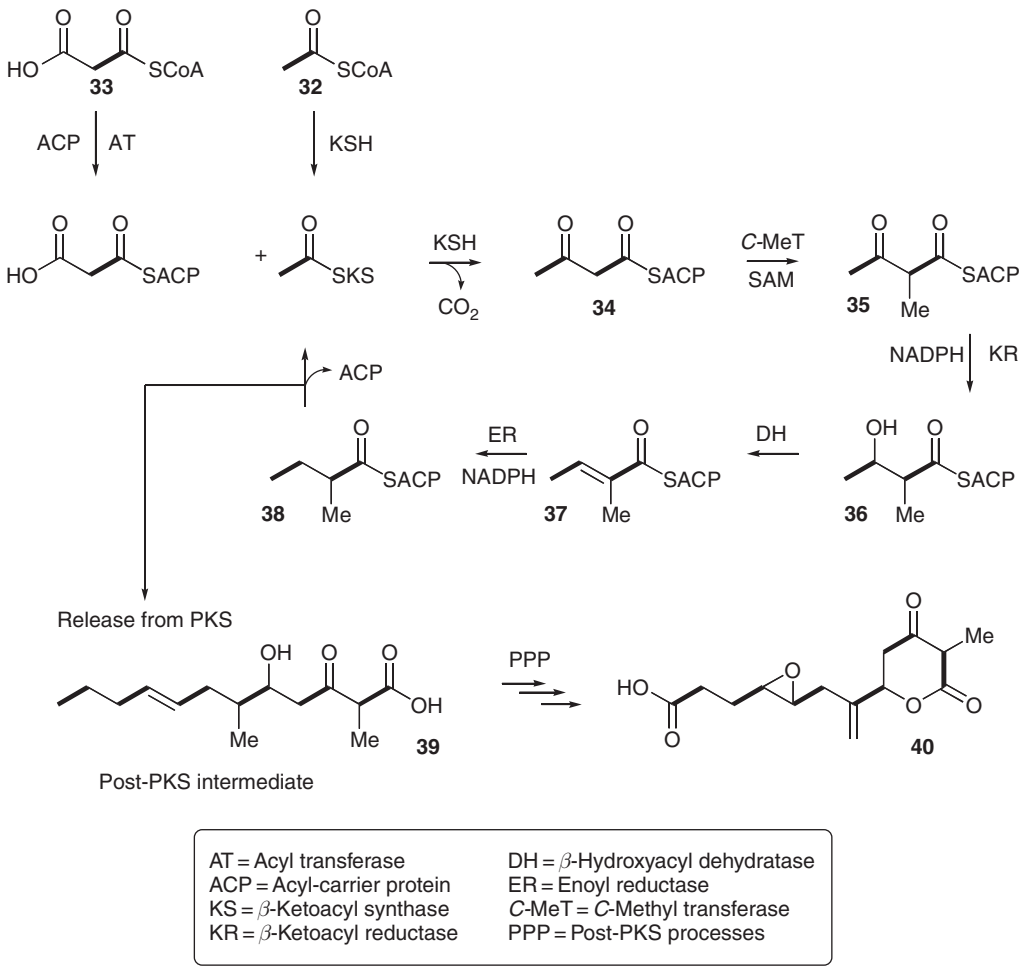


Figure 1 (Continued)


OH
OH
O
O
OH
O
O
OH
O
O
O

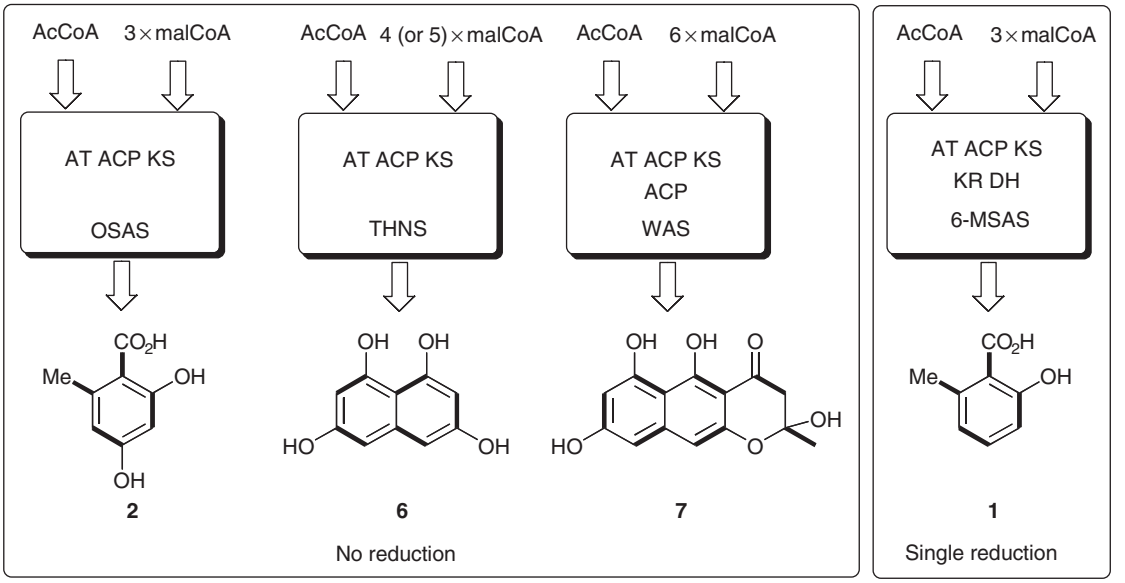


Scheme 1 The chemical reactions of fungal polyketide biosynthesis.

reductase (KR) can then afford the β-hydroxy acyl species (**36**), and the dehydration (DH) of which leads to the enoyl thiolester (**37**). Finally, action of an enoyl reductase (ER) domain forms the fully saturated chain (**38**). The cycle is then repeated to extend the chain.[5] Release of the fully grown and modified chain (e.g., (**39**)) can be achieved by a number of mechanisms, and such compounds are often subject to post-PKS processing, often, in the case of fungi, oxidatively, for example, to give compound **40**.

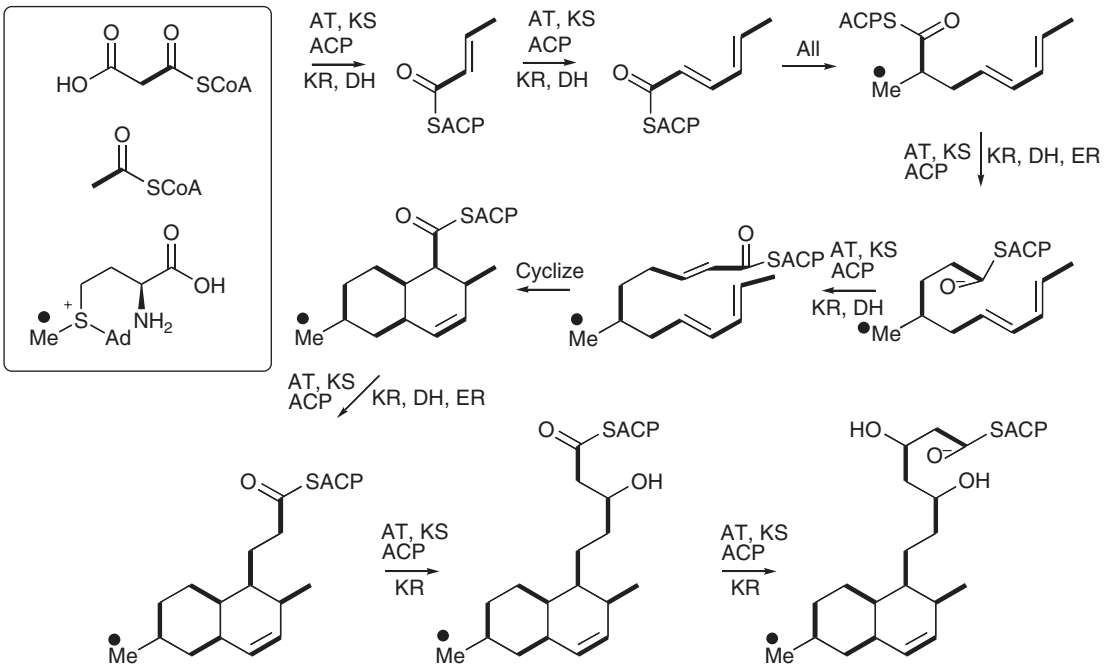
Fungal PKS proteins evidently play out a predetermined program. All the chemical reactions of fungal PKS (apart from methyl transferases) are the same as those catalyzed by fatty acid synthases. However, a PKS has the ability to coordinate its activities – in some way the synthase is 'programed'. The simplest example of this is represented by chain length selection. For example, the reactions carried out by orsellinic acid (orsellinate) synthase (OS) and by tetrahydroxynaphthalene (THN, **6**) synthase are identical except that OS catalyzes three extensions, giving the tetraketide (**2**), while THNS formally catalyzes four extensions (but see below) to give the pentaketide (**6**). Both synthases contain single KS, AT, and ACP domains, which must act iteratively – the difference is in the number of iterations (**Scheme 2**). Similarly, Ywa1 synthase (WAS) catalyzes six extensions to give the unreduced heptaketide (**7**). These, and similar synthases are referred to as nonreducing (NR) synthases that produce 'unreduced polyketides'.

The next level of complexity in programing is exemplified by 6-MSA synthase (6-MSAS). Here, KR and DH domains are present in addition to KS, AT, and ACP. However, the KR acts only once, while the KS domain acts thrice. The KR is active only after the second extension so that further extension and cyclization gives 6-MSA (**1**) (**Scheme 2**). Such compounds are referred to as 'partially reduced' polyketides.



Scheme 2 Chain length programing in fungal NR PKS and PR PKS.

In the case of lovastatin nonaketide biosynthesis (**Scheme 3**), programing is much more complex because the KR, DH, and ER domains must all be present and brought into play in different cycles during the eight rounds of chain extension. Methyl transferase domains are also subject to programed control, for example, during the biosynthesis of the lovastatin nonaketide, the methyl transferase acts only once, during the third round of chain extension. Lovastatin and related compounds belong to a large group of metabolites referred to as 'highly reduced' polyketides. It is the highly reduced (HR) compounds that often show significant chemical complexity and corresponding biological activity – in particular, these compounds are often likely drugs.



Scheme 3 Proposed steps during the biosynthesis of lovastatin nonaketide.

The programing of fungal PKS proteins is significantly more complex than that observed in bacterial PKS systems where each individual reaction is catalyzed by a different catalytic domain. This is exemplified in the case of 6-deoxy erythromycin B (6dEB) biosynthesis in *Saccharopolyspora erythrae*.[6] In this case the domains are arranged into six individual modules, each very similar to a single fungal PKS, to carry out the six rounds of chain extension and required reductive modifications. A single module from these bacterial systems usually acts only once and must pass on its product to the next module. In contrast, a fungal iterative PKS must retain the growing substrate until it is fully formed. Fungal PKSs could therefore be regarded as much more efficient than their modular bacterial analogues. However, the molecular mechanism of programing in fungal PKSs is not yet well understood.

1.09.2.2 Structural Types – Relationship between Compound, Protein, and Gene Structures

The hierarchy of fungal polyketide chemical structures is due to differences in programing of their PKS proteins – apparent increases in product structural complexity are due to increasing use and control of reductive, dehydrative, and methylating steps by the PKS. This must be due to differences in PKS protein sequence and structure. This fact has been exploited in the development of rapid methods for the cloning of fungal PKS genes associated with the biosynthesis of particular fungal polyketide types.

During the 1990s, it was realized that these subtle protein sequence differences should be reflected in DNA sequence and that polymerase chain reaction (PCR) primers could be designed to selectively amplify fragments of fungal PKS genes from fungal genomic DNA (or cDNA).[7] In the early work in this area, Simpson and Lazarus hypothesized that fungal polyketides could be grouped into two classes: nonreduced compounds such as orsellinic acid (**2**), THN (**6**), and norsolorinic acid (**8**), and partially reduced compounds such as 6-MSA (**1**). At that time very few fungal PKS genes were known, and based on very limited sets of sequences they designed degenerate PCR primers that were complementary to conserved DNA sequences in the KS domains in fungal PKS responsible for the biosynthesis of NR and PR compounds.[7] Later, the same analysis was extended to the KS domains of highly reduced compounds such as lovastatin (**19a**) when DNA sequence data became available for the lovastatin nonaketide (LNKS) and diketide synthases (LDKS).[8] The availability of these sequences also allowed the development of selective PCR primers for *C*-MeT domains.

Application of these selective primers in PCR reactions with diverse filamentous fungi yielded DNA products that were fragments of new PKS genes. Detailed sequence analysis showed that the newly generated sequences clustered with other PKS gene sequences from the expected classes. For example, PCR primers designed based on NR PKS sequences amplified fragments of other non-reducing (NR) PKS, and primers based on 6-MSA-type sequences amplified fragments of further examples of partially reducing (PR) PKS genes. Likewise, primers exploiting similarities between LNKS and LDKS gene sequences amplified the fragments of new highly reducing PKS genes.

This sequence analysis has been significantly extended as genomic approaches have been applied to fungi in the recent past.[9] Full genome sequences have now been obtained for more than a one hundred fungi (till January 2009).[10] In each organism many PKS genes have been discovered. For example, *Aspergillus niger* contains 34 PKS genes,[11] so there are now many hundreds of fungal PKS genes known. Sequence comparison of all these new PKS genes, however, shows that the three classes of fungal PKS genes predicted by Simpson and Lazarus are the same three classes observed in the most recent sequence comparisons.[9] Since the NR, PR, and HR nomenclature is useful for describing both the chemical products and their cognate PKS proteins and genes, it is sensible to survey the state of knowledge of fungal PKS in this way.

1.09.2.3 Biological Activities

Fungal PKS-derived compounds show a wide range of biological activities, although relatively few are used directly in the clinic or in agrochemistry. However, many fungal polyketides are known to be potent toxins, mutagens, and compounds active and useful in the new area of chemical genetics. One compound that is used in the clinic is griseofulvin (**22**), which shows unprecedented efficacy in treating fungal infections – particularly of the finger and toenails where it accumulates after oral administration.[12] Other compounds, such as lovastatin (**19a**) and strobilurin B (**23**), show potent biological properties *in vitro* but are unsuited for pharmacological or agrochemical use for other reasons. Lovastatin is a potent inhibitor of HMG-CoA reductase and has been used

as an anticholesterol drug in human medicine. However, more importantly, it has acted as an inspiration for the development of an entire class of semisynthetic and fully synthetic hypercholesterolemia treatments with improved bioavailability, and worth billions of dollars to the international pharmaceutical industry annually. Likewise, strobilurin B (**23**) is a potent inhibitor of quinol oxidation (Q_oI) in the cytochrome bc_1 complex, and has inspired the development of the entire class of Q_oI fungicides used extensively in agrochemistry.[13] Such compounds have stimulated the search for others, for example, the squalestatins also inhibit cholesterol biosynthesis, this time by inhibiting squalene synthase. Much is known of the mycotoxic effects of fungal polyketides. This area has been extensively reviewed elsewhere and the reader is directed to several excellent works in the area.[14,15]

1.09.3 Biosynthesis of Fungal Polyketides

The two-step biosynthesis of polyketides is a general rule: The PKS is responsible for assembling a 'linear' chain of carbons from acetate and malonate, often punctuated with methyl groups from SAM. The PKS also usually coordinates cyclization and release reactions, and once released the polyketide can be acted on by numerous post-PKS tailoring enzymes. It is the interplay between the PKS itself and the post-PKS enzymes, which generates the startling diversity of structures. As described above (Section 1.09.2.2), fungal polyketides and their biosynthesis are best described by considering highly reducing, partially reducing, and NR types of PKS. Post-PKS steps are considered briefly in Section 1.09.6.

1.09.3.1 Highly Reduced Polyketides

The most complex fungal polyketides are generally the HR compounds such as lovastatin (**19a**), T-toxin (**20**), fumonisin B1 (**21**), and squalestatin S1 (**31**). In all cases found to date, these fungal compounds are produced by iterative type I PKS. These PKS have an N-terminal KS domain, followed by AT and DH domains. In many cases the DH is followed by a *C*-MeT domain. Some HR PKS possess an ER domain, but those that lack this domain have a roughly equivalent length of sequence with no known function. This is followed by a KR domain and the PKS often terminates with an ACP. There appears to be no domain similar to the PT domain of the NR PKS (see Section 1.09.3.2.2) or the core domain of the PR PKS, and likewise there is no N-terminal SAT domain as found in the NR PKS (Section 1.09.3.3.1).

Many HR PKS genes are known from the numerous fungal genome sequences, but few gene sequences have been linked to the production of known compounds. However, of the few cases where both gene and chemical product are known (**Table 1**), some progress has been made in understanding function and programing.

Table 1 Examples of proven links between HR PKS and known compounds

Organism	*Gene*	*Protein*	*Final product*	*Reference*
Aspergillus terreus	*lovB*	LNKS	Lovastatin nonaketide (**19a**)	16
	lovF	LDKS	Lovastatin diketide (**19a**)	16
Penicillium citrinum	*mlcA*	CNKS	Compactin nonaketide (**19b**)	17
	mlcB	CDKS	Compactin diketide (**19b**)	17
Giberella zeae	*PKS4*	ZS-A	Zearalenone (**58**)	18
Phoma sp.	*PhPKS1*	SQTKS	Squalestatin tetraketide (**47**)	19
Cochliobolus heterostrophus	*pks1*	TTS1	T-toxin (**20**)	20
	pks2	TTS2	T-toxin (**20**)	21
Giberella fujikuroi	*fum1*	FUMS	Fumonisin B_1 (**21**)	22
Alternaria solani	*alt5*	PKSN	PKSN (**52**)	23
	pksF	PKSF	Aslanipyrone (**53**) Aslaniol (**54**)	24

1.09.3.1.1 The lovastatin polyketide synthases

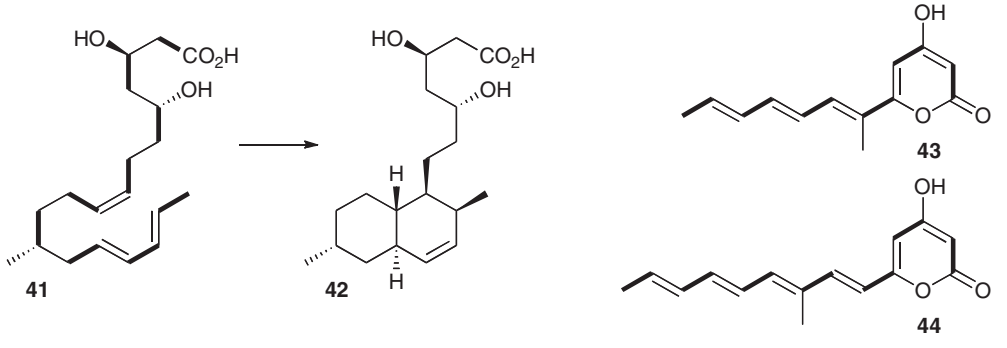
Lovastatin (**19a**) (also known as mevinolin) is produced by *Aspergillus terreus*. The related compound compactin (mevastatin) (**19b**), produced by *Penicillium citrinum*, is identical to lovastatin apart from the C-12 methyl group, which is absent in compactin. The biosynthesis of these compounds has been studied intensely – isotopic feeding experiments[25] have shown that two polyketide chains are required: a nonaketide and a methylated diketide. The requirement for two PKSs is evident in the gene clusters associated with **19a** and **19b** biosynthesis where two PKS genes are found. In the case of lovastatin (**19a**) biosynthesis, the two PKS genes are *lovB* and *lovF*,[26,16] encoding LNKS and LDKS, respectively. Two homologous genes are involved during compactin **19b** biosynthesis: *mlcA* (encoding CNKS) and *mlcB* (encoding CDKS).[17]

Isotopic feeding experiments had suggested that LNKS should synthesize a fully elaborated nonaketide such as **41**, which could undergo a biological Diels–Alder reaction to form dihydromonacolin L (**42**) – the observed first PKS-free intermediate. In order to confirm this, *lovB* was expressed in the heterologous fungal host *Aspergillus nidulans*.[16] Surprisingly, the polyunsaturated compounds **43** and **44** were isolated. These pyrones are related to the expected nonaketide, for example, the methylation has occurred at the correct position (i.e., after the third extension reaction), but it is evident that reductions – specifically enoyl reductions, and later keto reductions – have not occurred correctly and that chain extension has terminated prematurely (**Scheme 4**).

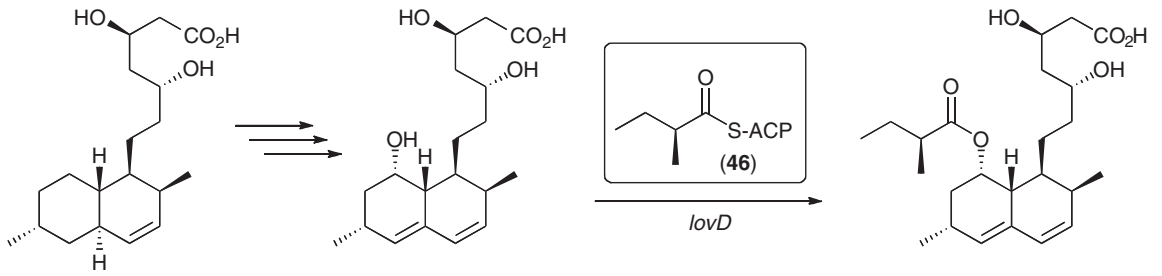
Close inspection of the sequence of LNKS suggests that the NADPH binding pocket of the ER domain might be dysfunctional, explaining the lack of observed ER activity. An enoyl reductase encoding gene, *lovC*, occurs downstream from *lovB*, and Vederas and coworkers showed that in coexpression experiments the lovC protein could complement the inactive ER domain and the expected dihydromonacolin L (**42**) was produced (**Scheme 4**). It is evident that the lovC protein must interact with LNKS intimately, drastically affecting programing by ensuring enoyl reduction at the correct positions and allowing complete chain extension. It now appears that lovC-type proteins are a common feature of HR PKS systems in fungi, although by no means ubiquitous.

The fact that 'defective' ER domains are not removed from the HR PKS where lovC-type proteins are used to achieve programed enoyl reductions may be significant. It is evident that the lovC protein must dock with the PKS in order to effect reduction, and the rump ER in the PKS sequence may act as a docking anchor; however, protein experiments have not yet been carried out to prove this hypothesis. Similar interactions between type I PKS and monofunctional proteins also occur in bacterial systems, for example, between multimodular and monofunctional proteins involved in mupirocin biosynthesis in *Pseudomonas fluorescens*.[27]

Recently, a further example of the effect of a lovC-type protein on programing of a fungal PKS has been described. In the case of tenellin biosynthesis (see Section 1.09.4), a similar pattern of sequence was observed in the tenellin biosynthetic gene cluster, that is, the ER domain of the PKS appeared defective and an 'extra' lovC-like ER gene was located nearby.[28] Expression of the tenellin synthetase alone led to the production of several compounds with altered polyketide programing, in particular numerous methyl isomers were observed by LCMS as well as less-reduced and chain-shortened analogues. Coexpression of the tenellin synthetase with the lovC homologue led to exclusive production of the correctly programed polyketide. It therefore appears that the exogenous ER domains may play a more general role in assisting in the programing of HR PKS proteins.



Scheme 4 Proposed and observed products of LNKS.



Scheme 5 Transfer of the lovastatin diketide (**46**) from its PKS to monacolin-J (**45**) by lovD.

In the bacterial modular PKS systems there is usually a C-terminal thiolesterase (TE) domain involved in off-loading the product, either to another PKS or to solution.[4] LNKS appears to possess a part of a nonribosomal peptide synthetase (NRPS) condensation (C) domain immediately downstream of the ACP, and this domain has been proposed to be involved in product release, presumably by either activating water as a nucleophile or the C-5 hydroxyl (rather than the nitrogen of an aminothiolester as activated by most NRPS C domains).

LDKS is closely related to LNKS, but its ER domain appears to be intact. LDKS is unusual among fungal PKS because it is not iterative – a single round of extension and processing affords the diketide (**46**) (**Scheme 5**). In this respect, LDKS closely resembles a single module of a bacterial modular PKS.[4] LDKS lacks an obvious product release domain, ending immediately after the ACP. A specialized AT, encoded by *lovD*, transfers **46** from LDKS onto the C-10 hydroxyl of monacolin J (**45**).[16] Tang and coworkers expressed *lovD* in *Escherichia coli* and showed that the recombinant protein will also transfer various acyl groups from CoA to the C-10 hydroxyl of **45** (**Scheme 5**).[29]

Recently, LNKS has been subjected to remarkable dissection by the group of Tang. LNKS itself is too difficult to work with, so Tang and coworkers cloned individual catalytic domains and expressed these in *E. coli*.[30] Two didomains consisting of the KS–AT and ACP–C regions, and three monodomains (AT, ACP, and C) were successfully expressed as soluble proteins. The KS, AT, and ACP domains appeared to be biochemically functional – interestingly, these studies showed that the AT domain selectively transfers malonyl groups to the KS, rather than acetyl, leading to the conclusion that the starter units may derive by decarboxylation of malonyl-ACP, as is observed during biosynthesis of octaketides by the bacterial (but also iterative) actinorhodin type II PKS.[31] This proposal is not, however, consistent with previous incorporation studies[25] with (1-^{13}C, 2-$^{2}H_3$)-acetate, which shows that both methyl starter groups in the lovastatin diketide and nonaketides retain all three hydrogens and thus are derived from acetate directly and not via malonate.

1.09.3.1.2 The squalestatin S1 polyketide synthases

Squalestatin S1 (**31**) is a potent inhibitor of mammalian squalene synthase that also possesses unprecedented curative properties toward prion-infected neurons.[32] It is produced by *Phoma* species, and like lovastatin (**19a**), consists of two polyketide chains: a main chain hexaketide and a side chain tetraketide. Like lovastatin, both chains are methylated, but unusually for a fungal HR polyketide, the main chain is formed from a nonacetate starter unit – benzoate is incorporated at this position.

Cox *et al.*[19] cloned an HR PKS gene from *Phoma* sp., *PhPKS1*. This was expressed in the heterologous fungal host *Aspergillus oryzae*, which produced the tetraketide (**47**) (**Figure 3**). The gene *PhPKS1* thus encodes the squalestatin tetraketide synthase (SQTKS). SQTKS is highly homologous to LDKS, lacking the C-terminal

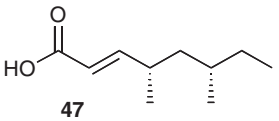


Figure 3 The squalestatin tetraketide.

NRPS condensation domain of LNKS. Like LDKS, SQTKS possesses a functional ER domain, but SQTKS carries out three extensions. Like LDKS, all modification reactions occur after the first extension. The stereochemistry of the branching methyl group is the same in each case. Two further extensions occur – all modifying reactions occur again after the first of these, but neither ER nor *C*-MeT is used after the final extension. Sequence comparisons of LDKS and LNKS, however, reveal few clues as to the structural basis for the difference in programing between the two.

1.09.3.1.3 The fumonisin and T-toxin polyketide synthases

T-toxin (**20**) and fumonisin B1 (**21**) are among the longest fungal polyketides. Fumonisin B1, a nonaketide, is produced by *Gibberella fujikuroi* (anamorph *Fusarium moniliforme*, syn. *Fusarium verticilloides*) and is a mycotoxin responsible for several human and animal diseases. T-toxin, a C_{41} linear polyketide, is produced by race-T of the maize pathogen *Cochliobolus heterostrophus*. T-toxin (**20**) is unbranched and its hydroxyls are unmodified, while fumonisin B1 (**21**) is methylated and esterified to two citric acid moieties.

T-toxin was among the first fungal compounds to be linked to a particular PKS gene.[20] More recently it has been shown that two PKS genes (*pks1* and *pks2* encoding TTS1 and TTS2, respectively) are involved in its biosynthesis; unusually *pks1* and *pks2* are not clustered on the *C. heterostrophus* genome.[21] Domain analysis of the two PKS genes indicates that both are typical HR PKS and that neither possesses a SAT domain – they differ in this respect from other fungal PKSs known to use multiple synthases, such as norsolorinic acid synthase (NSAS) and zearalenone synthase (ZR) (see Section 1.09.3.3.1). TTS1 (2528 aa) appears to possess a *C*-MeT domain (see Section 1.09.3.3.1), despite T-toxin not requiring C-methylation, while TTS2 (2144 aa) does not appear to possess a *C*-MeT domain.

The fumonisin biosynthetic genes of *G. fujikuroi* are clustered. The PKS fumonisin synthase (FUMS) is encoded by *fum1* (previously known as *fum5*), which appears to be a fairly typical PKS of the HR type, with KS, AT, DH, *C*-MeT, ER, KR, and ACP domains.[22] It appears that unlike the case of LNKS and the tenellin synthetase where the ER domains are inactive, the ER of FUMS is probably fully functional – consistent with this observation is the lack of a *lovC* homologue in the gene cluster. It appears that FUMS operates a fairly simple program – chain length is controlled to 18 carbons and, apart from the *C*-MeT, all activities operate in all cycles. Methylation is the only programed event, occurring during the second and fourth extension cycles only.[33]

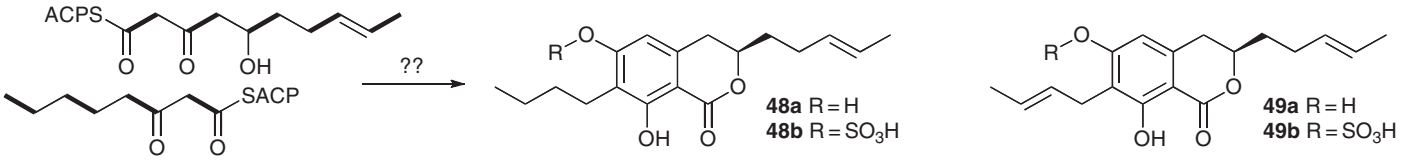
The function of the TTS and FUMS genes has been proven by knockout experiments, but none of the genes has yet been reported as expressed heterologously. However, FUMS has been the target for the first HR PKS-directed domain swap experiments. In the initial work, Du and coworkers described the development of a method to exchange regions of *fum1* with homologous regions of *C. heterostrophus pks1*. In effect the KS domain from TTS1 was inserted into FUMS. This domain swap appeared to have no effect on either levels of production of fumonisin B1 (**21**), or on programed events during biosynthesis.[34] Replacement of the whole of FUMS by TTS1 resulted in no production; however, as it is now known that two PKS are required for T-toxin production it is unsurprising that this experiment produced no new compounds.

Further work, however, in which the FUMS KS was replaced with the LDKS KS possibly resulted in significant reprograming of the PKS. The new chimeric PKS appeared to produce two unusual dihydroisocoumarins, **48** and **49** (**Scheme 6**).[35] These compounds appear to be the result of the condensation of two separate polyketide chains – in each case a tetraketide with a pentaketide. It is not clear what underlies these changes in programing, or the unusual dimerization, but it does suggest that the LDKS and FUMS KS domains are, at least partially, involved in the molecular interactions that help specify the chain length and the use of reductive domains. However, caution should be exercised in interpreting the results of these experiments as it cannot be proven that the compounds do arise from the modified PKS – it is possible that other PKS in the host organism have been upregulated in some way.

1.09.3.1.4 Highly reduced polyketide synthase from Alternaria solani

Alternaria solani is a plant pathogen and the causative agent of early blight in solanum species. It produces numerous polyketides such as solanapyrone A (**50**) and alternaric acid (**51**), and is thus an ideal target species for speculative PKS gene-fishing expeditions (**Figure 4**).

Fujii and Ebizuka have conducted such investigations, using PCR primers based on conserved PKS sequences as probes with genomic DNA libraries. An early investigation yielded two hits – one an HR PKS gene named *alt5*, and another, an NR PKS named *pksA*.[23] The *alt5* gene encodes a typical HR PKS, known as



Scheme 6 New compounds from genetic engineering of *F. verticilliodes*.

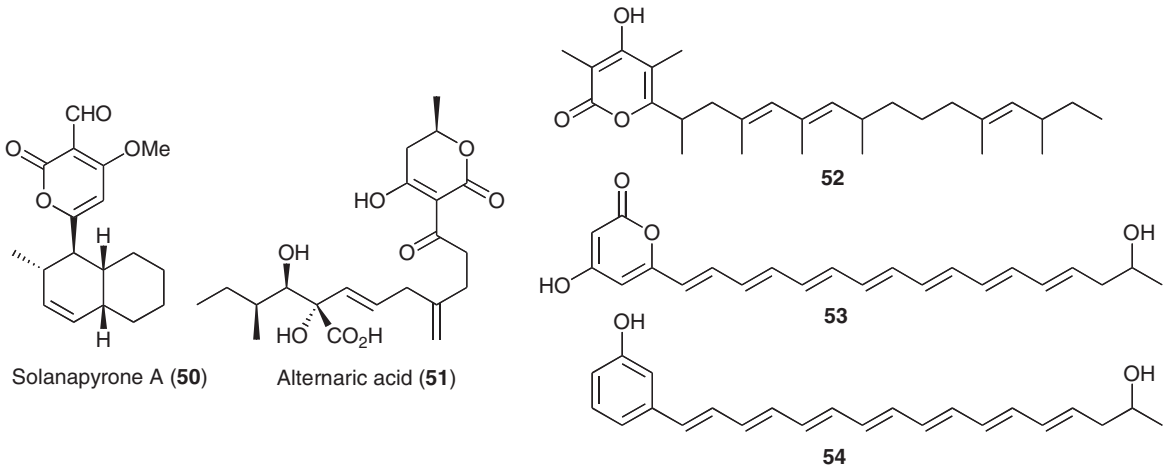


Figure 4 Compounds from *Alternaria* HR PKS.

PKSN, with the usual array of catalytic domains. Inspection of the ER sequence suggested that it should be functional like those from LDKS and SQTKS. Expression of *alt5* showed this to be correct – a single compound was synthesized in good yield ($\sim$15 mg l^{-1}), which proved to be the octamethylated decaketide pyrone (**52**), named alternapyrone. This compound is the most complex polyketide yet reported to be produced by an iterative PKS, showing programed chain length control, keto reduction, methylation, and enoyl reduction.

Further gene probing of *A. solani* afforded two more HR PKS, PKSF and PKSK encoded by *pksF* and *pksK*, respectively. Unlike PKSN, these synthases did not seem to possess *C*-MeT domains, but were otherwise similar to other HR PKS. PKSK proved to be inactive when expressed in *A. oryzae*, but PKSF did produce the new compounds aslanipyrone (**53**) and aslaniol (**54**) – albeit in low yields, together with nine other related minor compounds.[24] These polyunsaturated compounds must arise because of the inactivity of the ER domain, and detailed sequence analysis suggested that the NADPH binding motif of the ER domain may be defective in PKSF. It is not yet known whether the PKSF cluster possesses a *lovC* homologue as the lovastatin and tenellin clusters do, but coexpression of PKSF with such a gene could form new polyketides.

1.09.3.2 Partially Reduced Polyketides

Less is known about the enzymology of the PR PKS. The domain structure is much closer to mammalian fatty acid synthase (FAS), with an N-terminal KS followed by AT, and DH domains. A so-called 'core' domain follows the DH, and this is followed by a KR and the PKS terminates with an ACP domain but there is no obvious catalytic machinery for off-loading the product. The DNA sequence of the KS domain is distinguishable from NR PKS and FAS KS domains using selective PCR primers and DNA probes.[7,8] Overall, however, catalytic domains closely match those of the mammalian FAS, although the 'core' domain is different.[36] Remarkably, the fungal MSAS genes are closely related to bacterial genes for the synthesis of the NR tetraketide orsellinic acid (**2**).[37] For example, the iterative type I PKS from *Micromonospora echinospora*, calO5, appears to be homologous to the fungal PR PKS proteins apart from lacking a $\sim$450 amino acid region encompassing the KR domain.

Although a number of PR PKS genes are known from genome sequencing projects, only three genes have been matched to chemical products – in all cases the tetraketide 6-MCA **1** (**Table 2**). The first 6-MSAS to be discovered was from *Penicillium patulum*, encoded by *MSAS*.[36] The Ebizuka group have worked with the *atX* gene from *A. terreus*,[38] and most recently Tkacz and coworkers have described an MSAS gene (*pks2*) isolated from *Glarea lozoyensis*.[39]

MSAS from *P. patulum* is one of the smallest type I PKS at 191 kDa. It proved relatively easy to isolate, and many of the earliest *in vitro* studies of the enzymology of any PKS were carried out with this enzyme.[40,41,42] MSAS requires acetate as a starter unit – unlike THNS and the type II actinorhodin PKS, malonate cannot be

Table 2 Examples of proven links between PR PKS and known compounds

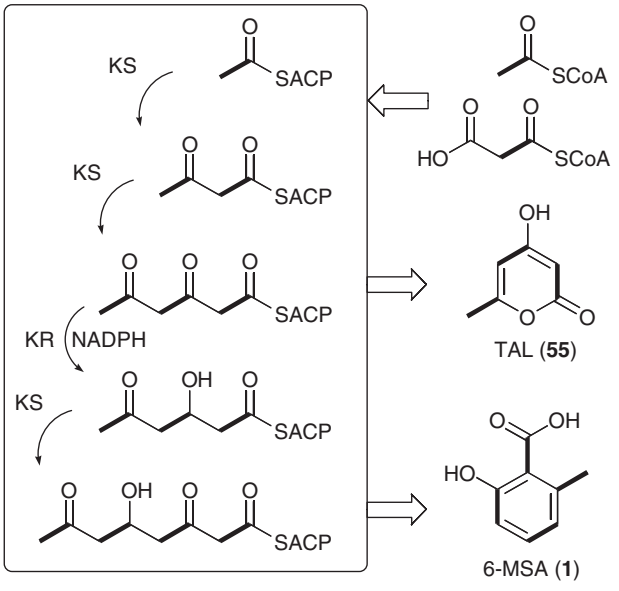
Organism	*Gene*	*Protein*	*Product*	*Reference*
Penicillium patulum	*MSAS*	MSAS	6-MSA (**1**)	36
Aspergillus terreus	*atX*	MSAS	6-MSA (**1**)	38
Glarea lozoyensis	*pks2*	MSAS	6-MSA (**1**)	39

used as a starter. The AT domain loads malonate from CoA to use as an extender. The PKS is evidently programed – acetate must be extended three times and the KR must only act once, after the second extension. When NADPH is unavailable the KR reaction cannot occur. Under these circumstances a tetraketide is not produced and triketide lactone (TAL) (**55**) is produced instead (**Scheme 7**). This reveals that chain extension and reduction are linked, and indicates that the KR must act during chain extension. This is different to the C-terminal R domain of 3-methylorcinaldehyde synthase (see Section 1.09.3.3.8), which must act after extension and cyclization.[43] Mammalian FAS also produces TAL (**55**) under the same circumstances. The similarities between MSAS and FAS were further exemplified by Spencer and Jordan who showed that the KS and ACP components of different peptide chains in dimeric MSAS species could be cross-linked using thiol-specific reagents.[44]

Chain length determination in MSAS appears to use a 'counting' mechanism. When acetyl-CoA is unavailable as a starter unit other acyl-CoAs can substitute and be extended. In the case of the type II act PKS, experiments of this type show that the length of the starter unit is important so that C_{16} chains are preferred products despite the length of the starter unit.[45] In the case of MSAS denied NADPH and acetyl-CoA, but incubated with malonyl-CoA extender units and various acyl-CoA starter units, two extensions always occurred to produce substituted TALs.[46]

The expression of MSAS in heterologous hosts has allowed more detailed experiments to be performed. Hopwood, Khosla, and coworkers first described the expression of *P. patulum* MSAS in the Gram-positive bacterium *Streptomyces coelicolor*.[47] Ebizuka, Fujii, and coworkers expressed *A. terreus atX* in *A. nidulans*.[38] However, the most useful expression system to date has proven to be the yeast *Saccharomyces cerevisiae*.[48,49]

Fujii and coworkers developed a *S. cerevisiae* expression system in which two independent copies of *atX* can be produced simultaneously. This allowed them to mutate each copy of *atX* in complementary ways (**Figure 5**). Initial experiments determined how much of the N- and C-termini of MSAS could be deleted without affecting



Scheme 7 Chemical reactions performed by MSAS.

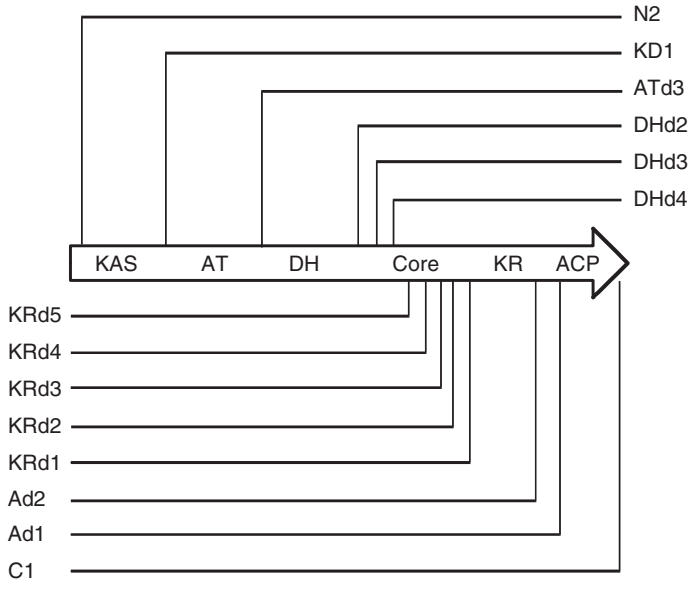


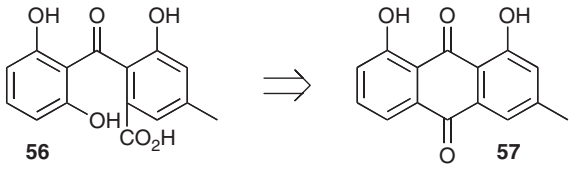
Figure 5 Approximate locations of truncation points in MSAS.

6-MSA production. It was found that at the N-terminus a deletion mutant lacking 44 amino acids (*atXN1*) was still active, but that a deletion of 81 amino acids (*atXN2*) prevented 6-MSA formation by removal of key KS residues. At the C-terminus, deletion of as few as nine amino acids (*atXC1*) rendered the MSAS inactive by removal of key ACP residues.

Both *P. patulum* and *A. terreus* MSAS form homotetramers. Fujii and coworkers exploited this fact in the reconstitution of functional MSAS complexes formed from combinations of the deletion mutants. For example, coexpression of *atXN2* with *atXC1*, both of which are inactive in isolation, led to the reconstitution of an active MSAS complex. This shows that the KS of one peptide chain must interact with the ACP of another peptide chain. The number of deletion constructs was then increased to examine a total of six different length deletions from the N-terminus and eight different length deletions from the C-terminus (**Figure 5**). Remarkably, almost all combinations of the truncated MSAS proteins proved functional, showing that very severe truncations can be complemented if the activity is present on another peptide chain. However, a short region of the core domain was identified, the presence of which proved essential for successful complementation. It is hypothesized that this region of 122 amino acids probably forms a motif required for subunit–subunit interaction. This core sequence is present in other fungal PR PKS, and in the bacterial PKS such as CalO5. Mammalian FAS also has a central core region, which has been shown to mediate the assembly of the synthase;[50] however, there are no sequence similarities between the FAS and MSAS core regions.

These truncation studies were then complemented with site-directed mutations in the yeast-expressed *atX* genes.[51] As expected, removal of key residues in KS (C216A), AT (S667A), and ACP (S1761A) yielded inactive synthases. Interference with the KR domain (GxGxxG to AxPxxA) led to production of TAL (**55**), as expected, but a mutation in the DH domain (H972A) did not lead to the production of an expected triketide alcohol, or other detectable products, indicating the previously unsuspected key role of H972. Next, coexpression of dual *atX* clones was achieved so that each clone carried a different active site mutation. In all cases, 6-MSA (**1**) biosynthesis was reconstructed, confirming the previous results. Remarkably, further experiments showed that activity could be reconstructed for mutations in which one clone of *atX* carried a single mutation while the other clone carried the four complementary mutations.

Genetic engineering has also been used to increase the titer of MSAS in yeast. Nielsen and coworkers engineered *S. cerevisiae* such that it contained a copy of the *acc1* gene encoding acetyl-CoA carboxylase under the control of a strong promoter, in order to increase the metabolic supply of malonyl-CoA. The strain also contained the PPTase from *A. nidulans*, and when *P. patulum* MSAS was expressed a final titer of 6-MSA (**1**) of $>550\,\mathrm{mg\,l^{-1}}$ was obtained, 60% greater than in an unmodified host.[52]



Scheme 8 Possible biosynthetic route to octaketide benzophenone (**56**).

Tkacz and coworkers expressed the *G. lozoyensis pks2* gene in *A. nidulans*,[39] replicating the earlier work of Fujii and coworkers with *atX*.[38] Tkacz showed that the expression of *pks2* led to the formation of 6-MSA (**1**), but a minor metabolite, the octaketide benzophenone (**56**), was also produced. It is not yet clear how this compound arises. Tkacz and coworkers suggested several possibilities such as induction of an as yet unidentified *A. nidulans* octaketide synthase gene by 6-MSA (although this did not happen in Fujii's experiments with *atX*) or possible positional effects of the insertion of the *trpC* promoter. However, a very intriguing possibility is that an N-terminal intron in *pks2* can be processed in two different ways (Tkacz showed that 42% of *pks2* transcripts have an alternative mode of intron removal) leading to two different MSAS N-terminal sequences. It is conceivable that one of these could be programed differently to produce an octaketide such as **57**. Intriguingly, **57** is also partially reduced (**Scheme 8**).

1.09.3.3 Nonreduced Polyketides

One of the first discovered fungal PKS, orsellinic acid synthase (OSAS), was isolated from *Penicillium madriti* and reported in 1968.[53] The tetraketide orsellinic acid (**2**) is the simplest tetraketide, requiring no reductions during its biosynthesis. Despite the early work with the protein, however, the OSAS encoding gene has not yet been discovered and nothing is known of the catalytic domains or their organization. However, genes involved in the biosynthesis of a number of other NR polyketides are now known (**Table 3**) and a general pattern of domain organization has emerged. In all cases known to date these genes encode type I iterative PKS proteins. At the N-terminus is a starter unit, ACP transacylase (SAT) domain, which mediates the loading of a starter unit (**Figure 6**). It appears that the starter unit can derive from either a dedicated FAS, another PKS or an

Table 3 Examples of proven links between NR PKS and known compounds

Organism	*Gene*	*Protein*	*Product*	*Reference(s)*
Aspergillus parasiticus	*pksA*	NSAS	Aflatoxin B_1	54
Aspergillus nidulans	*pksST*	NSAS	Sterigmatocystin	55
Dothistroma septosporum	*pksA*	NSAS	Dothistromin	56
Aspergillus nidulans	*wA*	WAS	YWA1 (**7**)	55
Aspergillus fumigatus	*alb1*	alb1p	YWA1 (**7**)	57
Colletotrichum lagenarium	*PKS1*	THNS	Tetrahydroxy naphthalene[a] (**6**)	58
Wangiella dermatitidis	*WdPKS1*	THNS	Tetrahydroxy naphthalene[a] (**6**)	59,60
Gibberella zeae	*PKS13*	ZS-B	Zearalenone (**58**)	18
Monascus purpureus	*pksCT*	CitS	Citrinin (**4**)	61
Cercospora nicotianae	*CTB1*	CRCS	Cercosporin	62
Giberella fujikuroi	*pks4*	BIKS	Bikaverin (**72**)	63

[a] Likely product is the hexaketide acetyltetrahydroxynaphthalene (**68**).

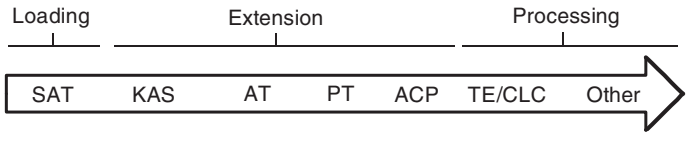


Figure 6 The general domain organization of an NR PKS.

acyl-CoA. The SAT domain is followed by typical KS and AT domains responsible for chain extension and malonate loading. Beyond the AT is a conserved domain originally proposed to be a product template domain (PT)[64] and later demonstrated to catalyze specific cyclization(s) of the mature ACP-bound poly β-keto intermediate.[65] This is followed, in turn, by an ACP. Some NR PKS appear to terminate after the ACP, but many feature a diverse range of different domains including cyclases, methyl transferases, and reductases. It thus appears that these synthases are arranged with an N-terminal 'loading component,' a 'chain extension,' and 'cyclization component' consisting of KS, AT, PT, and ACP domains with possible control over the number of extensions, and a C-terminal 'processing component'.

These three components correspond to the three important elements of programing – starter unit selection, chain length determination and cyclization, and post-assembly PKS processing. These features are exemplified below. There are now several examples of fungal NR PKS genes, which have been definitively linked to chemical structure (**Table 3**).

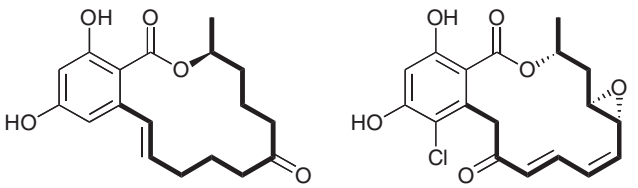
1.09.3.3.1 Nonreduced polyketide synthase loading component

Feeding experiments with isotopically labeled,[66] and in some cases ^{19}F-labeled,[67] precursors have shown that many NR fungal polyketides are formed by the use of 'advanced' starter units. For example, in the classic case of norsolorinic acid (**8**) biosynthesis it has long been known that hexanoate forms the starter unit. Differential specific incorporations of acetate into the early and late positions in compounds such as citrinin (**4**) have been used to argue that these compounds may have been formed by more than one PKS so that one PKS makes an advanced starter unit, which is passed to a second PKS for further extension.[68]

Dehydrocurvularin[69] (**17**), monocerin[70] (**18**), and zearalenone[71] (**58**) are all good potential examples of this; here, the structures of the compounds indicate high levels of reduction early during biosynthesis and no reductions during later steps. In the case of dehydrocurvularin (**17**), advanced labeled tetraketide precursors were incorporated during biosynthesis, possibly indicating that a tetraketide formed by an HR PKS forms the starter unit for a second NR PKS.[72,73] More recently, definitive examples of this type of HR–NR dual PKS have been elucidated, for example, in the cases of radicicol (**59**) and related compounds where an HR and an NR PKS are found in the gene clusters (**Figure 7**)[74] . In the case of zearalenone (**58**) the activity of the NR PKS component has been confirmed *in vitro*.[75] The NR zearalenone PKS, encoded by *G. zeae pks13*, was expressed in *E. coli*, apparently without the need to engineer codon usage, and obtained as pure protein in 2 mg l^{-1} yield. The PKS is fully functional *in vivo* and *in vitro* (indicating compatibility with the *E. coli* phosphopantetheine transferase), utilizing a number of acyl SNACs and malonyl-CoAs as substrates. It has selectivity for C_{10} acyl starter groups, but can process starter units of between 6 and 16 carbons.

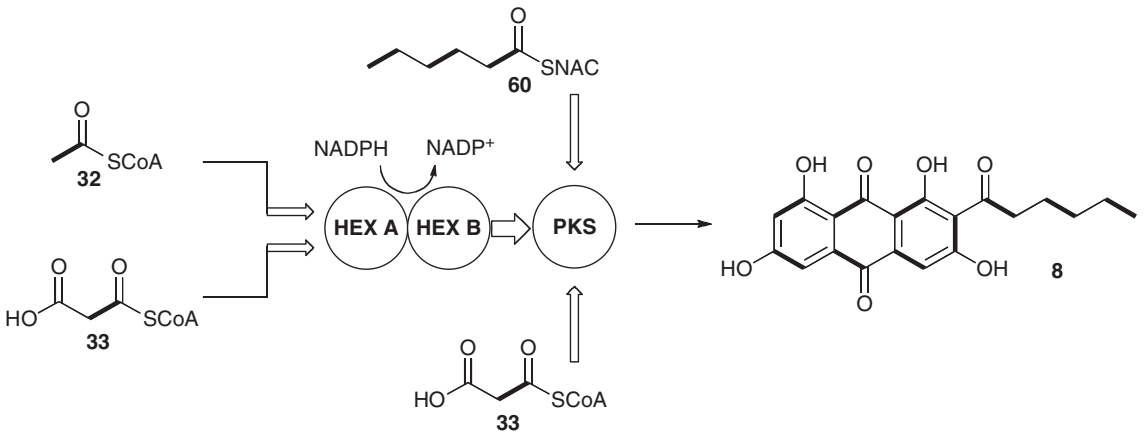
Townsend has defined the molecular basis for the ability of NR PKS to utilize starter units derived from other FAS or PKS systems. Two genes in the aflatoxin biosynthetic gene cluster of *A. parasiticus* (*aflA* and *aflB*) encoded the α and β components of a typical fungal FAS (HexA and HexB).[76] Clustering of these FAS genes with the NSAS PKS suggested that HexA and HexB probably produced hexanoate for use as the norsolorinic acid starter unit.

The protein complex formed between NSAS, HexA, and HexB, known as NorS, was then isolated and characterized.[77] This ~1.4 MDa protein complex synthesizes norsolorinic acid (**8**) from acetyl-CoA, malonyl-CoA, and NADPH. Townsend and coworkers were unable to detect hexanoyl-CoA as a free intermediate produced in NorS reconstitution experiments, suggesting that hexanoate produced as an ACP derivative by the



Zearalenone (**58**) Radicicol (**59**)

Figure 7 Zearalenone and radicicol, products of combined HR and NR PKS. Bold bonds indicate likely product of HR PKS.

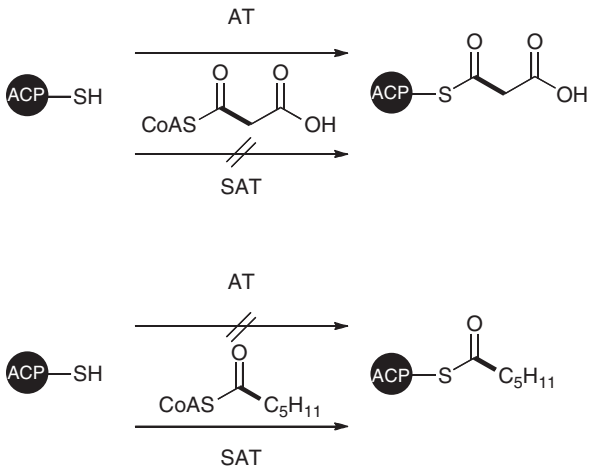


Scheme 9 Biosynthesis of norsolorinic acid by the NorS complex.

HexA/HexB FAS must be passed directly to NSAS. In the absence of NADPH, hexanoate cannot be formed by the FAS components and thus no norsolorinic acid is formed. Norsolorinic acid biosynthesis is restored, however, when exogenous hexanoyl SNAC (**60**) is used. However, the extreme difficulty of working with a protein complex of ~1.4 MDa consisting of three very large proteins at WT concentrations precluded further studies into the transfer mechanism (**Scheme 9**).

Complementary experiments carried out by Cox and coworkers involved the cloning and heterologous expression of the NSAS AT and ACP catalytic domains.[78] These were shown to be properly folded and catalytically active *in vitro*. The AT domain catalyzed the transfer of malonyl extender groups from CoA onto ACP, but did not appear to transfer potential acyl starter groups from CoA or FAS ACP ruling out its involvement in starter unit provision.

Townsend realized that the N-terminal domain of NSAS possesses canonical AT sequence motifs and this domain could be a candidate for the required starter unit transferase.[79] The N-terminal domain and ACP were cloned and heterologously expressed. *In vitro*, the N-terminal protein was shown to catalyze the transfer of hexanoate from CoA onto the ACP. Site-directed mutagenesis experiments to remove the proposed catalytic cysteine of the transferase resulted in the loss of catalytic activity. The N-terminal transferase showed a significant selectivity for the transfer of hexanoate over longer or shorter acyl chains. Thus, the N-terminal domain of NSAS acts as a SAT component (**Scheme 10**).



Scheme 10 Complementary activities of AT and SAT domains in NSAS.

Sequence comparison with other known NR PKS suggests that such SAT domains are common.[80] In the few cases where the PKS sequence has been correlated with product structure, the presence of SAT domains now explains prior results from feeding experiments that suggested the use of advanced starter units. For example, it is now known that two PKS genes are involved in the biosynthesis of zearalenone (**58**)[18] – one of these is an HR PKS and probably provides an HR hexaketide as a starter unit. The second zearalenone PKS is an NR PKS possessing an N-terminal SAT domain, which likely loads the hexaketide ready for three further extensions. Although genetic studies have not been carried out, labelling studies with advanced precursors provide strong evidence that the biosynthesis of the related dehydrocurvularin (**17**) does indeed involve the use of a HR polyketide as the starter unit.[81]

Similar SAT domains are present in the citrinin (**4**) PKS, making it likely that citrinin is formed by a diketide starter unit, extended three times by an NR tetraketide synthase – this validates results from isotopic feeding experiments.[68] Further work by Townsend and coworkers have extended this theory. Genome mining of known fungal genomes revealed PKS genes homologous to NSAS in two *Coccidioides* species. Heterologous expression of the SAT domains and *in vitro* assay showed that these domains selected and transferred octanoate fourfold faster than hexanoate and indicate that the *Coccidioides* species probably produce NR polyketides derived from C_8 precursors.[80]

However, most NR PKS appear to possess potential SAT domains whether they require an acetate starter unit or not; for example, the PKS involved in the biosynthesis of Ywa1 (WAS) (**7**) and tetrahydroxynaphthalene (THNS) (**6**). In the case of THNS from *C. lagenarium*, *in vitro* experiments suggested that the purified protein utilizes malonyl-CoA as the starter unit.[82] This initially suggested that the THNS SAT domain may be involved with loading of malonate to use as a starter unit. Bacterial type II PKSs use malonate as a starter unit, but here the KS_β component decarboxylates it before chain extension.[83] However, Townsend's work has again been instrumental in exposing the underlying chemistry of acetyl starter unit selection in this case. Four SAT domains were cloned from PKSs associated with the biosynthesis of THN (**6**), Ywa 1 (**7**), cercosporin,[62] and bikaverin (**Table 3**). All four SAT domains were purified and their activities analyzed *in vitro*.[84] All four transferred acetate, but not malonate. This result, therefore, conflicts with the prior report of Ebizuka and coworkers' *in vitro* experiments with partially purified THNS.[82]

A unifying model for THN (**6**) biosynthesis, proposed by Townsend and coworkers, would have THNS as a hexaketide synthase, producing acetyl-THN (**68**) as its product (see **Figure 9**, Section 1.09.3.3.5). The pendant acetyl group, derived from the starter acetate, could be removed by retro-Claisen reaction *in vivo*, similar to the Ayg1p chemistry discussed in Section 1.09.3.3.5. This idea has been validated by Wheeler, Ebizuka, and coworkers, who have shown that THNS from *Wangiella dermatitidis* most likely synthesizes acetyl-THN (**68**) itself, as this compound accumulates when the gene *WdYG1*, encoding a homologue of Ayg1p, is knocked out.[59]

1.09.3.3.2 Nonreduced polyketide synthase chain extension component

The extension components of NR PKSs consist of KS, AT, and ACP domains. Little is known about the enzymology of the chain extension components – certainly in comparison to the KS, AT, and ACP components of the bacterial type I modular PKS and the type II iterative PKS. Once again the problem of investigating these systems lies in the large size of the type I proteins and the difficulty of obtaining sufficient protein for enzymology studies.

These problems have begun to be overcome, however, by the heterologous expression of individual catalytic domains. Heterologous expression offers the possibility of obtaining large quantities of protein, but ensuring correct folding of the recombinant proteins can be a hit-and-miss process. It appears that correct determination of cut sites in the linker regions near domain boundaries is important for ensuring correct folding during *E. coli* expression. Two approaches have been taken to determining the likely C-terminal and N-terminal boundaries. Cox and coworkers have used a simplistic model in which the peptide sequences of type II PKS components are aligned with the type I NR PKS and the N and C termini of the type II proteins are assumed to correspond with the type I domain boundaries.[78] Townsend has described a more sophisticated model involving a bioinformatic approach that assesses possible secondary structure, hydrophobicity, and family homology to identify the flexible regions found between domains.[64]

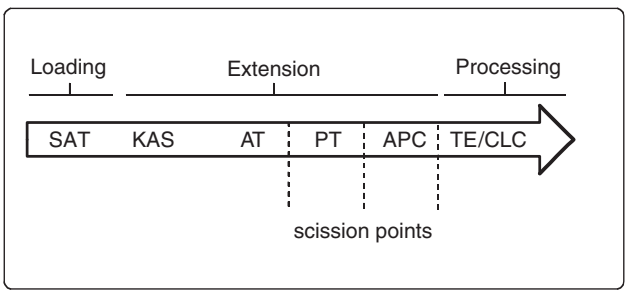
The Cox group has successfully expressed and studied the AT and ACP components of NSAS.[78] The AT appears to act as a relatively standard malonyl transferase, transferring malonate from CoA onto the NSAS

holo-ACP, but not transferring other acyl groups such as acetate or hexanoate. The ACP component reacted with bacterial ACPS enzymes to generate *holo*-ACP. The *holo*-ACP was also able to receive malonyl groups via the bacterial *S. coelicolor* AT (also known as MCAT) as well as via the NSAS AT. Remarkably, the excised NSAS *holo*-ACP domain is also active in polyketide biosynthesis assays with the KS components of the bacterial type II actinorhodin PKS. All these facts suggest that the NSAS AT and ACP behave much as do the type II iterative PKS systems found in bacteria.

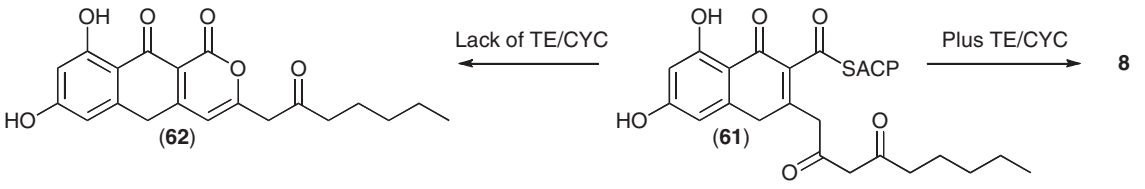
The type II iterative PKS systems found in bacteria, however, differ from the type I systems in that they contain a chain initiation factor (CIF, also known as KS_β or chain length factor) protein, which is involved in the decarboxylation of malonate to form acetate, which then acts as the starter unit for polyketide biosynthesis.[31,83] The CIF forms an $\alpha\beta$ dimer with the KS protein and this dimeric species is able to determine chain length (i.e., the number of extensions).[85] The type I PKS of fungi lacks the CIF component, but it is conceivable that the PT domain is involved in chain length determination. Sequence analysis[43] of PT domains (450–550 residues) from a range of NR PKS genes where the chemical products are known suggests that this may well be the case. For example, comparison of the PT domains from the *Acremonium strictum* PKS1, citrinin PKS, zearalenone PKS-B, NSAS, sterigmatocystin PKS, dothistromin PKS, THNS from *C. lagenarium* and *W. dermatitidis*,[60] and WAS suggests that these group into clades that correspond with chain length. Thus, the citrinin, ASPKS1, and zearalenone-B group correspond to tetraketide synthases; the NSAS, sterigmatocystin, and dothistromin PKS are all octaketide synthases; the THNSs are hexaketide synthases (Section 1.09.3.3.1); and the wA synthase (WAS) forms an outgroup – this is a heptaketide.

Townsend has investigated the role of the PT and other domains, again by domain expression in *E. coli* followed by *in vitro* assays.[65] The previous work with individual domains of NSAS had shown the difficulty in obtaining individual isolated domains of NSAS, so combinations of mono-, di-, and tri-domains were produced and combined *in vitro* with hexanoyl and malonyl-CoA. HPLC analysis was then conducted to examine the products of the reactions (**Scheme 11**). From these results it was shown that likely intermediate (**61**) (corroborated by high-resolution mass spectrometry) can form either the pyrone (**62**) or norsolorinic acid (**8**) itself.

These experiments clearly show the effect of the PT domain – its presence vastly improves the efficiency of NSAS by strongly driving cyclization and dehydration to form the first two rings of **61**, leaving control of the final cyclization required to produce norsolorinic acid to the TE/Claisen cyclase (CLC) domain (see Section 1.09.3.3.4). Thus, PT appears to act similarly to the aromatase/cyclase family of PKS enzymes from bacterial type II systems. The exact roles of the PT and KS domains in control of chain length individually or in concert remain to be established.



SAT-KS-AT + PT + ACP + TE/CLC	: **8**
SAT-KS-AT + PT + ACP – TE/CLC	: **8**
SAT-KS-AT + PT–ACP + TE/CLC	: **8**
SAT-KS-AT + PT–ACP – TE/CLC	: **8**
SAT-KS-AT + PT + ACP	: **62**
SAT-KS-AT + PT – ACP	: **62**
SAT-KS-AT + ACP + TE/CLC	: **8 (min)**
SAT-KS-AT + ACP – TE/CLC	: **8 (min)**
SAT-KS-AT + ACP	: **8 + 62 (min)**
SAT-KS-AT + PT + TE/CLC	: no products
SAT-KS-AT	: no products



Scheme 11 Results of *in vitro* reconstruction of NSAS. Table shows major products formed.

1.09.3.3.3 Nonreduced polyketide synthase processing component

The domains found after the chain extension components at the C-termini of fungal NR PKS are highly varied. These include putative CLC/TE, methyl transferases (MeT), reductases (R), and additional ACP domains. All evidence collected to date suggests that these processing components act after chain assembly to modify either a polyketo or a cyclized intermediate, which is probably still bound to the ACP. This raises the intriguing possibility (see Section 1.09.3.3.7) that C-methylation in these systems may occur after rather than during polyketide assembly in these systems.

1.09.3.3.4 Claisen cyclase/thiolesterase domains

The first fungal NR PKS gene to be cloned was *A. nidulans* wA (encoding WAS) first reported in 1992.[86,87] At that time limited domain analysis was carried out to determine the presence of KS, AT, and ACP domians, but N- and C-terminal catalytic domains had not been recognized. The *A. nidulans stcA*[54] (encoding the sterigmatocystin NSAS) and the *A. parasiticus aflC*[55] (encoding the aflatoxin NSAS) were then reported in 1995. Yu and Leonard[54] realized that the 3′ sequence of *stcA* encoded a protein domain with homology to known TE domains of FAS and the ACV synthetase involved in β-lactam biosynthesis. Later, it was shown that *wA* and *pksA* also possess similar domains and TE domains were recognized as one of the most common processing components of NR PKS.

From the outset it was realized that the TE domain could either operate as a standard thiolesterase, or be involved in a cyclization release mechanism. The first experiment that investigated the role of the TE domain occurred when Ebizuka and coworkers expressed *A. nidulans wA* (encoding WAS) in *A. oryzae*.[88] In an initial experiment the expression strain produced the citreoisocoumarin (**63**) and the parent compound **64** indicating that WAS is a heptaketide synthase (**Figure 8**). The secondary alcohol of **63** is presumably formed by an adventitious reduction *in vivo* – WAS could not catalyze this reaction as it lacks reductive domains.

However, it was later realized that the expression construct used in this experiment had a deletion resulting in the expressed WAS missing the final 67 amino acids of the C-terminal TE domain. When the complete *wA* gene was expressed, however, the heptaketide naphthopyrone YWA1 (**7**) was produced.[89] A series of experiments involving step-wise shortening of the C-terminus of WAS showed that deletion of as few as 32 amino acids resulted in the production of the citreoisocoumarin (**64**). Site-directed mutagenesis of a conserved serine and histidine in the C-terminal domain also resulted in a switch from naphthopyrone production to citreoisocoumarin production.[90]

Isotopic feeding experiments using ^{13}C-labeled acetate showed both the citreoisocoumarin (**64**) and the naphthopyrone (**7**) must result from cyclization of the common intermediate.[90] This observation suggests that WAS produces a heptaketide and catalyzes the cyclization and aromatization of the first ring prior to product release by the TE/CLC. The C-terminal domain of WAS must therefore catalyze a second (Claisen) cyclization reaction to form the observed naphthopyrone (**7**). Thus, the TE domain has been renamed as CLC. These domains also occur in the known NSAS and THNS proteins, where the same chemistry must occur to provide the observed products.

1.09.3.3.5 Chain-shortening reactions

The *C. lagenarium PKS1* gene is involved in the biosynthesis of the previously assumed pentaketide-derived THN (**6**).[58] Ebizuka and Watanabe expressed *C. lagenarium PKS1* in the heterologous fungal host *A. oryzae* and, as expected, observed the production of THN (**6**). However, three additional compounds were produced: the tetraketide orsellinic acid (**2**), the pentaketide α-acetylorsellinic acid (**65**), and the derived dihydroisocoumarin (**69**).[91] Similar results were obtained *in vitro* using purified THNS and malonyl-CoA as the substrate.[82] In this case the system produced THN (**6**) and α-acetylorsellinic acid (**65**). In a subsequent study, heterologous

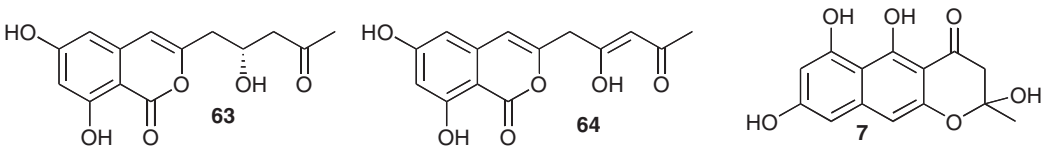


Figure 8 Products of expression of truncated WAS.

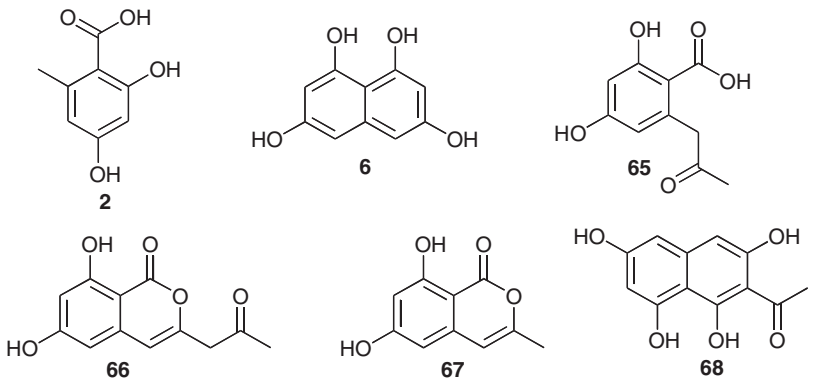
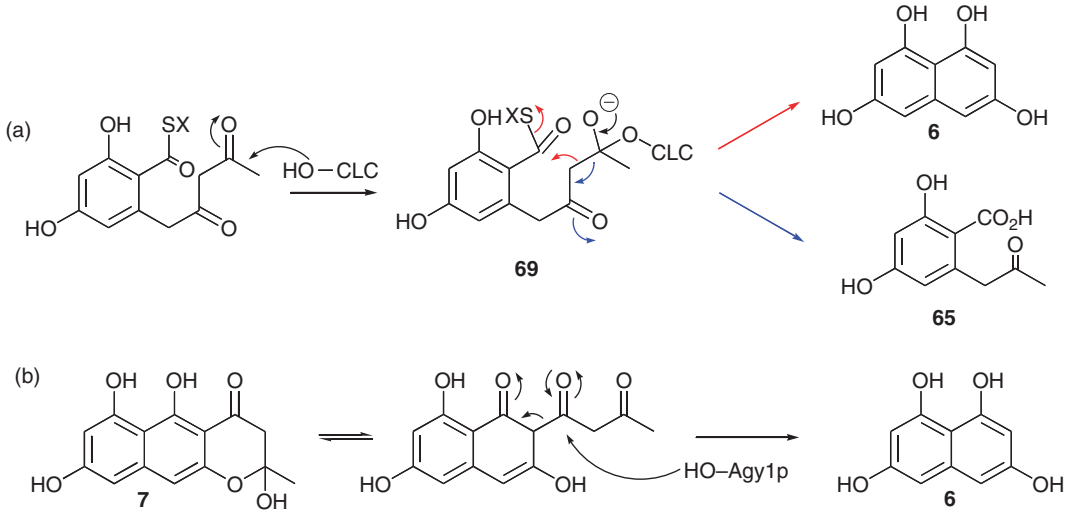


Figure 9 Compounds produced by THNS *in vivo* and *in vitro*.

expression of PKS1 produced a hexaketide, acetyl-THN (**68**, 15%), 'pentaketide' THN (**6**, 50%), pentaketide α-acetylorsellinic acid (**65**, 25%), and the tetraketide orsellinic acid (**2**, 10%) (**Figure 9**).[92]

Watanabe and Ebizuka[92] then generated two mutants of the *C. lagenarium PKS1* in which the CLC domain was completely removed or the active site serine of the CLC was mutated to alanine. In both cases, expression of the modified *C. lagenarium PKS1* genes resulted in the production of >95% hexaketide isocoumarin (**66**) and ~5% of pentaketide isocoumarin (**67**). It was proposed that the CLC domain was involved in chain length determination – that is, the prevention of chain extension to form a hexaketide. This mechanism must involve the active site serine of the CLC. However, consistent with other work that suggests that THN is actually formed via a hexaketide acetyl-THN (**68**),[84] it is likely that the role of the CLC is to generate the carbanion equivalent required for cyclization to give **6** by loss of acetyl as shown in **Scheme 12**(**a**) via a retro-Claisen reaction from the enzyme-bound monocyclic intermediate **69**, formed by attack of the CLC active site serine on the acetyl carbonyl group. Deacetylation rather than cyclization would also explain the formation of α-acetylorsellinic acid (**65**). Ebizuka and coworkers did in fact rationalize the formation of both compounds via alternative decarboxylative cyclization or simple decarboxylation from a common intermediate formed from a malonate-primed polyketide.[82,91] Watanabe and Ebizuka[93] further investigated the relationship between chain length and cyclization by synthesizing a fused PKS consisting of the N-terminus of *C. lagenarium* THNS and the C-terminus of *A. nidulans* WAS. The fusion site was chosen between the AT and ACP domains. The resulting chimeric PKS, known as SWB, produced the hexaketides **66** and **67** as well as the pentaketides **65** and **68**.[93]



Scheme 12 Chemistry catalyzed by (a) CLC and (b) Agy1p.

This result shows that the fundamental hexaketide selectivity of *C. lagenarium* THNS is preserved and that the WAS CLC domain is capable of cyclizing both hexaketides and heptaketides.

The original hypothesis that CLC domains are required for chain release must now be modified. Although it appears that CLC domains can clearly affect the mode and timing of chain release, they are not necessary as indicated by the above experiments where a functional NR PKS can be made lacking a CLC. It is not clear which component of the NR PKS is involved in chain release when the CLC is absent; however, the situation closely mirrors the situation in the bacterial type II PKS where there is no obvious mechanism to release the chain from the ACP when it has reached its final length.

A related process of chain shortening has been observed during the biosynthesis of the pentaketide THN (**6**) in *Aspergillus fumigatus*. In this organism the NR PKS gene *alb1* encodes a PKS (alb1p), which is closely related to the heptaketide synthase WAS. When *alb1* was expressed in *A. oryzae* the heptaketide YWA 1 (**7**) was produced, together with a minor component that proved to be a dehydration product of **7**.[57] This result was initially puzzling because *alb1* had previously been shown to be involved in the biosynthesis of THN (**6**).

The conundrum was resolved by the analysis of other genes in the *alb1* cluster in *A. fumigatus*. Knockout of the gene *ayg1* (encoding the protein Ayg1p) led to loss of the production of THN (**6**) and its derivatives, but production of YWA 1 (**7**) was observed. It was therefore concluded that Ayg1p must convert **7** to **6**, presumably by a retro-Claisen mechanism (**Scheme 12**(**b**)).[94] This was demonstrated by treating YWA 1 (**7**) with partially purified Ayg1p resulting in the formation of **6** *in vitro*. A similar mechanism has been shown to be functional in *W. dermatitidis* where acetyl-THN (**68**) is the intermediate used for the retro-Claisen reaction.[59]

1.09.3.3.6 Acyl carrier protein domains

Additional ACP domains are a common feature of the processing components of fungal NR PKS. For example, the *A. nidulans* WAS and NSAS and the *C. lagenarium* THNS all possess an additional ACP domain between the ACP of the chain extension component and the C-terminal CLC/TE domain. In the case of the NSAS involved in dothistromin biosynthesis (biosynthesis is thought to proceed through a norsolorinic acid intermediate) in *Dothistroma septosporum*, there are two additional ACP domains.[56] It is clear that these additional ACP domains are not required because the NSAS from *A. parasiticus* requires only a single ACP.

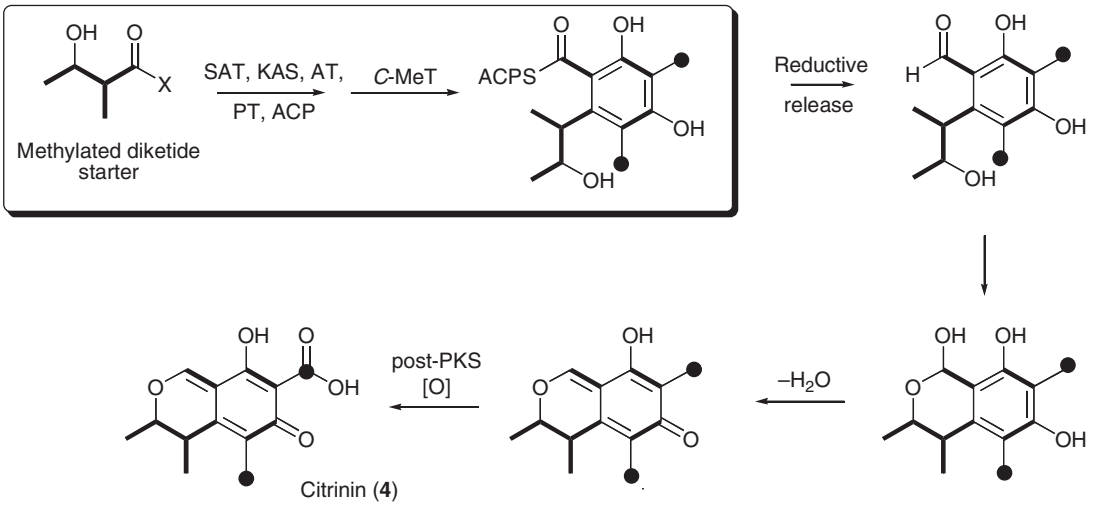
In the case of the WAS from *A. nidulans*, the role of the two ACP domains has been studied by Ebizuka and coworkers.[90] For an ACP to be active during polyketide biosynthesis it must have been appended with PP at its conserved serine. The *A. nidulans wA* gene encoding WAS was expressed in the heterologous fungal host *A. oryzae*. WAS was then modified to replace the two ACP serines with alanines. When either ACP was inactivated singly the synthase was still capable of producing the expected naphthopyrone (**7**), but inactivation of both ACP domains resulted in total loss of polyketide production. This experiment indicates that both ACP domains were capable of interacting with the chain extension and processing components.

Ebizuka and coworkers have also fused the ACP didomain and CLC from WAS onto the SAT, KS, AT, and PT domains of THNS. This chimeric PKS was active indicating that the ACPs are not 'matched' to the other components – they appear to function independently of any of the other program elements.[93]

1.09.3.3.7 C-MeT domains

Few NR PKS are known to possess C-methylation domains, although a number of known fungal NR polyketides are C-methylated such as 3-methylorcinaldehyde (**14**). A small group of NR PKS have been identified by Kroken *et al.*,[9] which feature a *C*-MeT domain located after the ACP. One of the first correlations between a gene sequence in this class and a compound comes from the case of citrinin (**4**) where the PKS involved in citrinin biosynthesis in *Monascus purpureus* has been described.[61]

Here the *C*-MeT domain must also be programed as it acts twice during polyketide biosynthesis when a probable C_5 starter unit is extended.[68] It is not yet clear whether the *C*-MeT domain acts during extension, after chain extension but before aromatization, or after aromatization. 1,3-Dihydroxyaromatics are known to tautomerize easily to keto forms[95] and it is conceivable that this could act as the nucleophile for the reaction with SAM (**Scheme 13**). This would maintain the pattern observed for the other processing component domains where these act after the extension components are finished. Other types of programing remain to be investigated, for example, 3,5-dimethylorsellinic acid methyl ester is a known fungal metabolite[96] and this could conceivably be synthesized by a very similar PKS with a differently programed *C*-MeT domain.



Scheme 13 Reactions proposed to be performed by CitS.

1.09.3.3.8 R domains

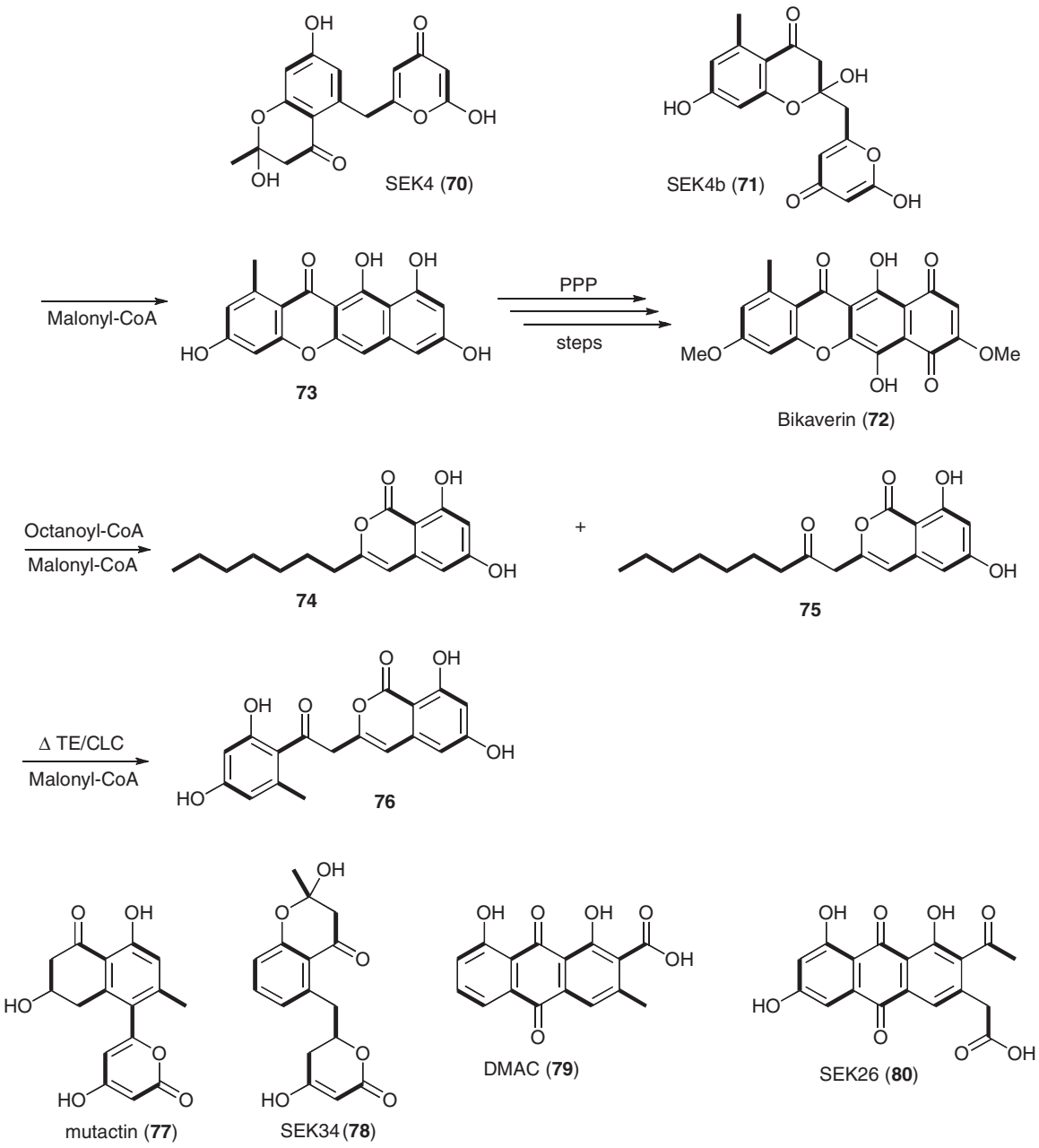
Once again, reductases are currently rare as part of the processing component of NR PKS. Although not described in the literature, sequence analysis of the *M. purpureus pksCT* sequence[61] shows that it possesses a C-terminal thiolester reductase domain.[97] A bona fide C-terminal reductase domain has been discovered in the case of 3-methylorcinaldehyde synthase (MOS) from *A. strictum*. Heterologous expression experiments using *A. oryzae* showed the reductive release mechanism to be operative in this case.[43]

Similar domains are known from NRPS systems where reductase domains are sometimes used as chain release mechanisms, releasing an aldehyde or primary alcohol. In the case of citrinin biosynthesis the reductive release mechanism makes good sense as this provides the product with C-1 at the correct oxidation state (**Scheme 13**).

1.09.3.3.9 Interaction of isolated NR PKS domains with components of bacterial type II PKS

As yet, extremely limited structural information exists regarding the 3D distribution of the active domains in fungal PKS enzymes. Structures of the mammalian FAS and of components of bacterial modular PKS serve as useful models. However, in terms of chemical catalysis and programing, the fungal NR PKS closely resemble the iterative type II PKS of bacteria – indeed, sequence comparisons with type II components have been used to delimit potential domain boundaries of the NR PKS.[78] The simplest hypothesis is that fungal NR PKS are simply 'joined-up' type II systems. As discussed above, early work by Cox and coworkers with the isolated AT and ACP domains of NSAS showed that the sequence similarities translate to biochemical activity. For example, NSAS *apo*-ACP interacts productively with the phosphopantetheinyl transferase (PPTase) from *S. coelicolor*[98] to give both NSAS *holo*-ACP and acylated ACP species. In turn, NSAS *holo*-ACP can interact with *S. coelicolor* MCAT[99] (a malonyl transferase) and become malonylated, and then interact with the $KAS_{\alpha\beta}$ components of the actinorhodin PKS to give SEK4 (**70**) and SEK4b (**71**) (see **Scheme 14**) indicating that the NSAS ACP provides no component of programing.[78]

The gene *pks4* from *G. fujikuroi* encodes bikaverin (**72**) synthase (BIKS) and is a typical NR PKS consisting of SAT, KS, AT, PT, ACP, and TE/CLC domains.[63] Tang and coworkers have expressed *pks4* in *E. coli* in active form – again indicating successful interaction with the bacterial PPTase.[100] The purified BIKS was then incubated with malonyl-CoA, and this yielded the nonaketide bikaverin precursor (**73**). Interestingly, efforts to determine the starter unit selectivity of the intact synthase showed that acetyl-CoA, butyryl-CoA, and hexanoyl-CoA were not accepted, but that octanoyl-CoA was an effective substrate, giving the 'octaketide' and 'nonaketide' benzopyrones (**74**) and (**75**), respectively. Townsend and coworkers, however, showed that there was an error in the *pks4* gene sequence and the former BIKS protein harbored an inactivated SAT domain where the active site Cys was inadvertently removed. *In vitro* biochemical analysis using the correct, purified



Scheme 14 Products of BIKS under various circumstances.

SAT domain from BIKS showed it was specific for acetyl-CoA and showed no activity toward malonyl-CoA.[84] Octanoyl-CoA is presumably preferred by the downstream KS domain when the SAT domain is inactivated *in vitro*. Remarkably, BIKS shows very similar behavior to the act PKS when presented with long unsaturated starter units, which also maintains its carbon count (16 carbons in the case of the octaketide actinorhodin, 18 carbons for (**72**) and (**75**)).[45]

The role of the BIKS TE/CLC domain was then investigated by deleting its active site serine – this yielded the benzopyran (**76**).[101] The same result was obtained from BIKS in which the entire TE/CLC domain had been removed. Complementation of either mutant with a stand-alone TE/CLC domain partially restored **73** biosynthesis, showing that the assembled (or partially assembled) poly-β-keto intermediate is accessible to

exogenous proteins. Tang and coworkers then added components of type II PKS, such as the act KR (+ NADPH), gris ARO/CYC (an aromatase cyclase), and oxyN (the oxytetracycline second ring cyclase), resulting in the production of the known type II compounds, for example, **77–80** (**Scheme 14**).

In a further set of remarkable experiments, Tang and coworkers have emphasized the similarities between the fungal type I PKS and their bacterial type II counterparts. A minimal fungal PKS was synthesized by removing the SAT, PT, and TE/CLC domains from BIKS, leaving a simple KS AT ACP of ~129 kDa.[102] This simple PKS is capable of synthesizing nonaketides and, again, the synthase can interact productively with various bacterial cyclases to produce new compounds *in vivo* and *in vitro*.

The results of these experiments clearly show that there is enough structural similarity between the fungal PKS and the type II bacterial PKS to allow the proteins to associate and interact catalytically. Importantly, the bacterial proteins are clearly able to affect the nascent program of the fungal PKS, and this also reveals something of their activity with their cognate synthase components. Overall, the experiments show simply how similar the two sets of proteins are and thus far support the simplistic assumption that the fungal type I proteins can be regarded as conjoined type II enzymes.

1.09.4 Mixed Polyketide/Nonribosomal Peptides

Fungi produce a wide range of bioactive compounds derived from polyketides fused to amino acids. Examples include fusarin C (**24**) and chaetoglobosin A (**29**) (**Table 4**). Fusarin C (**24**) consists of a tetramethylated heptaketide fused to homoserine[103] and is produced by strains of the plant pathogens *F. moniliforme* and *Fusarium venenatum*. Genomic DNA libraries from these organisms were used to isolate a gene cluster centered around a 12-kb ORF encoding an HR PKS fused to an NRPS module .[104] The PKS region is homologous to LNKS: KS, AT, and DH domains are followed by *C*-MeT, a defective ER, KR, and ACP domains. Like LNKS, the ACP is upstream of an NRPS condensation (C) domain, but in this case the NRPS module is complete, featuring downstream adenylation (A), thiolation (T), and C-terminal thiolester reductase (R) domains. These R domains are homologous to others known in nonribosomal peptide biosynthesis and in other PKS (Section 1.09.3.3.8).

Directed knockout of the PKS–NRPS gene proved it to be involved in the biosynthesis of **24** and it was thus named fusarin synthetase (FUSS). The dysfunctional ER domain and the fact that there appears to be no *lovC* homologue in the cluster, is consistent with the polyunsaturated nature of the polyketide moiety.

A highly homologous PKS–NRPS gene has been shown to be involved in the biosynthesis of equisetin (**80**) (**Figure 10**) in *Fusarium heterosporum*.[105] Equisetin synthase (EQS) possesses the same catalytic domains as FUSS, but examination of the structure of **80** indicates that the pyrollidinone carbon derived from the carboxylate of the amino acid (serine in this case) is not reduced during the release, indicating either a reoxidation mechanism, or the fact that the R domain does not reductively release in this case. Other fungal tetramic acids and related compounds have also been linked to PKS–NRPS. These include cyclopiazonic acid (**30**),[108,109] pseurotin A (**81**),[110] and the cytochalasins such as (**29**).[107]

PKS–NRPSs are also involved in the biosynthesis of as yet unknown compounds in the plant pathogen *Magnaporthe grisea*.[111,112] The *Ace1* PKS–NRPS gene forms part of the biosynthetic gene cluster, which is fleetingly expressed only during apressorium formation during the initial penetration event of pathogenesis.

Table 4 Examples of proven links between HR PKS–NRPS and known compounds

Organism	*Gene*	*Protein*	*Final product*	*Reference*
Fusarium moniliforme	*ORF3*	FUSS	Fusarin C (**24**)	104
Fusarium heterosporum	*eqiS*	EQS	Equisetin (**80**)	105
Beauveria bassiana	*tenS*	TENS	Tenellin (**82**)	28
Aspergillus nidulans	*apdA*	ASPS	Aspyridone A (**83**)	106
Penicillium expansum	*cheA*	CGS	Chaetoglobosin A (**29**)	107
Aspergillus oryzae	*pks-nrps*	CPAS	Cyclopiazonic acid (**30**)	108
Aspergillus flavus	*pks-nrps*	CPAS	Cyclopiazonic acid (**30**)	109
Aspergillus fumigatus	*psoA*	PSES	Pseurotin A (**81**)	110

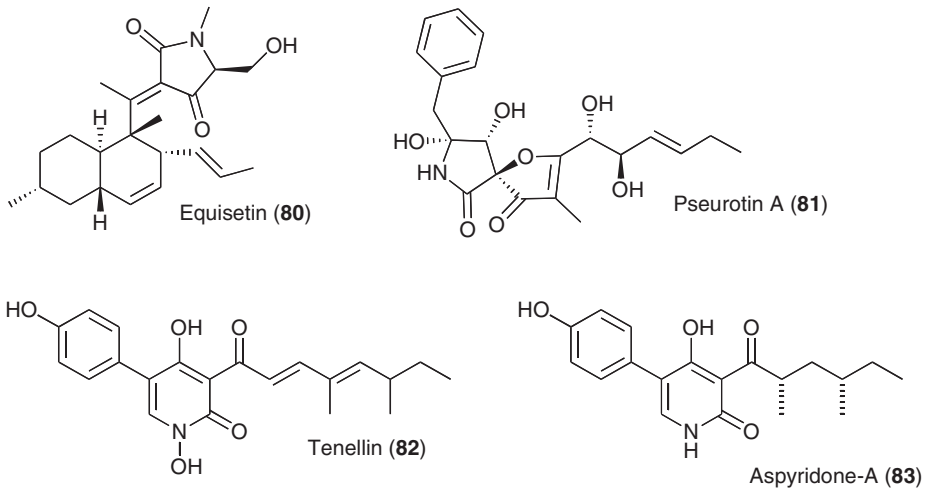


Figure 10 Products of mixed PKS–NRPS.

Other genes in the *Ace1* cluster encode cytochrome P-450 oxidases as well as lovC-type ER homologues. The compound produced by this cluster can be detected by specific genotypes of rice and, if detected, this produces a signaling cascade resulting in the effective mounting of resistance to further fungal penetration. Thus, when *Ace1* and its neighboring genes are expressed, *M. grisea* is rendered avirulent.[111]

PKS–NRPS proteins are also involved in the biosynthesis of fungal 2-pyridones. For example, the yellow pigment tenellin (**82**) from the insect pathogen *Beauveria bassiana* is produced by a PKS–NRPS, named TENS, homologous to FUSS.[113] Similarly, a PKS–NRPS hybrid is involved during the biosynthesis of the novel aspyridone A (**83**).[106] Like FUSS, the ER domain of TENS appears to be inactive. However, the TENS gene cluster also contains a *lovC* homologue, ORF3, and this provides the programed ER event required during the first round of polyketide biosynthesis. Cox and coworkers expressed *tenS* in *A. oryzae* and this experiment provided two key clues regarding the programing of these multifunctional synthases.[28] Three compounds were produced, namely, **84–86** (**Figure 11**). None of these could be true precursors of tenellin because the side chains had not been constructed correctly. However, coexpression of *tenS* with *tenC* led to the exclusive production of pretenellin A (**87**) showing that the presence of the ER guided polyketide programing in a similar way to that shown by lovC during lovastatin biosynthesis. However, tenC is clearly also involved in programing methylation fidelity, an activity not displayed by lovC.

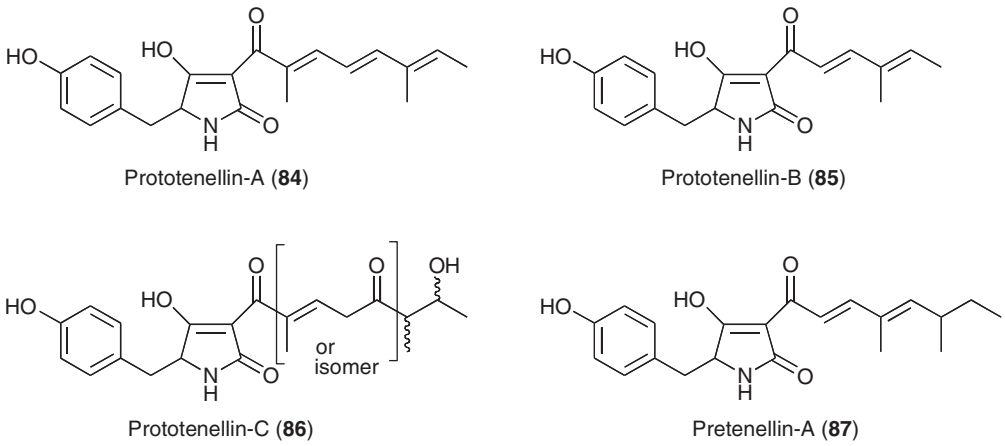


Figure 11 Compounds produced by TENS.

The tenellin heterologous expression experiments also showed that the product of *tenS* and *tenC* is the acyl tetramic acid (**87**), and not a product of the proposed reductive release and Knoevenagel cyclization. This experiment showed that the so-called R-domain must catalyze a Dieckmann-type cyclization (DKC). This observation was subsequently confirmed by Sims and Schmidt who showed that the expressed and purified DKC domain did indeed cyclize tetramic acid precursors rather than perform reductive chemistry.[114] Two other genes (*tenA* and *tenB*) in the TENS cluster encode cytochrome P-450 oxidases and have been shown to catalyze oxidative ring expansion to form the 2-pyridone, and N-hydroxylation, respectively (see Section 1.09.6).[115]

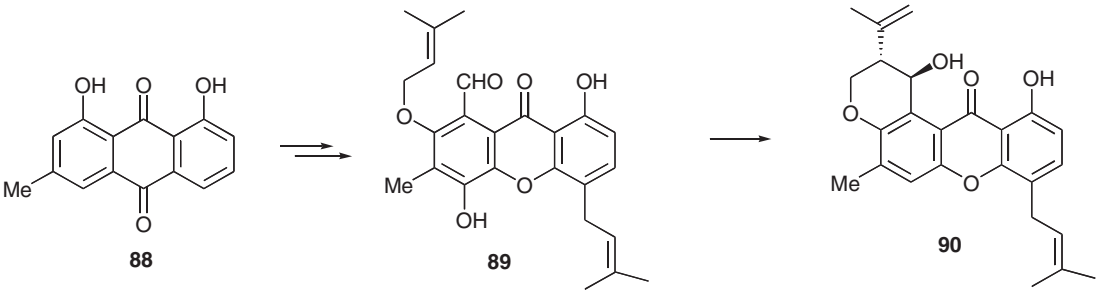
1.09.5 Meroterpenoids

Polyketide-derived moieties are very common in fungal metabolites of mixed biosynthetic origin, a particularly important group being the fungal meroterpenoids in which the polyketide-derived portion is attached to a terpenoid moiety – this can vary from a simple prenyl group derived from dimethylallyl diphosphate to geranylgeranyl diphosphate-derived diterpenoids. Shamixanthone (**90**), a metabolite of *Aspergillus variecolor*, contains both *C*- and *O*-prenyl residues attached to a xanthone skeleton, itself derived (**Scheme 15**) via oxidative ring cleavage of the central ring of the anthraquinone chrysophanol (**88**).[116] One prenyl residue remains intact, the other undergoes a Prins cyclization with an aldehyde (**89**) formed in the anthraquinone ring-cleavage process to form the dihydropyran ring. In the cometabolite, tajixanthone, the dimethylallyl group is epoxidized, a commonly observed modification in a wide range of plant- and fungal-derived prenyl phenolics.

The so-called reverse prenyl groups are also a common feature, for example, in the phenalenone deoxyherqueinone (**91**) (**Figure 12**) isolated from *Penicillium herquei*, the prenyl group is used to form the 2-methyl-3,3-dimethyl-dihydrofuran ring, by addition of a phenolic hydroxyl to a DMAPP-derived 1,1-dimethylprop-2-enyl group. Interestingly, the correct of three possible modes of folding of the NR heptaketide intermediate was established by incorporation of [1,2-$^{13}C_2$]-acetate in one of the earliest applications of stable isotope labeling to fungal polyketides.[117]

Viridicatumtoxin (**92**) isolated from *Penicillium expansum* contains a geranyl moiety attached in the form of a spiro arrangement to a tetracyclic nonaketide-derived moiety very similar to that found in the tetracycline antibiotics produced by *Streptomyces* species. Biosynthetic studies[118] are consistent with a malonate-derived starter unit as in the tetracyclines,[119] but incorporation of doubly labeled ^{13}C-acetate indicates a different direction of folding of the nonaketide precursor to that found in the tetracyclines.[120] Another interesting group of metabolites contain tetraketide-derived tropolone moieties fused to the sesquiterpene, humulene, a recent example being the anthelmintic noreupenifeldin (**93**), isolated from an unknown ascomycete.[121] In addition to the tropolone ring, noreupenifeldin contains a similarly fused six-membered phenolic ring. Baldwin has shown via model studies that these fusions may well proceed through a hetero-Diel–Alder cyclization of *ortho*-quinone methide intermediates with humulene.[122]

Xenovulene A (**15**), isolated from *A. strictum*, a potent inhibitor of benzodiazepine binding to the gamma-aminobutyric acid (GABA) receptor[123,124] contains a cyclopentenone fused to humulene. Minor cometabolites contain similarly fused phenolic and tropolone moieties[124] and it has been shown that the cyclopentone is formed by initial ring expansion onto the SAM-derived *C*-methyl substituent of a 3-methylorcinaldehyde-derived



Scheme 15 Biosynthesis of shamixanthone (**90**).

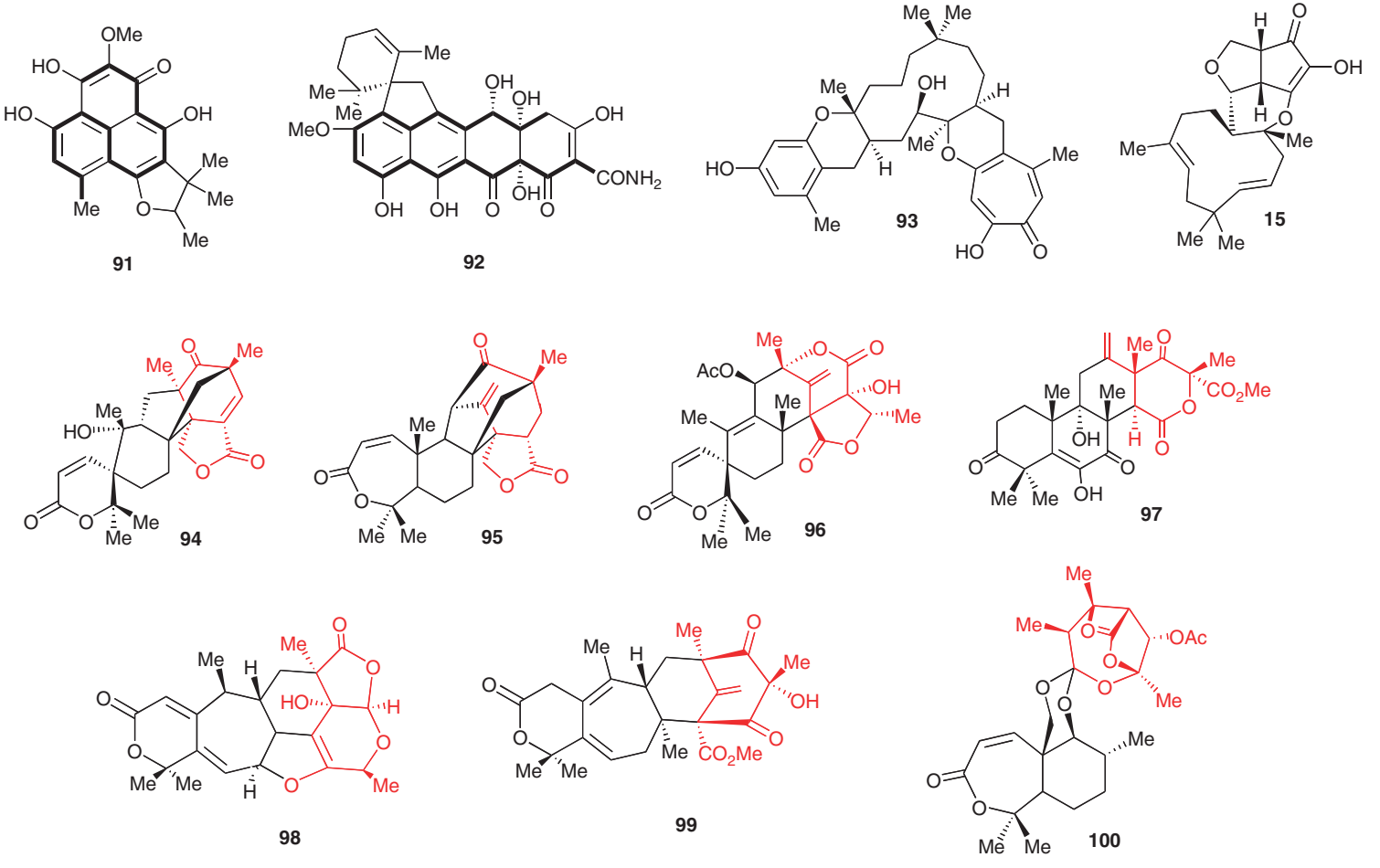
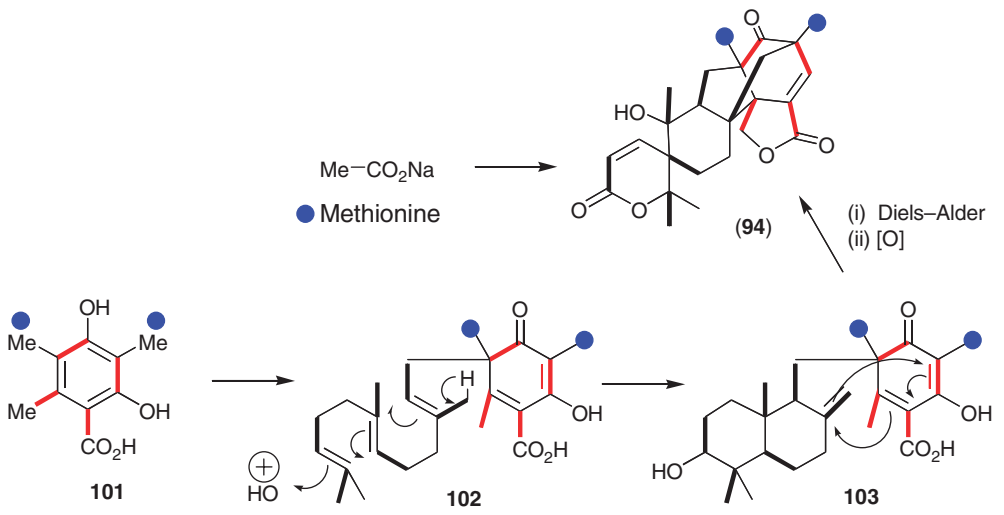


Figure 12 Meroterpenoid metabolites. Bold lines indicate linear polyketide folding; red indicates part derived from 3,5-dimethylorsellinate.



Scheme 16 Biosynthesis of andibenin B via 3,5-dimethylorsellinate and farnesyl diphosphate.

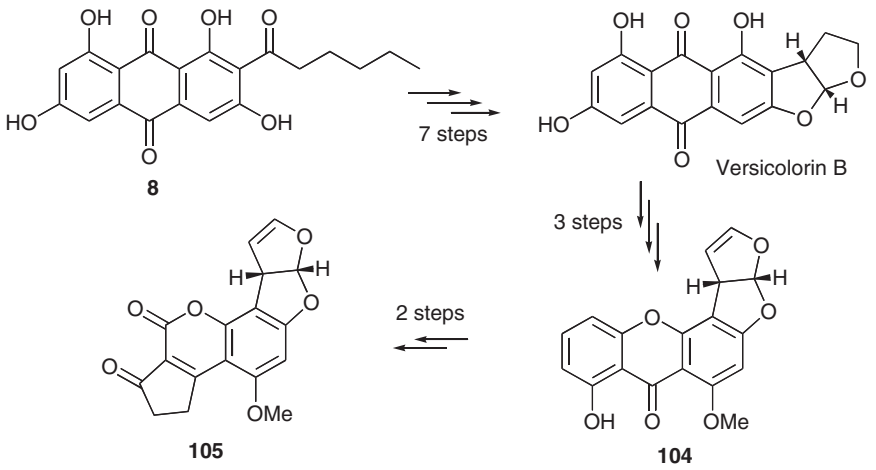
tetraketide moiety to form the tropolone, followed by successive ring contractions via the phenol to form the cyclopentenone.[125] The co-occurrence of these metabolites suggests that the initial fusion must occur before the subsequent ring expansion and contraction processes. As discussed above, an NR PKS gene encoding methylorcinaldehyde synthase (MOS) has been isolated from *A. strictum* (see Section 1.09.3.3.7).[43]

Andibenin B (**94**) is a meroterpenoid metabolite of *A. variecolor* formed via alkylation of 3,5-dimethylorsellinate (**101**) by farnesyl diphosphate to give **102**, followed by further fusion of the terpenoid and polyketide moieties via an intramolecular Diels–Alder of the key intermediate (**103**) to form the highly substituted bicyclo-2,2,2-octane ring system (**Scheme 16**).[126] Further extensive oxidative modifications of both the polyketide and terpenoid portions then occur.[127]

Subsequent to the original isolation of andbenin B, related metabolites anditomin (**95**) (*Aspergillus stellatus*),[128] austin (**96**) (*Aspergillus ustus*),[129] and terretonin (**97**) (*A. terreus*)[130] showing progressively more drastic rearrangements and oxidative cleavages of the 3,5-dimethylorsellinate and farnesyl-derived portions were reported. The involvement of 3,5-dimethylorsellinate was established by its synthesis with a variety of ^{13}C, ^{2}H, and ^{18}O labels and incorporation of the variously labeled precursors into the metabolites.[131] The pathway is now known to be widespread in a number of *Aspergillus* and *Penicillium* species giving rise to a wide range of increasingly complex structures all formed via a common precursor (**102**). Further representative examples include fumigatonin (**98**) from *A. fumigatus*,[132] berkleydione (**99**) from *Penicillium rubrum*,[133] and paraherquonin (**100**) from *Penicillium paraherquei*.[133,134] The structural and biosynthetic relationships among this large group of complex metabolites has been discussed in a recent comprehensive review of fungal meroterpenoids.[135]

1.09.6 Post-PKS Reactions in Fungi

Once the PKS has finished its work, the newly synthesized polyketide is generally released from the synthase. For example, in the case of aflatoxin B_1 (**105**), norsolorinic acid (**8**) is the first product of the pathway.[66] The released compounds are then acted on by a series of other, usually discrete, enzymes. In fungi, the reactions are frequently oxidative and include epoxidation, hydroxylation, and dehydrogenation, but other common reactions also include methylation of heteroatoms, cyclization, and rearrangement reactions. In the case of aflatoxin and the related sterigmatocystin (**104**) biosynthesis, the entire biosynthetic gene clusters have been sequenced and annotated: at least 12 individual post-PKS reactions occur (**Scheme 17**).[136] The post-PKS enzymes of fungi are often more diverse and more frequently occurring than their bacterial counterparts, again leading to the observation of increased chemical diversity and complexity of the polyketides derived from fungi.

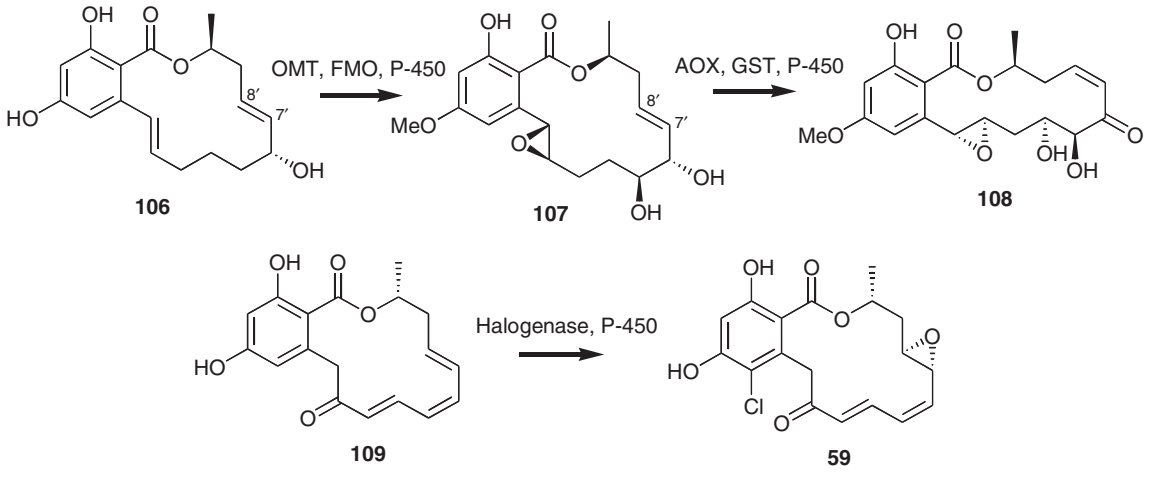


Scheme 17 Conversion of norsolorinic acid to aflatoxin B_1 via versicolorin B and sterigmatocystin.

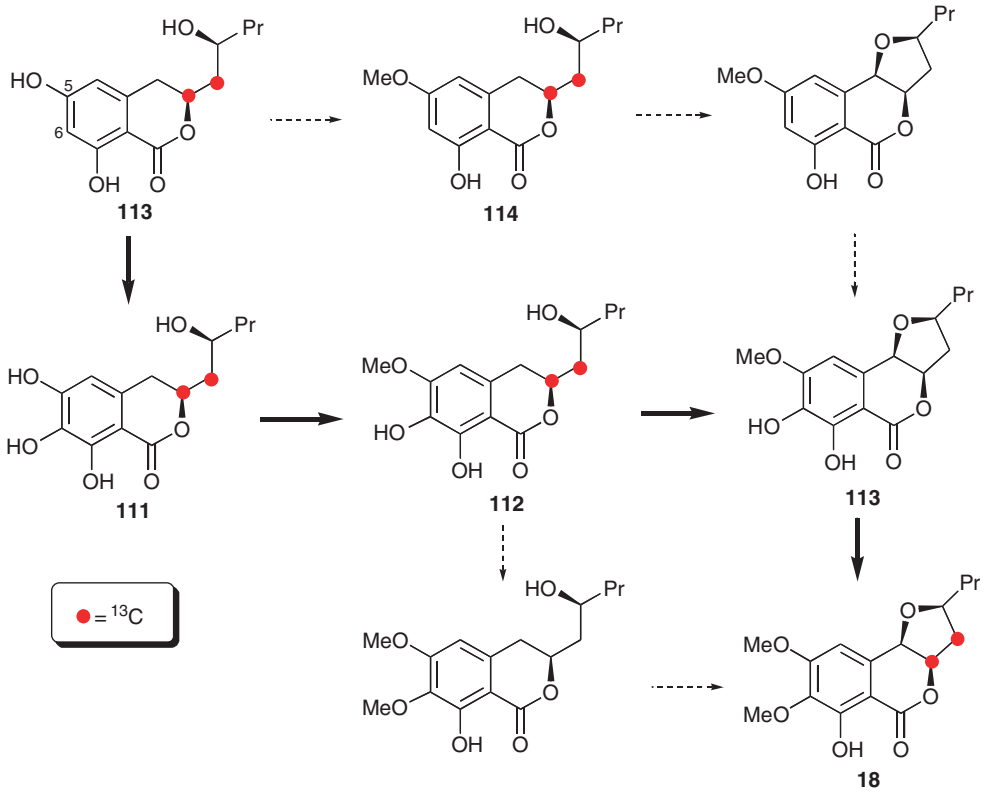
Gene clusters for the biosynthesis of the resorcylic lactones hypothemycin (**108**) from *Hypomyces subiculosis* and radicicol (**59**) from *Pochonia chlamydosporia* in addition to both HR and NR PKS genes as expected for these zearalenone analogues, contain genes for putative post-PKS reactions.[137] Thus, *O*-methyltransferase (OMT), flavin-dependent monooxygenase (FMO), cytochrome P-450, glutathione-*S*-transferase (GST), and alcohol oxidase (AOX-like) activities responsible for the necessary methylation, epoxidation, hydroxylation, E to Z double bond isomerization, and oxidation steps were identified. These were studied by a combination of gene knockout and *in vitro* heterologous expression studies indicated that the PKS product 7′,8′-dehydrozearalenol (**106**) was converted to hypothemycin (**108**) via the known compound aigialomycin C (**107**) as shown in **Scheme 18**. The radicicol gene cluster shows the presence of halogenase and P-450 homologues required to epoxidize and chlorinate the polyunsaturated PKS product (**109**) to give radicicol (**59**).

In the above cases, in common with the majority of fungal polyketides, the exact sequence of post-PKS steps in the biosynthetic pathway is uncertain. Indeed, in some cases, it may be that lack of substrate specificity allows a metabolic grid of alternate pathways to operate. This was thought to be the case for monocerin (**18**) biosynthesis as indicated in **Scheme 19**.

The synthesis of doubly ^{13}C-labeled dihydroisocoumarin (**110**) and its exceptionally high (~60%) intact incorporation into monocerin by cultures of *Dreschlera ravenelii* established its role as the first PKS-free



Scheme 18 Post-PKS modifications in the biosynthesis of hypothemycin and radicicol.



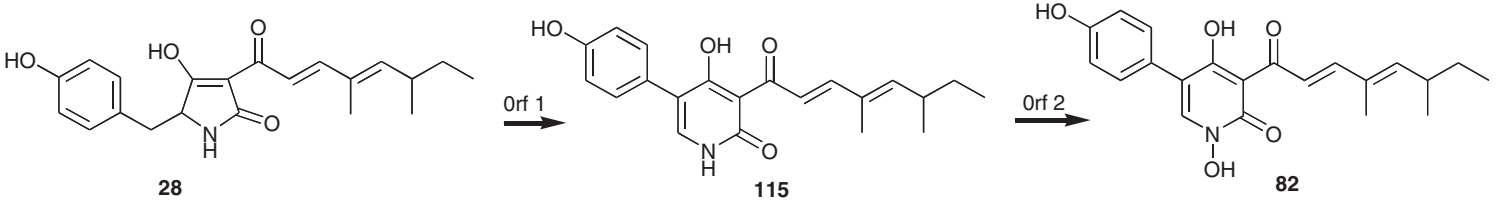
Scheme 19 Incorporation of ^{13}C-labeled dihydroisocoumarins into monocerin.

intermediate on the pathway.[138] It was proposed that the next steps would be 5-O-methylation, followed by hydroxylation and subsequent O-methylation at C-6. However, incorporation studies with doubly ^{13}C-labeled methyl ether (**111**) and triol (**112**) showed no detectable incorporation for the former but again very high incorporation of triol (**112**). Subsequent feeding of the doubly labeled triol methyl ether (**113**) again gave very high incorporation to establish the pathway via the known cometabolite (**114**) shown (bold arrows) in **Scheme 19**.[139]

As discussed above, heterologous coexpression studies in *A. oryzae* of the genes encoding the PKS–NRPS TENS and the LovC homologue, ORF3 from the tenellin gene cluster of *B. bassiana* results in the production of the tetramic acid pretenellin (**28**), the presumed precursor of the pyridone tenellin (**82**). The gene cluster contains two cytochrome P-450 encoding genes, *orf1* and *orf2*. By a combination of *in vivo* knockout and selective gene silencing and cell-free studies these have been shown to encode, respectively, the activities required for ring expansion of the tetramic acid to pyridone (**115**), pyridone N-hydroxylation to complete the biosynthesis of tenellin (**Scheme 20**).[115]

1.09.7 Conclusions

A combination of structure elucidation, classical biosynthetic labeling, molecular genetic, and enzymological studies is beginning to reveal how the huge variety of fungal polyketide structures are produced. In particular, the relatively recently established links between degree of reductive modifications (and methylation) and the structures of corresponding PKS genes and encoded synthases allow three main classes of polyketide structures and modes of assembly to be discerned. In comparison to bacterial polyketides, studies at the enzymological level are much less developed but a number of groups in Japan, North America, and Europe have begun to make significant advances in this. Likewise, very little structural information on fungal polyketides has been



Scheme 20 Conversion of pretenellin A to tenellin.

obtained but from sequence comparisons of individual catalytic domains at the amino acid level and modeling studies some structural insights can be made. Some progress is being made in rational manipulations at the genetic level to effect the production of modified or indeed novel structures, but 'combinatorial biosynthesis' of fungal polyketides remains in its relative infancy. All of these areas present scope for future studies but the key future challenge must be achieving an understanding of how programing of iterative assembly steps is controlled in these complex multidomain multifunctional proteins – the number of chain lengthening condensations, the degree of reductive modification and methylation that occurs in each elongation cycle, the selection of starter groups, and chain release and cyclization mechanisms. This is arguably the holy grail – the largest and the most ambitious challenge remaining in the whole field of secondary metabolism. It is a major intellectual challenge and addresses some of the most fundamental problems in enzymology: understanding how such complex enzyme systems actually work – how they recognize and transform their ever changing substrates. It is also a challenge with great potential for exploitation – how to use and manipulate these 'molecular factories' for the design and efficient production of new chemical entities and new bioactive molecules, building on the already exquisite structural variety displayed by fungal polyketides. This will require determining the function of individual catalytic domains as well as entire synthases, and the individual and cooperative roles of the domains in the overall process catalyzed by the PKS. The ultimate goal will be to relate the genetic sequence and the structure of these iterative PKSs to the chemistry that they control and to allow accurate prediction of the structure of the final metabolite. This will become of increasing necessity as the wealth of information on PKS structure continues to burgeon from genome sequencing efforts. The role of the PKS in relation to other genes in the biosynthetic gene clusters also needs to be determined to allow expression and modification of entire gene clusters to achieve full predictive control of metabolite production, particularly in the case of heterologous expression of entire cryptic pathways identified in increasing abundance in genomic sequences. Realization of this goal will require a synergistic combination of genetic engineering, multidomain protein expression and structural biology, study of intra- and interprotein–protein and substrate–protein interactions, individual and collective enzyme mechanisms, natural product isolation, synthesis and structure elucidation, and detailed bioinformatics analysis to relate the extremely subtle differences in gene sequence to overall function and final metabolite structure.

References

1. D. L. Hawksworth, *Mycol. Res.* **2001**, *105*, 1422–1432.
2. A. J. Birch; R. A. Massy-Westropp; C. J. Moye, *Aust. J. Chem.* **1955**, *8*, 539–544.
3. E. Stocker-Wörgötter, *Nat. Prod. Rep.* **2008**, *25*, 188–190.
4. J. Staunton; K. J. Weissman, *Nat. Prod. Rep.* **2001**, *18*, 380–416.
5. T. J. Simpson, *Chem. Ind.* **1995**, *11*, 407–411.
6. B. Wilkinson; G. Foster; B. A. M. Rudd; N. L. Taylor; A. P. Blackaby; P. J. Sidebottom; D. J. Cooper; M. J. Dawson; A. D. Buss; S. Gaisser; I. U. Bohm; C. J. Rowe; J. Cortes; P. F. Leadlay; J. Staunton, *Chem. Biol.* **2000**, *7*, 111–117.
7. L. E. H. Bingle; T. J. Simpson; C. M. Lazarus, *Fungal Genet. Biol.* **1999**, *26*, 209–223.
8. T. P. Nicholson; B. A. M. Rudd; M. Dawson; C. M. Lazarus; T. J. Simpson; R. J. Cox, *Chem. Biol.* **2001**, *8*, 157–178.
9. S. Kroken; N. L. Glass; J. W. Taylor; O. C. Yoder; B. G. Turgeon, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15670–15675.
10. see http://fungalgenomes.org/wiki/Fungal_Genome_Links
11. H. J. Pel; J. H. de Winde; D. B. Archer; P. S. Dyer; G. Hofmann; P. J. Schaap; G. Turner; R. P. de Vries; R. Albang; K. Albermann; M. R. Andersen; J. D. Bendtsen; J. A. Benen; M. van den Berg; S. Breestraat; M. X. Caddick; R. Contreras; M. Cornell; P. M. Coutinho; E. G. Danchin; A. J. Debets; P. Dekker; P. W. van Dijck; A. van Dijk; L. Dijkhuizen; A. J. Driessen; C. d'Enfert; S. Geysens; C. Goosen; G. S. Groot; P. W. de Groot; T. Guillemette; B. Henrissat; M. Herweijer; J. P. van den Hombergh; C. A. van den Hondel; R. T. van der Heijden; R. M. van der Kaaij; F. M. Klis; H. J. Kools; C. P. Kubicek; P. A. van Kuyk; J. Lauber; X. Lu; M. J. van der Maarel; R. Meulenberg; H. Menke; M. A. Mortimer; J. Nielsen; S. G. Oliver; M. Olsthoorn; K. Pal; N. N. van Peij; A. F. Ram; U. Rinas; J. A. Roubos; C. M. Sagt; M. Schmoll; J. Sun; D. Ussery; J. Varga; W. Vervecken; P. J. van de Vondervoort; H. Wedler; H. A. Wösten; A. P. Zeng; A. J. van Ooyen; J. Visser; H. Stam, *Nat. Biotechnol.* **2007**, *25*, 221–231.
12. D. I. Williams; R. H. Marten; I. Sarkany, *Lancet* **1958**, *2*, 1212–1213.
13. D. W. Bartlett; J. M. Clough; J. R. Godwin; A. A. Hall; M. Hamer; B. Parr-Dobrzanski, *Pest Manag. Sci.* **2002**, *58*, 649–662.
14. M. J. Sweeney; A. D. W. Dobson, *Int. J. Food Microbiol.* **1998**, *43*, 141–158.
15. M. J. Sweeney; A. D. W. Dobson, *FEMS Microbiol. Lett.* **1999**, *175*, 149–163.
16. J. Kennedy; K. Auclair; S. G. Kendrew; C. Park; J. C. Vederas; C. R. Hutchinson, *Science* **1999**, *284*, 1368–1372.
17. Y. Abe; T. Suzuki; C. Ono; K. Iwamoto; M. Hosobuchi; H. Yoshikawa, *Mol. Genet. Genomics* **2002**, *267*, 636–646.
18. I. Gaffoor; F. Trail, *Appl. Environ. Microbiol.* **2006**, *72*, 1793–1799.
19. R. J. Cox; F. Glod; D. Hurley; C. M. Lazarus; T. P. Nicholson; B. A. M. Rudd; T. J. Simpson; B. Wilkinson; Y. Zhang, *Chem. Commun.* **2004**, *20*, 2260–2261.

20. G. Wang; M. S. Rose; B. G. Turgeon; O. C. Yoder, *Plant Cell* **1996**, *8*, 2139–2150.
21. S. E. Baker; S. Kroken; P. Inderbitzin; T. Asvarak; B.-Y. Li; L. Shi; O. C. Yoder; B. G. Turgeon, *Mol. Plant Microbe Int.* **2006**, *19*, 139–149.
22. R. H. Proctor; A. E. Desjardins; R. D. Plattner; T. M. Hohn, *Fungal Genet. Biol.* **1999**, *27*, 100–112.
23. I. Fujii; N. Yoshida; S. Shimomaki; H. Oikawa; Y. Ebizuka, *Chem. Biol.* **2005**, *12*, 1301–1309.
24. K. Kasahara; I. Fujii; H. Oikawa; Y. Ebizuka, *ChemBioChem* **2006**, *7*, 920–924.
25. R. N. Moore; G. Bigam; J. K. Chan; A. M. Hogg; T. T. Nakashima; J. C. Vederas, *J. Am. Chem. Soc.* **1985**, *107*, 3694–3701.
26. L. Hendrickson; C. R. Davis; C. Roach; D. K. Nguyen; T. Aldrich; P. C. McAda; C. D. Reeves, *Chem. Biol.* **1999**, *6*, 429–439.
27. A. K. El-Sayed; J. Hothersall; S. M. Cooper; E. Stephens; T. J. Simpson; C. M. Thomas, *Chem. Biol.* **2003**, *10*, 419–430.
28. L. M. Halo; J. W. Marshall; A. A. Yakasai; Z. Song; C. P. Butts; M. P. Crump; M. Heneghan; A. M. Bailey; T. J. Simpson; C. M. Lazarus; R. J. Cox, *ChemBioChem* **2008**, *9*, 585–594.
29. X. Xie; K. Watanabe; W. A. Wojcicki; C. C. C. Wang; Y. Tang, *Chem. Biol.* **2006**, *13*, 1161–1169.
30. S. Ma; Y. Tang, *FEBS J.* **2007**, *274*, 2854–2864.
31. P. Beltran-Alvarez; R. J. Cox; J. Crosby; T. J. Simpson, *Biochemistry* **2007**, *46*, 14672–14681.
32. C. Bate; M. Salmona; L. Diomede; A. Williams, *J. Biol. Chem.* **2004**, *279*, 14983–14990.
33. R. A. E. Butchko; R. D. Plattner; R. H. Proctor, *J. Agric. Food Chem.* **2006**, *54*, 9398–9404.
34. X. Zhu; F. Yu; R. S. Bojja; K. Zaleta-Rivera; L. Du, *J. Ind. Microbiol. Biotechnol.* **2006**, *33*, 859–868.
35. X. Zhu; F. Yu; X.-C. Li; L. Du, *J. Am. Chem. Soc.* **2007**, *129*, 36–37.
36. J. Beck; S. Ripka; A. Siegner; E. Schiltz; E. Schweizer, *Eur. J. Biochem.* **1990**, *192*, 487–498.
37. J. Ahlert; E. Shepard; N. Lomovskaya; E. Zazopoulos; A. Staffa; B. O. Bachmann; K. Huang; L. Fonstein; A. Czisny; R. E. Whitwam; C. M. Farnet; J. S. Thorson, *Science* **2002**, *297*, 1173–1176.
38. I. Fujii; Y. Ono; H. Tada; K. Gomi; Y. Ebizuka; U. Sankawa, *Mol. Gen. Genet.* **1996**, *253*, 1–10.
39. P. Lu; A. Zhang; L. M. Dennis; A. M. Dahl-Roshak; Y.-Q. Xia; B. Arison; Z. An; J. S. Tkacz, *Mol. Genet. Genomics* **2005**, *273*, 207–216.
40. R. J. Light, *J. Biol. Chem.* **1967**, *242*, 1880–1886.
41. P. Dimroth; H. Walter; F. Lynen, *Eur. J. Biochem.* **1970**, *13*, 98–110.
42. J. B. Spencer; P. M. Jordan, *Biochem. J.* **1992**, *288*, 839–846.
43. A. M. Bailey; R. J. Cox; K. Harley; C. M. Lazarus; T. J. Simpson; E. Skellam, *Chem. Commun.* **2007**, *Oct 21*, 4053–4055.
44. C. J. Child; J. B. Spencer; P. Bhogal; P. M. Shoolingin-Jordan, *Biochemistry* **1996**, *35*, 12267–12274.
45. T. P. Nicholson; C. Winfield; J. Westcott; J. Crosby; T. J. Simpson; R. J. Cox, *Chem. Commun.* **2003**, *Mar 21*, 686–687.
46. I. D. G. Campuzano; P. M. Shoolingin-Jordan, *Biochem. Soc. Trans.* **1998**, *26*, S284–S1284.
47. D. J. Bedford; E. Schweizer; D. A. Hopwood; C. Khosla, *J. Bacteriol.* **1995**, *177*, 4544–4548.
48. J. T. Kealey; L. Liu; D. V. Santi; M. C. Betlach; P. J. Barr, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 505–509.
49. T. Moriguchi; Y. Ebizuka; I. Fujii, *ChemBioChem* **2006**, *7*, 1869–1874.
50. A. Witkowski; A. K. Joshi; S. Smith, *Biochemistry* **2004**, *43*, 10458–10466.
51. T. Moriguchi; Y. Ebizuka; I. Fujii, *ChemBioChem* **2008**, *9*, 1207–1212.
52. S. Wattanachaisaereekul; A. E. Lantz; M. L. Nielsen; J. Nielsen, *Metab. Eng.* **2008**, *10*, 246–254.
53. G. M. Gaucher; M. G. Shepherd, *Biochem. Biophys. Res. Commun.* **1968**, *32*, 664.
54. J.-H. Yu; T. J. Leonard, *J. Bacteriol.* **1995**, *177*, 4792–4800.
55. P. K. Chang; J. W. Cary; J. Yu; D. Bhatnagar; T. E. Cleveland, *Mol. Gen. Genet* **1995**, *248*, 270–277.
56. R. E. Bradshaw; H. P. Jin; B. S. Morgan; A. Schwelm; O. R. Teddy; C. A. Young; S. G. Zhang, *Mycopathologica* **2006**, *161*, 283–294.
57. A. Watanabe; I. Fujii; H.-F. Tsai; Y. C. Chang; K. J. Kwon-Chung; Y. Ebizuka, *FEMS Microbiol. Lett.* **2000**, *192*, 39–44.
58. Y. Takano; Y. Kubo; K. Shimizu; K. Mise; T. Okuno; I. Furusawa, *Mol. Gen. Genet.* **1995**, *249*, 162–167.
59. M. H. Wheeler; D. Abramczyk; L. S. Puckhaber; M. Naruse; Y. Ebizuka; I. Fujii; P. J. Szaniszlo, *Eukaryotic cell* **2008**, *7*, 1699–1711.
60. B. Feng; X. Wang; M. Hauser; S. Kaufmann; S. Jentsch; G. Haase; J. M. Becker; P. Szaniszlo, *Infect. Immun.* **2001**, *69*, 1781–1794.
61. T. Shimizu; H. Kinoshita; S. Ishihara; K. Sakai; S. Nagai; T. Nihira, *Appl. Environ. Microbiol.* **2005**, *71*, 3453–3457.
62. M. Choquer; K. L. Dekkers; H.-Q. Chen; L. Cao; P. P. Ueng; M. E. Daub,; K.-R. Chung, *Mol. Plant Microbe Interact.* **2005**, *18*, 468–476.
63. P. Linnemannstöns; J. Schulte; M. Prado; R. H. Proctor; J. Avalos; B. Tudzynski, *Fungal Genet. Biol.* **2002**, *37*, 134–148.
64. D. W. Udwary; M. Merski; C. A. Townsend, *J. Mol. Biol.* **2002**, *323*, 585–598.
65. J. M. Crawford; P. M. Thomas; J. R. Scheerer; A. L. Vagstad; N. L. Kelleher; C. A. Townsend, *Science* **2008**, *320*, 243–246.
66. R. E. Minto; C. A. Townsend, *Chem. Rev.* **1997**, *97*, 2537–2555.
67. D. S. J. McKeown; C. McNicholas; T. J. Simpson; N. J. Willett, *Chem. Commun.* **1996**, 301–302.
68. H. Hajjaj; A. Klaebe; M. O. Loret; G. Goma; P. J. Blanc; J. Francois, *Appl. Environ. Microbiol.* **1999**, *65*, 311–314.
69. K. Arai; B. J. Rawlings; Y. Yoshizawa; J. C. Vederas, *J. Am. Chem. Soc.* **1989**, *111*, 3391–3399.
70. F. E. Scott; T. J. Simpson; L. A. Trimble; J. C. Vederas, *J. Chem. Soc., Chem. Commun.* **1984**, 756–758.
71. S. V. Pathre; P. V. Khadikar; C. J. Mirocha, *Appl. Environ. Microbiol.* **1989**, *55*, 1955–1966.
72. Y. Yoshizawa; Z. Li; P. B. Reese; J. C. Vederas, *J. Am. Chem. Soc.* **1990**, *112*, 3212–3213.
73. Z. Li; F. M. Martin; J. C. Vederas, *J. Am. Chem. Soc.* **1992**, *114*, 1531–1533.
74. S. Wang; Y. Xu; E. A. Maine; E. M. K. Wijeratne; P. Espinosa-Artiles; A. A. L. Gunatilaka; I. Molnár, *Chem. Biol.* **2008**, *15*, 1328–1338.
75. H. Zhou; J. Zhan; K. Watanabe; X. Xie; Y. Tang, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6249–6254.
76. T. S. Hitchman; E. W. Schmidt; F. Trail; M. D. Rarick; J. E. Linz; C. A. Townsend, *Bioorg. Chem.* **2001**, *29*, 293–307.
77. C. M. H. Watanabe; C. A. Townsend, *Chem. Biol.* **2002**, *9*, 981–988.
78. Y. Ma; L. H. Smith; R. J. Cox; P. Beltran-Alvarez; C. J. Arthur; T. J. Simpson, *ChemBioChem* **2006**, *7*, 1951–1958.
79. J. M. Crawford; B. C. R. Dancy; E. A. Hill; D. Udwary; C. A. Townsend, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16728–16733.
80. J. M. Crawford; A. L. Vagstad; K. C. Ehrlich; C. A. Townsend, *Bioorg. Chem.* **2008**, *36*, 16–22.

81. Y. Liu; Z. Li; J. C. Vederas, *Tetrahedron* **1998**, *54*, 15937–15958.
82. I. Fujii; Y. Mori; A. Watanabe; Y. Kubo; G. Tsuji; Y. Ebizuka, *Biochemistry* **2000**, *39*, 8853–8858.
83. C. Bisang; P. F. Long; J. Cortes; J. Westcott; J. Crosby; A. L. Matharu; R. J. Cox; T. J. Simpson; J. Staunton; P. F. Leadlay, *Nature* **1999**, *401*, 502–505.
84. J. M. Crawford; A. L. Vagstad; K. P. Whitworth; K. C. Ehrlich; C. A. Townsend, *ChemBioChem* **2008**, *9*, 1019–1023.
85. A. Keatinge-Clay; D. A. Maltby; K. F. Medzihradszky; C. Khosla; R. M. Stroud, *Nat. Struct. Mol. Biol.* **2004**, *11*, 888–893.
86. M. E. Mayorga; W. E. Timberlake, *Genetics* **1990**, *126*, 73–79.
87. M. E. Mayorga; W. E. Timberlake, *Mol. Gen. Genet.* **1992**, *235*, 205–212.
88. A. Watanabe; Y. Ono; I. Fujii; U. Sankawa; M. E. Mayorga; W. E. Timberlake; Y. Ebizuka, *Tetrahedron Lett.* **1998**, *39*, 7733–7736.
89. A. Watanabe; I. Fujii; U. Sankawa; M. E. Mayorga; W. E. Timberlake; Y. Ebizuka, *Tetrahedron Lett.* **1999**, *40*, 91–94.
90. I. Fujii; A. Watanabe; U. Sankawa; Y. Ebizuka, *Chem. Biol.* **2001**, *8*, 189–197.
91. I. Fujii; Y. Mori; A. Watanabe; Y. Kubo; G. Tsuji; Y. Ebizuka, *Biosci. Biotechnol. Biochem.* **1999**, *63*, 1445–1452.
92. A. Watanabe; Y. Ebizuka, *Chem. Biol.* **2004**, *11*, 1101–1106.
93. A. Watanabe; Y. Ebizuka, *Tetrahedron Lett.* **2002**, *43*, 843–846.
94. H.-F. Tsai; I. Fujii; A. Watanabe; M. H. Wheeler; Y. C. Chang; Y. Yasuoka; Y. Ebizuka; K. J. Kwon-Chung, *J. Biol. Chem.* **2001**, *276*, 29292–29298.
95. T. J. Simpson; M. K. B. Weerasooriya, *J. Chem. Soc., Perkin Trans. 1* **2000**, 2771–2775.
96. A. G. Soman; J. B. Gloer; D. T. Wicklow, *J. Nat. Prod.* **1999**, *62*, 386–388.
97. R. J. Cox, unpublished results.
98. R. J. Cox; J. Crosby; O. Daltrop; F. Glod; M. E. Jarzabek; T. P. Nicholson; M. Reed; T. J. Simpson; L. H. Smith; F. Soulas; A. E. Szafranska; J. Westcott, *J. Chem. Soc., Perkin Trans. 1* **2002**, 1644–1649.
99. A. E. Szafranska; T. S. Hitchman; R. J. Cox; J. Crosby; T. J. Simpson, *Biochemistry* **2002**, *41*, 1421–1427.
100. S. M. Ma; J. Zhan; K. Watanabe; X. Xie; W. Zhang; C. C. Wang; Y. Tang, *J. Am. Chem. Soc.* **2007**, *129*, 10642–10643.
101. S. M. Ma; J. Zhan; X. Xie; K. Watanabe; Y. Tang,; W. Zhang, *J. Am. Chem. Soc.* **2008**, *130*, 38–39.
102. W. Zhang; Y. Li; Y. Tang, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20683–20688.
103. D. O. Rees; N. Bushby; R. J. Cox; J. R. Harding; T. J. Simpson; C. L. Willis, *ChemBioChem* **2007**, *8*, 46–50.
104. Z. S. Song; R. J. Cox; C. M. Lazarus; T. J. Simpson, *ChemBioChem* **2004**, *5*, 1196–1203.
105. J. W. Sims; J. P. Fillmore; D. D. Warner; E. W. Schmidt, *Chem. Commun.* **2005**, *Jan 14*, 186–188.
106. S. Bergmann; J. Schümann; K. Scherlach; C. Lange; A. A. Brakhage; C. Hertweck, *Nat. Chem. Biol.* **2007**, *3*, 213–217.
107. J. Schümann; C. Hertweck, *J. Am. Chem. Soc.* **2007**, *129*, 9564–9565.
108. M. Tokuoka; Y. Seshime; I. Fujii; K. Kitamoto; T. Takahashi; Y. Koyama, *Fungal Genet. Biol.* **2008**, *45*, 1608–1615.
109. P.-K. Chang; B. W. Horn; J. W. Dorner, *Fungal Genet. Biol.* **2009**, *46*, 176–182.
110. S. Maiya; A. Grundmann; X. Li; S.-M. Li; G. Turner, *ChemBioChem* **2007**, *8*, 1736–1743.
111. H. U. Bohnert; I. Fudal; W. Dioh; D. Tharreau; J. L. Notteghem; M. H. Lebrun, *Plant Cell* **2004**, *16*, 2499–2513.
112. J. Collemare; A. Billard; H. U. Böhnert; M.-H. LeBrun, *Mycol. Res.* **2008**, *112*, 207–215.
113. K. L. Eley; L. M. Halo; Z. Song; H. Powles; R. J. Cox; A. M. Bailey; C. M. Lazarus; T. J. Simpson, *ChemBioChem* **2007**, *8*, 289–297.
114. J. W. Sims; E. W. Schmidt, *J. Am. Chem. Soc.* **2008**, *130*, 11149–11155.
115. L. M. Halo; M. N. Heneghan; A. A. Yakasai; Z. Song; K. Williams; A. M. Bailey; R. J. Cox; C. M. Lazarus,; T. J. Simpson, *J. Am. Chem. Soc.* **2008**, *130*, 17988–17996.
116. S. A. Ahmed; E. Bardshiri; C. R. McIntyre; T. J. Simpson, *Aust. J. Chem.* **1992**, *45*, 249–274.
117. T. J. Simpson, *J. Chem. Soc., Chem. Commun.* **1975**, 258–260.
118. A. E. De Jesus; W. E. Hull; P. S. Steyn; F. R. van Heerden; R. Vleggar, *J. Chem. Soc., Chem. Commun.* **1982**, 902–904.
119. R. Thomas; D. J. Williams, *J. Chem. Soc., Chem. Commun.* **1983**, 677–679.
120. R. Thomas; D. J. Williams, *J. Chem. Soc., Chem. Commun.* **1983**, 128–130.
121. S. Ayers; D. L. Zink; J. S. Powell; C. M. Brown; A. Grund; G. F. Bills; G. Platas; D. Thompson; S. B. Singh, *J. Nat. Prod.* **2008**, *71*, 457–459.
122. R. M. Adlington; J. E. Baldwin; G. J. Pritchard; A. J. Williams; D. J. Watkin, *Org. Lett.* **1999**, *1*, 1937–1939.
123. P. Thomas; H. Sundarum; B. J. Krishek; P. Chazot; X. Xie; P. Bevan; S. J. Brocchini; C. J. Latham; P. Charlton; M. Moore; S. J. Lewis; D. M. T. Horton; F. A. Stephenson; T. G. Smart, *J. Pharmacol. Exp. Ther.* **1997**, *282*, 513–520.
124. A. M. Ainsworth; M. I. Chicarelli-Robinson; B. R. Copp; U. Faith; P. J. Hylands; J. A. Holloway; M. Latif; G. B. O'Breirne; N. Porter; D. V. Renno; M. Richards; N. Robinson, *J. Antibiot.* **1995**, *48*, 568–573.
125. M. E. Raggatt; T. J. Simpson; M. I. Chicarelli-Robinson, *Chem. Commun.* **1997**, 2245–2246.
126. J. S. E. Holker; T. J. Simpson, *J. Chem. Soc., Chem. Commun.* **1978**, 626–627.
127. T. J. Simpson; S. A. Ahmed; C. R. McIntyre; F. E. Scott; I. H. Sadler, *Tetrahedron* **1997**, *53*, 4013–4034.
128. T. J. Simpson, *Tetrahedron Lett.* **1981**, *22*, 3785–3788.
129. K. K. Chexal; J. P. Springer; J. Clardy; R. J. Cole; J. W. Kirksey; J. W. Dorner; H. G. Cutler; B. J. Strawter, *J. Am. Chem. Soc.* **1976**, *98*, 6748–6750.
130. J. P. Springer; J. W. Dorner; R. J. Cole; R. H. Cox, *J. Org. Chem.* **1979**, *44*, 4852–4854.
131. F. E. Scott; T. J. Simpson; L. A. Trimble; J. C. Verderas, *J. Chem. Soc., Chem. Commun.* **1986**, 214–215.
132. E. Okuyama; M. Yamazaki, *Tetrahedron Lett.* **1984**, *25*, 3233–3234.
133. D. B. Stierle; A. A. Stierle; J. D. Hobbs; J. Stokken; J. Clardy, *Org. Lett.* **2004**, *6*, 1049–1052.
134. E. Okuyama; M. Yamazaki, *Tetrahedron Lett.* **1983**, *24*, 3113–3114.
135. R. Geris; T. J. Simpson, *Nat. Prod. Rep.* **2009**, *26*, 1063–1094.
136. D. Hoffmeister; N. P. Keller, *Nat. Prod. Rep.* **2007**, *24*, 393–416.
137. C. D. Reeves; Z. Hu; R. Reid; J. T. Kealey, *Appl. Environ. Microbiol.* **2008**, *74*, 5121–5129.
138. L. C. Axford; T. J. Simpson; C. L. Willis, *Angew. Chem. Int. Ed.* **2004**, *43*, 727–730.
139. A. Kanari, Synthesis of Putative Biosynthetic Intermediates to Monocerin. Ph.D. Thesis, University of Bristol, Bristol, UK, 2009.

Biographical Sketches

Dr. Russell Cox is a Professor of Organic and Biological Chemistry in the School of Chemistry at the University of Bristol, UK. He was born in the New Forest in 1967. He studied for his first Degree in Chemistry at the University of Durham, where he stayed to study for a Ph.D. in Bioorganic Chemistry with Professor David O'Hagan. Postdoctoral positions with Professor John Vederas FRS at the University of Alberta, Canada, and Professor David Hopwood FRS at the John Innes Centre, UK, were followed by his first academic appointment at the University of Bristol in 1996. His main interests focus on the chemistry of iterative polyketide synthases from bacteria and fungi.

Thomas J. Simpson graduated from the University of Edinburgh in 1969, and received his Ph.D. from the University of Bristol in 1973 working with Jake MacMillan. After postdoctoral work in the University of Liverpool with Stan Holker and the Australian National University with Arthur Birch, he was appointed to a lectureship in chemistry, University of Edinburgh in 1978. He moved to professorships in organic chemistry in the University of Leicester in 1988 and Bristol University in 1990, where he holds the Alfred Capper Pass Chair of Chemistry. His research, which covers all aspects of the chemistry and biosynthesis of microbial natural products, has led to ~180 papers. He has received a number of awards including the Royal Society of Chemistry Corday Morgan Medal (1984), the Simonsen (2002), Tilden (2001), and Hugo Muller (2004) Lectureships, and the Microbial Chemistry Medal of the Kitatsato Institution in 2005. He was elected Fellow of the Royal Society in 2001 and Fellow of the Royal Society of Edinburgh in 2006.

1.10 Type I Modular PKS

Alison M. Hill, University of Exeter, Exeter, UK

James Staunton, University of Cambridge, Cambridge, UK

© 2010 Elsevier Ltd. All rights reserved.

1.10.1	**Introduction to Polyketides**	386
1.10.1.1	Development of the Biosynthetic Theory	386
1.10.2	**Fatty Acid Biosynthesis – Reactions and Enzymes**	388
1.10.2.1	Synthetic Operations	388
1.10.2.2	Enzymes of Fatty Acid Biosynthesis	389
1.10.3	**Polyketide Biosynthesis – Reactions and Enzymes**	392
1.10.3.1	Assembly of the Basic Carbon Skeleton	392
1.10.3.2	Biosynthesis of the Polyketide Chain of Erythromycin	392
1.10.3.3	Identification of the Megasynthases Involved in Erythromycin Biosynthesis	395
1.10.3.4	Generation of Truncated Versions of the DEBS	398
1.10.4	**Mechanism and Structural Specificity of Precursor Loading within the Modules of the DEBS**	399
1.10.4.1	Specificity of Transfer of Acyl Group Building Blocks from External CoA Thioesters onto the DEBS Assembly Lines	399
1.10.4.2	Kinetic Studies of the Structural Specificity of AT Domains	400
1.10.4.3	Mass Spectrometric Studies of the Structural Specificity of AT Domains	401
1.10.5	**Experimental Investigation of Stereospecificity of the Reactions of the DEBS Chain-Extension Cycles**	403
1.10.5.1	Celmer's Rules	403
1.10.5.2	Deleting Catalytic Activities	404
1.10.5.3	Determination of the Chirality of the Methylmalonate Precursors	405
1.10.5.4	Studies of the Stereochemistry of Reactions using Intact Modules and Isotopically Labeled Precursors	405
1.10.5.5	Studies of the Stereochemistry of Reactions using Reconstructed DEBS Modules	408
1.10.5.6	Conclusion	409
1.10.6	**Studies of the Quaternary Structure of the Type I FAS**	409
1.10.6.1	Early Investigations of the Type I FAS Structure	410
1.10.6.1.1	The homodimeric character and cross-linking studies	410
1.10.6.1.2	The Wakil head-to-tail model for the type I FAS	410
1.10.6.2	Recent Studies of the Type I FAS Structure	410
1.10.6.2.1	Mutant complementation studies, leading to the Smith structure for the FAS	411
1.10.6.2.2	Direct observations of the complete type I FAS structure by X-ray crystallography: The Ban structure for the animal FAS	412
1.10.6.3	Comparison of the Smith and Ban Proposals for the Type I FAS Structure	414
1.10.7	**Studies of the Structure of the DEBS Modules**	415
1.10.7.1	Isolation of the DEBS Multienzymes	415
1.10.7.2	Proteolysis Studies on the DEBS Multienzymes	415
1.10.7.3	The Homodimeric Character of the Multienzymes and Individual Domains	416
1.10.7.4	Mutant Complementation Studies	416
1.10.7.5	NMR Studies of Docking Domains	416
1.10.7.6	Characterization of a DEBS Didomain, KS–AT, by X-Ray Crystallography	418
1.10.8	**Current Proposals for the Topology of the DEBS Modules and Multienzymes**	418
1.10.8.1	Structures Based on X-ray Images	418
1.10.8.2	The Cambridge Topology for the PKS Module	419
1.10.9	**Other Polyketide Synthases**	423

1.10.9.1	Variation in the Packaging of Modules into Multienzymes	424
1.10.9.2	The Monensin PKS	425
1.10.9.2.1	Origin of the core structure of monensin	425
1.10.9.3	The Rapamycin PKS	427
1.10.9.4	The Rifamycin PKS	430
1.10.9.5	The Mupirocin PKS, an AT-less System with Special Mechanisms for Generating C-1 Branch Points	430
1.10.9.6	Methymycin and Pikromycin	434
1.10.9.7	Borrelidin	436
1.10.9.8	Conclusion	438
1.10.10	**Commercial Applications of Genetic Engineering of Modular Polyketide Synthases**	438
1.10.10.1	Development of New Versatile Super Hosts and Combinatorial Biosynthesis of Aromatic Compounds by Type II PKS Pathways	438
1.10.10.2	Strategies for Engineering the Erythromycin PKS	439
1.10.10.2.1	Creation of hybrid versions of the DEBS	439
1.10.10.2.2	Mutasynthesis experiments with a DEBS 1 containing an inactivated KS1	441
1.10.10.2.3	Genetic engineering of the chain-extension modules of the DEBS	441
1.10.10.3	Generation of Analogues of Rapamycin	442
1.10.10.4	Insertion of a Complete Module into the DEBS Assembly Line	443
1.10.10.5	Prospects for Future Commercial Exploitation	444
1.10.11	**Future Perspectives**	444
References		447

1.10.1 Introduction to Polyketides

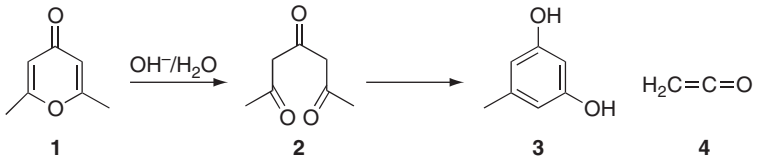
1.10.1.1 Development of the Biosynthetic Theory

The term polyketide (PK) was coined more than 100 years ago by the organic chemist J. Collie to describe a class of aromatic molecules produced synthetically while studying pyrones.[1] For example, in **Scheme 1**, on treatment with aqueous sodium hydroxide, the pyrone (**1**) was converted into the phenolic compound, orcinol (**3**). Collie rightly proposed that the triketone (**2**) was an intermediate. He also noted that the substitution pattern of hydroxyl groups on alternate carbons of the aromatic ring of orcinol was characteristic of many phenolic natural products.[2] He therefore proposed that such β-polyketones might be produced in living cells as biosynthetic precursors of phenolic natural products with a 1,3 pattern of hydroxyl groups. He also noted that the proposed polyketone intermediates could be hypothetically considered as polymers of ketene (**4**). He therefore coined the term PK for both the polyketones and the derived phenolic natural products.

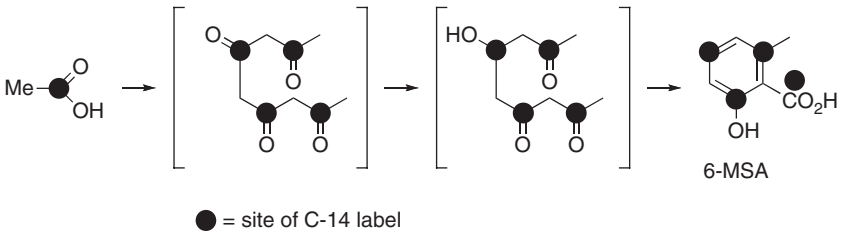
With these biosynthetic speculations, Collie was well in advance of his contemporaries. His ideas failed to achieve the impact they deserved. They and the term PK therefore became buried in the literature for more than 50 years.

In the mid-1950s, Birch independently had essentially the same idea.[3] By this time much more was known about the molecules and chemical processes occurring in living cells, and so Birch was able to extend the proposal with the suggestion that the proposed polyketones might arise in nature by condensation of acetate units. Unaware of the term PK, some natural product chemists coined the term acetogenin for these classes of natural products (See Chapters 2.04, 2.07, 2.17, 2.19, 3.03, 3.06, 3.08).

Working in the 1950s, Birch also had a key technical advantage over Collie in that radiotracer elements including carbon-14 had become commercially available. He therefore was able to test his acetate hypothesis by feeding labeled acetic acid to cells of a fungus that produce 6-methylsalicylic acid (6-MSA).[3] The resulting natural product molecules were isotopically labeled as shown in **Scheme 2** in the manner predicted by Birch's acetate hypothesis, thereby vindicating the biosynthetic proposal. Birch went on to justify his hypothesis with many other phenolic systems.



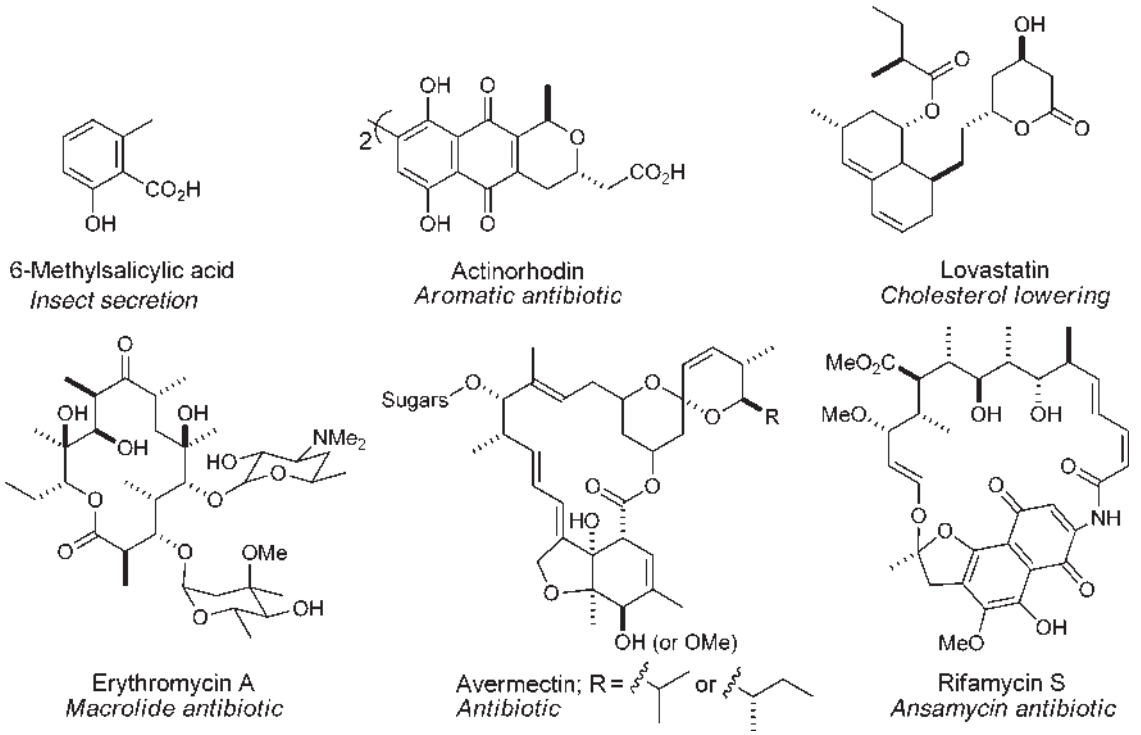
Scheme 1 Collie's chemical studies of polyketones and polyketides.



Scheme 2 Biosynthetic study of 6-MSA biosynthesis.[3]

Later, when Collie's pioneering contributions were rediscovered, the research community had to choose what to call this class of acetate-derived natural products. Both terms, acetogenin and PK, had their adherents for a while, but in the end PK gained universal acceptance.

Since the middle of the last century, there has been a vigorous search for new natural products and as a result, the number and diversity of PK structures has rapidly increased.[4–11] Some high-profile compounds are shown in **Scheme 3**. As can be seen, some are aromatic compounds whereas others are aliphatic. Collie would have been surprised to see the latter included under his chosen heading, but with the benefit of further knowledge accumulated over half a century, Birch was able to produce a unified hypothesis, which accounted for the genesis of both types of PK. He based his ideas partly on mechanistic reasoning and partly on analogy



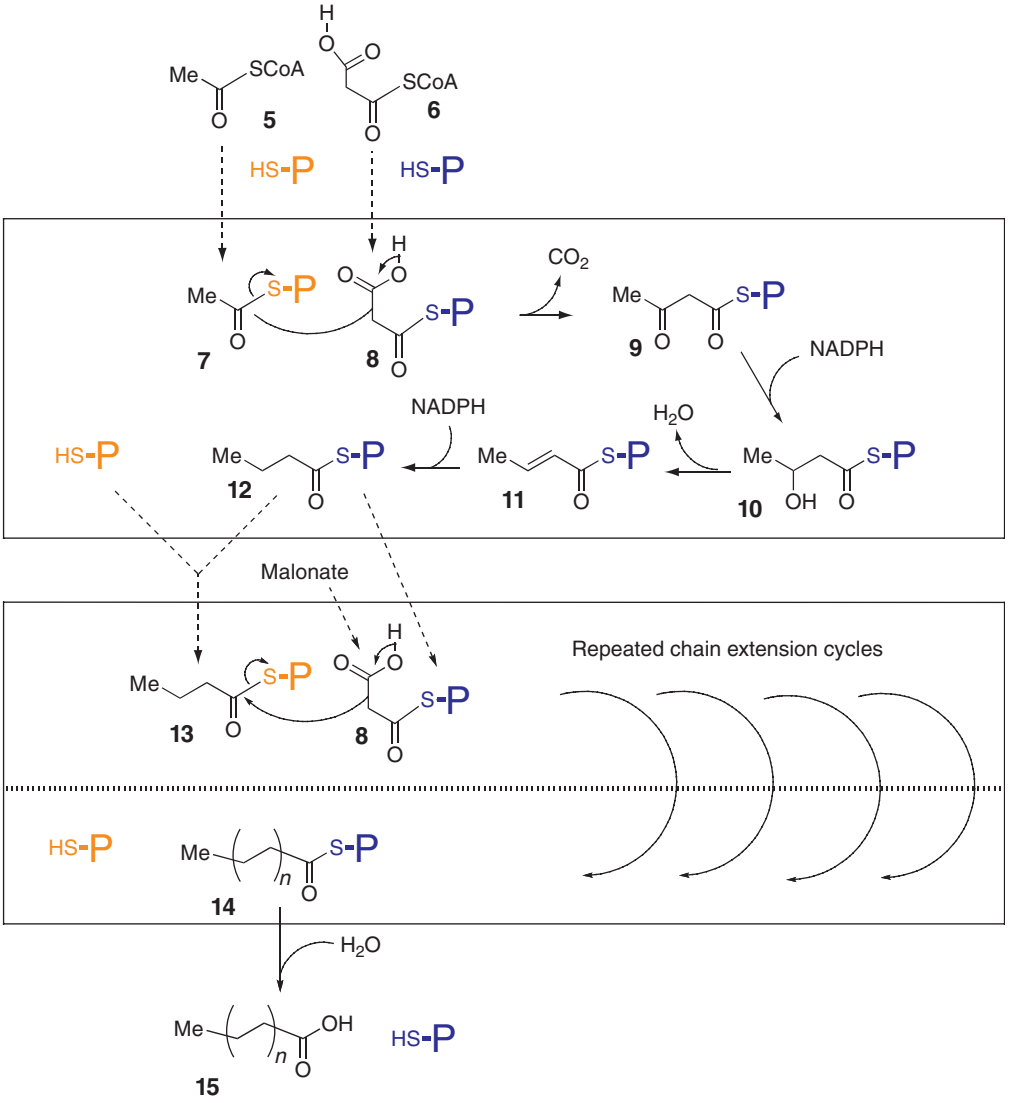
Scheme 3 Polyketide structures.

with new knowledge of fatty acid biosynthesis. The relationship between fatty acid biosynthesis and that of PKs continues to influence thinking in the PK field to this day, so it is appropriate to give a brief account of the biosynthetic reactions used in fatty acid biosynthesis.

1.10.2 Fatty Acid Biosynthesis – Reactions and Enzymes

1.10.2.1 Synthetic Operations

Saturated fatty acids such as stearic acid are produced by repeated condensation of units of acetate to give a chain of the requisite length. The detailed steps are shown in **Scheme 4**. On mechanistic grounds, it is not surprising that the carboxyl groups of reacting acyl units are derivatized as thioesters using active thiol groups at the active sites of the participating ketosynthase (KS) enzymes and acyl carrier proteins (ACPs). This derivatization activates the carbonyl toward nucleophilic attack, and so helps to stabilize enolate anion derivatives where they occur in the



Scheme 4 The interactions between proteins and intermediates during the successive cycles of chain extension in fatty acid biosynthesis.

sequence of reactions. The proteins will be introduced later. For this scheme, the letter P is used to designate the protein partner, and they are color coded to indicate that two different types of proteins are used.

Initially, the two thiol groups in the two proteins are primed with building blocks derived from the metabolic pools of precursors, acetyl CoA (**5**) and malonyl CoA (MalCoA) (**6**). An acetate unit is added to the first protein to serve as a starter acyl unit, and malonate is added to the second to serve as the chain-extender unit. These thiol transfers are essentially organizing operations, so they are designated by dotted arrows to distinguish them from the synthetic operations. The first step of the synthesis, in which the new carbon–carbon bound is formed, involves nucleophilic attack on the carbonyl group of the starter acetic acid (**7**). This type of reaction gives a β-ketoester, in this case acetoacetate (**9**), could be achieved through formation of the enolate of a second acetate unit, as in the classic Claisen condensation reaction widely used in organic synthesis. In fatty acid biosynthesis, however, an alternative strategy is employed in which the building block for chain-extension unit is malonate.[12–14] Carbon–carbon bond formation can then be driven by decarboxylation, either with simultaneous carbon–carbon bond formation, as indicated by the curly arrows on (**7**) and (**8**), or through an enolate intermediate. In terms of the isotopic-labeling outcome in Birch's classic pioneering labeling study,[3] the decarboxylation strategy for forming a carbon–carbon bond is equivalent to the direct Claisen condensation.

With the chain extended, the next steps involve multistep reduction of the new keto group in (**9**) leading to a methylene group in the intermediate (**12**). First, hydride transfer from a nicotinamide coenzyme, NADPH, reduces the keto group of (**9**) to form the hydroxyl group in (**10**). Dehydration follows, presumably through an enolate intermediate triggering an E1cb elimination mechanism. The resulting conjugated bond in (**11**) is then reduced to form the saturated acyl chain in (**12**).

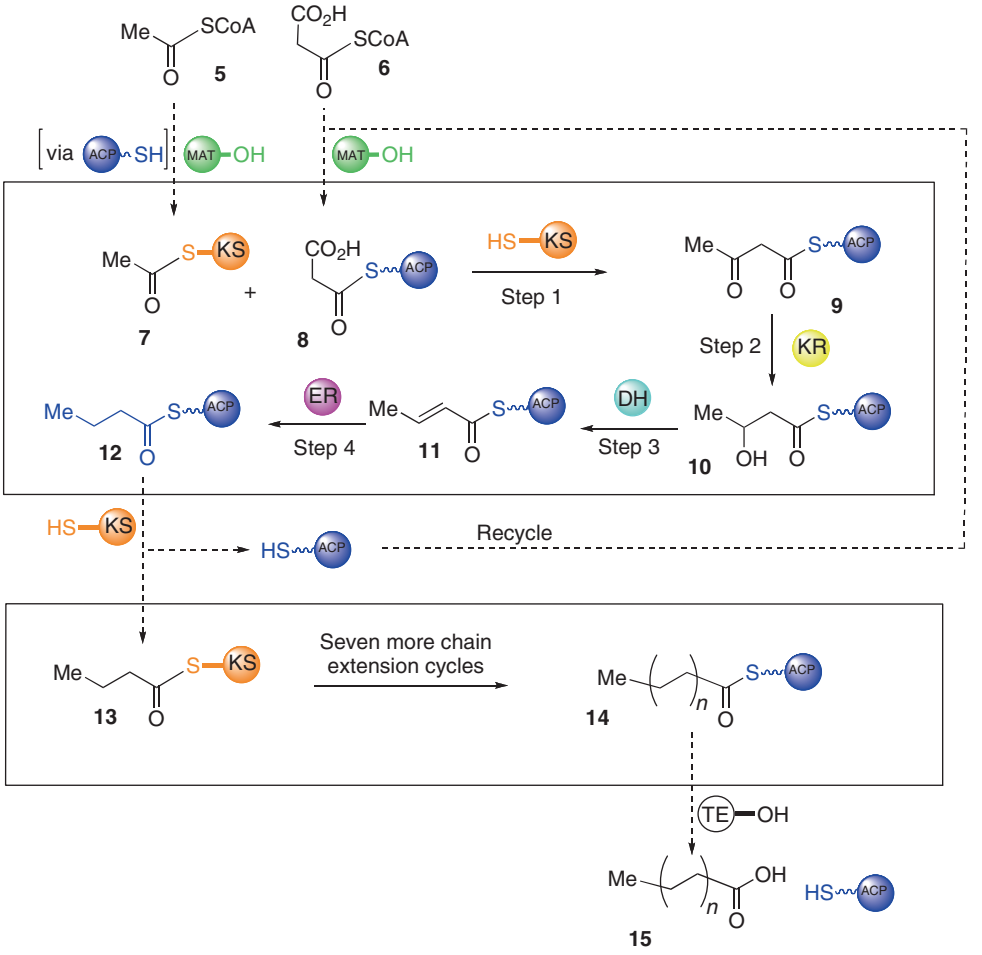
At this stage, the end of the first chain-extension cycle, the chain is extended by two methylene groups. The extended chain is then transferred from the blue protein to the orange protein, to give (**13**), which is now ready to commence the second chain-extension cycle. A new unit of methylmalonate is transferred to the now vacant thiol on the blue protein to regenerate intermediate (**8**). Repetition of the set of reactions can then proceed through further cycles of chain extension until the acyl chain has reached the required length. The final enzyme-bound intermediate (**14**) is then hydrolyzed from the orange thiol, rather than being transferred to the blue one, to yield the free fatty acid (**15**). In the case of stearic acid ($n = 8$), eight chain-extension cycles are required.

1.10.2.2 Enzymes of Fatty Acid Biosynthesis

The nature of the enzymes that catalyze the chemical reactions in **Scheme 4** was also first established for fatty acid systems.[15] **Scheme 5** shows the various types of enzymes making up a typical FAS system. The orange-colored protein in **Scheme 4** is called a ketosynthase, KS, because it acts as the catalyst for ketoester formation in step 1 of each chain-extension cycle. This class of enzyme is characterized by a cysteine thiol at its active site. The protein blue-colored, which carries the malonate reactant in carbon–carbon bond formation, is called an acyl carrier protein, ACP. As its name implies, this protein is noncatalytic but it serves an essential role in the chain-extension processes as a carrier of the acyl chain between the various catalytic domains. After step 1, the acyl chain leaves the active site of the KS but it remains attached to the ACP as a thioester for the next three steps, in which the keto group is reduced to the methylene level. The keto group reduction in step 2 is carried out by a nicotinamide coenzyme, NADPH, by a standard hydride transfer mechanism. The catalyst, called a ketoreductase (KR), is a member of a very wide family of enzymes known generally as alcohol dehydrogenases, which use NADPH to drive a reduction step. The enzyme that operates in step 3, called a dehydratase, DH, catalyzes a dehydration step. Finally, in step 4, an enzyme called an enoyl reductase, ER, catalyzes the reduction of the double bond, again using a hydride from the coenzyme NADPH to effect reduction.

This cycle of operations is then repeated, starting with transfer of the growing fatty acyl chain to the active thiol of the KS enzyme in a standard acyl transfer process. The enzymes of the fatty acid cycle are sufficiently nonspecific toward substrate so that the same set of enzymes can carry out repeated cycles until a chain of the required length has been reached. At that stage, a different enzyme called a thioesterase, TE, hydrolyzes the thioester link to liberate the chain as the free carboxylic acid.

Before closing this discussion of the FAS systems, it is worth summarizing what is known about the structure and function of the various proteins that were introduced in **Scheme 5**. Many of the enzymes, isolated from

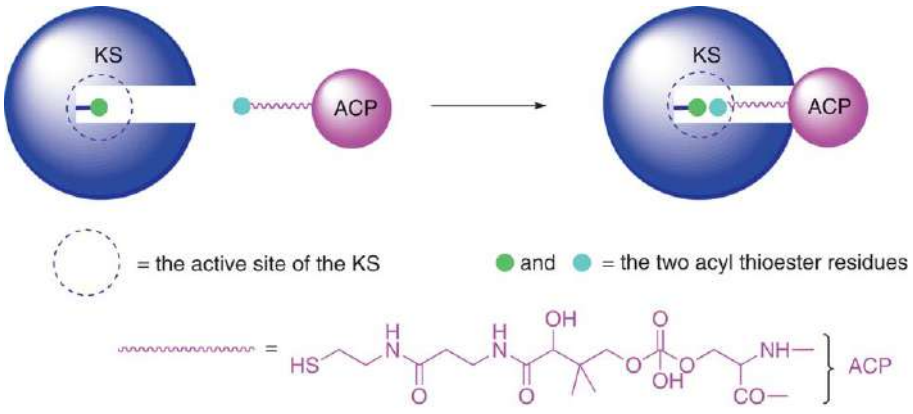


Scheme 5 Fatty acid biosynthesis.

various sources, have been studied by X-ray crystallography. Information concerning tertiary structures can be found in other chapters devoted to the molecular biology of FAS and PKS systems.

However, some brief comments on the general structure of the various domains will be helpful for later discussions. First, the KS proteins have the active thiol residue buried in the interior of the tertiary structure.[16] In contrast, the ACP has its active thiol as part of a long, flexible phosphopantetheine chain attached to the hydroxyl group of a serine residue on the exterior of the folded protein.[17] A phosphopantetheine chain also provides the active thiol of coenzyme A (CoA). It is thought that this long flexible chain can be inserted into a channel in the tertiary structure of the KS leading to the active site.[18] Any acyl residues on either one or both of the thiols is then subjected to the catalytic mechanisms of the active site of the KS. A cartoon showing this mode of interaction is presented in **Scheme 6**. Note that it assumes that the surfaces of the KS and ACP have to make intimate contact for this interaction to occur.

The acyl chain does not make any covalent interactions with the active sites of any of the next three enzymes in the cycle. When inserted into the active site by the phosphopantetheine chain, the acyl group is held by conventional noncovalent interactions. In the reduction by the KR in step 2, a hydride is transferred from a bound nicotinamide coenzyme, NADPH. The hydroxy ester then dissociates and is next bound to the DH, where it is dehydrated using the basic imidazole residue of a histidine residue as a base to catalyze the E1cb elimination mechanism. Finally, in the reduction by the ER, a bound NADPH transfers a hydride to the β-carbon of the enoyl system, followed by protonation to give the saturated acyl chain.



Scheme 6 Delivery of an intermediate attached to the thiol of an ACP into the active site of a KS domain.

All the enzyme active sites should have a broad specificity with respect to size of the substrate to cope with the steadily growing chain.[15] Eventually, however, the KS reaches the limit of its elasticity,[19,20] and at that point, further chain extension becomes very difficult. Fortunately, another enzyme, TE, comes into play as a catalyst for the cleavage of the thioester. Its substrate specificity is complementary to the KS in that the shape of the active site is optimized to accept a chain of the required length but to reject shorter chains.[21] At this chain length, therefore, the favored process switches from chain extension to chain release. The result of this complementarity of substrate specificity is an impressive degree of precision in the control over chain length achieved by a typical FAS (Chapters 1.08 and 8.07).

Scheme 5 also shows that a typical FAS depends on many acyl transfer reactions, indicated by dotted lines, in which an acyl residue is transferred from one thiol to another. Two very important examples are the loading of the starter acid unit and chain-extension unit from the CoA thioester pool onto active thiols of KS and ACP. The acyltransferase enzyme, MAT (malonyl acyl transferase), which catalyzes these steps, has an active serine hydroxyl at its active site. In the first step of the transfer, the acyl group is transferred from its thioester to form an oxygen ester on the AT. It is then transferred onto the destination thiol by a further ester transfer step. The TE enzyme employs a similar two-step mechanism in cleaving the completed chain, with an oxygen ester intermediate being formed on an active serine hydroxyl.

Finally, note the different symbols used to represent bonds to the active thiols and hydroxyl groups in the cartoon representations of the enzymes. A thick solid line is used for the KS, AT, and TE bonds, to symbolize the fact that the active residue is attached directly to the protein backbone of the structure. In the case of the ACP, where the active thiol is attached to the protein by a long, flexible chain, the link is symbolized by a wavy line to denote flexibility. A cartoon is presented in **Scheme 6** to show how an ACP might be visualized as it interacts with a KS in the condensation step. An insight into how the phosphopantetheine arm delivers the acyl chain to the KS can be gleaned from the X-ray crystal structure of yeast FAS stalled at the KS.[22] Conserved, charge-complementary surfaces allow the ACP to position itself near the KS deep catalytic cleft. The flexible phosphopantetheine arm, initially buried within the hydrophobic ACP core, is proposed to 'flip out' using a switchblade mechanism to enable the acyl chain to be threaded into the KS cleft.

Two very different degrees of organization have been discovered for the components of fatty acid synthases. In bacteria, such as *Escherichia coli*, the individual enzymes can be isolated as separate entities. In contrast, the FAS systems isolated from animals are large multienzymes, in which the individual catalytic activities (domains) are covalently linked through sections of relatively unfolded protein chain (linkers). The linkers can vary in length from tens of amino acids to hundreds. The nondissociable systems are called type I, and the systems with separable components are called type II. In recent reviews, the nondissociable multienzymes have been called megasynthases. A more detailed account of their structure is given later.

1.10.3 Polyketide Biosynthesis – Reactions and Enzymes

1.10.3.1 Assembly of the Basic Carbon Skeleton

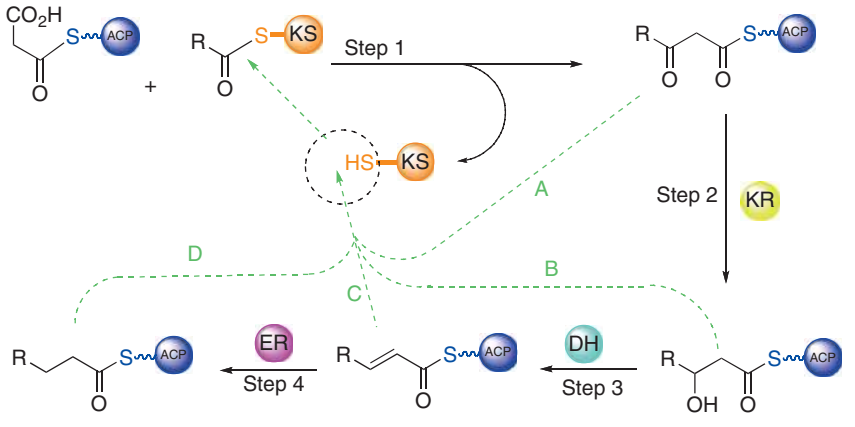
The very basic set of organic reactions used by the FAS may lack the exciting diversity and mechanistic thrill of some modern organic synthetic methodology, but the synthetic strategy is simple and direct, and achieves its synthetic goal in an elegant and potentially very versatile manner. This potential versatility can be appreciated by considering the basic assembly processes used in PK biosynthesis.

First, the poly-β-ketones envisaged in the Collie–Birch hypothesis for biosynthesis of phenolic PKs can be made by omitting all the steps of keto group reduction in a fatty acid synthetic cycle. This type of PK biosynthesis is the subject of Chapter 1.03 and so it will not be considered in detail here. **Scheme 7** shows the general manner by which a variation of the FAS biosynthetic cycle can serve to assemble the requisite poly-β-ketone chains: each new β-ketoacyl intermediate is passed back to the KS thus avoiding the reductive steps (the enzymes may be absent). Chains of differing length are possible depending on the number of cycles, leading ultimately to many different types of aromatic rings.

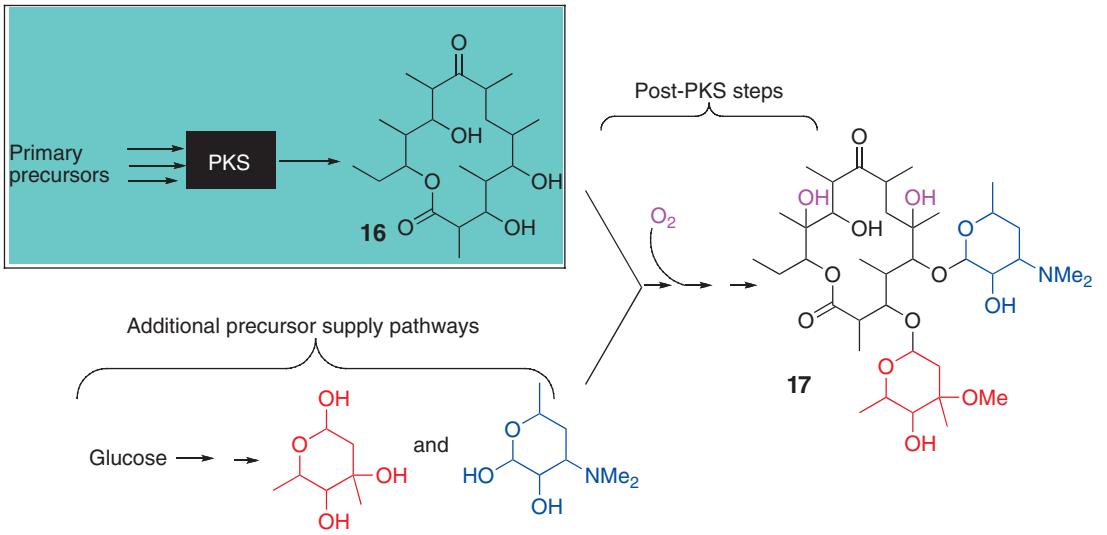
The compounds relevant to this chapter are the aliphatic systems such as erythromycin (ery). In such compounds, the process of keto group modification in individual cycles can be stopped at any stage, leading to the possibility of generating an enormous diversity of structure. The principle of this varied operation of the FAS cycle to produce PK structures is shown in **Scheme 7**, where dotted lines indicate all three possible shortcuts, A, B, C, which can limit the extent of keto group modification in individual cycles. The fourth pathway, D, is the normal path of fatty acid biosynthesis, which is also used. This setup provides an elegant and efficient strategy for achieving the biosynthesis of a complex PK chain by varying the operations of a standard fatty acid synthetic cycle. It also raises a very challenging question: how is the operation of such a synthase programmed to ensure that each new ketoacyl intermediate is modified to the appropriate extent in successive chain-extension cycles? It is this puzzling concept of programming that has inspired many leading research groups to investigate the area of aliphatic PK biosynthesis.

1.10.3.2 Biosynthesis of the Polyketide Chain of Erythromycin

As the best-studied example, the antibiotic erythromycin A (**17**) from *Saccharapolyspora erythraea* makes a logical choice to illustrate the principles of aliphatic PK biosynthesis. The first enzyme-free intermediate on this pathway is the macrolactone 6-deoxyerythronolide B (6-dEB) (**16**).[23] The free intermediate is converted into the final product erythromycin (**17**) A by a series of conventional elaboration reactions as indicated in **Scheme 8**.[24–26] The character of such post-PKS steps will be considered in a later section. The present focus is on the processes of PK chain assembly on the ery PKS, which for the moment is represented as a black box.



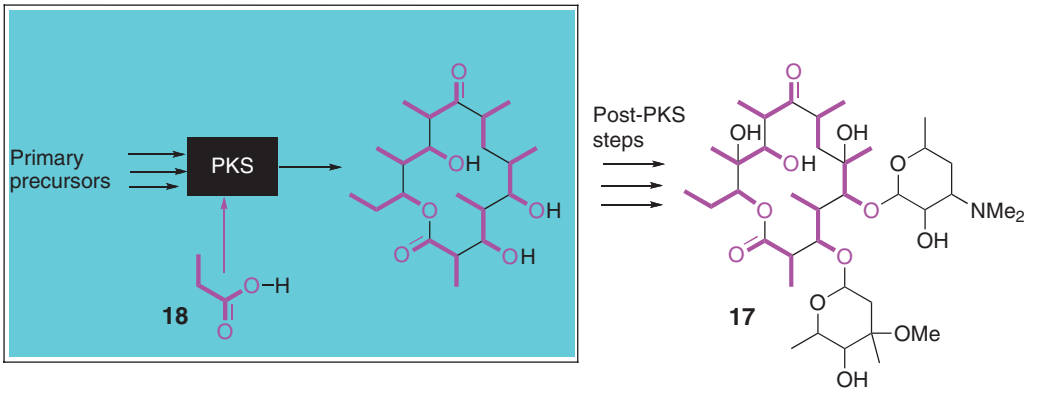
Scheme 7 The polyketide biosynthetic cycle showing the possible stages at which the cycle might be short circuited.



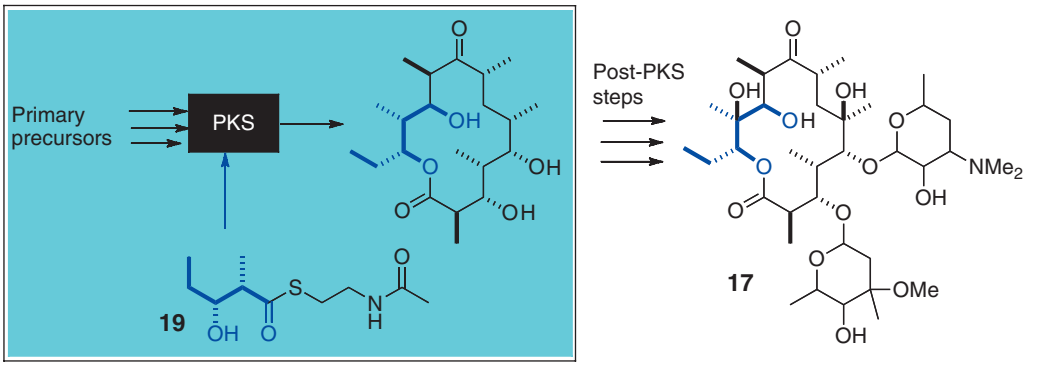
Scheme 8 Overview of the erythromycin biosynthetic pathway.

Two aspects of the PK chain of ery set it apart from the structure of simple fatty acid chains. First, it has a variety of functional groups in its individual chain-extension units, and so the chain assembly process needs to be programmed as discussed above. Second, in contrast to the fatty acids, the carbon backbone is not a straight chain of carbons, but has branching methyl groups strategically distributed at alternative sites along its length. Significantly, the methyl groups are attached to carbons that correspond to carbons in a simple fatty acid that derive from the methyl group of acetate.

In general, branching methyls can arise in a PK structure in two different ways. First, the methylene residue of a β-keto ester can be methylated using the methyl group derived from *S*-adenosyl methionine (SAM). Alternatively, the branching methyl pattern can arise by using propionate units instead of acetate units in the chain-elongation process. Confirmation that the latter strategy is used in ery biosynthesis derived from feeding various labeled forms of propionate.[27–29] The most revealing of such experiments was the feeding by the Cane group of (**18**), a version of propionic acid multiply labeled at strategic sites by carbon-13 and oxygen-18.[27] The resulting pattern of labeling in the core macrocyclic ring structure of ery is shown in **Scheme 9**. In addition to confirming earlier evidence that seven intact C_3 units of propionate are incorporated head-to-tail in the PK manner, the labeling pattern showed that key oxygen atoms also derive from propionate. It was assumed, of course, by analogy with the FAS systems, that the C_3 units introduced as extenders would be incorporated initially in the form of methylmalonate.



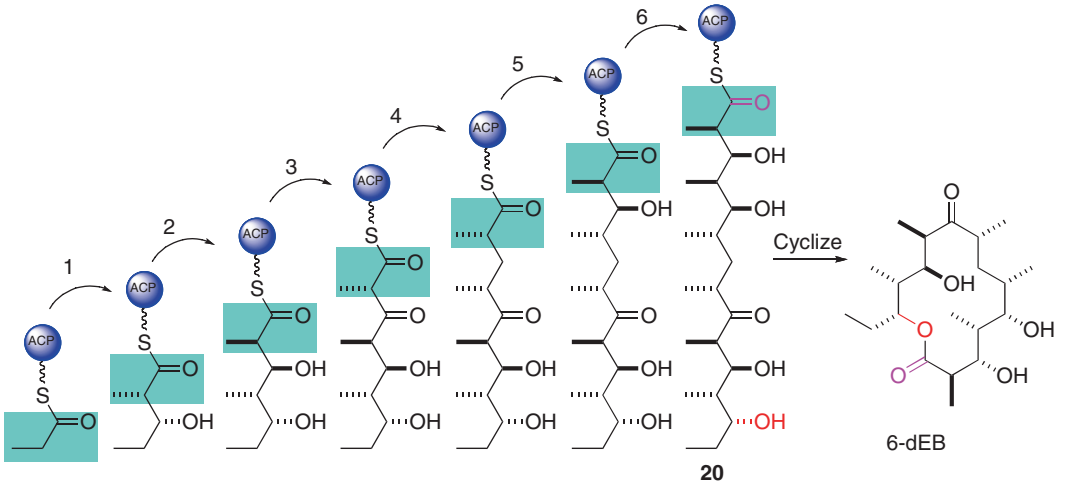
Scheme 9 Incorporation of propionate carbons and oxygens into erythromycin, based on isotopic-labeling studies.[27]



Scheme 10 Specific incorporation of a diketide unit into the core of erythromycin.[30]

Another milestone experiment from the Cane group was the feeding of the putative diketide intermediate resulting from the first cycle of chain extension.[30] This was derivatized as a thioester with *N*-acetylcysteamine (**19**), to aid acyl group transfer onto the appropriate thiol of the PKS (**Scheme 10**). The isotopic labels appeared at the expected sites in the macrolide ring without any significant randomization into other C_3 units. This seminal experiment confirmed that the chain-extension process follows the so-called processive strategy, under which each new keto group is modified to the extent found in 6-dEB, before commencement of the next cycle of chain extension.

It was now possible to speculate with confidence concerning the steps of assembly of the ery macrolactone core, based on retro-biosynthetic analysis. In the first chain-extension cycle, as shown in **Scheme 11**, a propionate unit as an ACP thioester is condensed with a unit of methylmalonate accompanied by decarboxylation, to form a keto ester. This is then reduced to a hydroxyester, but no further. The hydroxy ester starts the next cycle and the new keto group is again reduced as far as a hydroxyl group. In the next cycle, the new keto group is left unreduced, whereas in cycle 4 the new keto group is reduced all the way to a methylene. The final two cycles repeat the pattern of the first two, leaving again two new hydroxyl functions. In summary, the same set of reactions is potentially available at every stage, but the synthase somehow chooses how far to reduce each new keto group. This is the first source of diversity in these aliphatic PKs.



Intermediates formed in the six successive chain extension cycles;
each newly added residue is marked by a blue box

Scheme 11 Putative intermediates produced by processive chain assembly of 6-dEB. Intermediates formed in the six successive chain-extension cycles; each newly added residue is marked by a blue box.

The many chiral centers generated in the PK chain provide a second source of diversity. Both the methyl groups and the hydroxyl groups are capable of being either up or down as drawn, and all four arrangements are found in the final uncyclized heptaketide intermediate (**20**). The number of potential stereochemical variants that could be formed in principle by this synthetic strategy is 2^{10} (1024). That only one product (**20**) is formed in significant amounts is a convincing tribute to the control and specificity of the biosynthetic apparatus. The challenge in discovering, and possibly emulating, these synthetic operations lies not in the character of the reactions themselves, but in the processes by which the enzymes achieve such high and varied specificity in each successive extension cycle.

1.10.3.3 Identification of the Megasynthases Involved in Erythromycin Biosynthesis

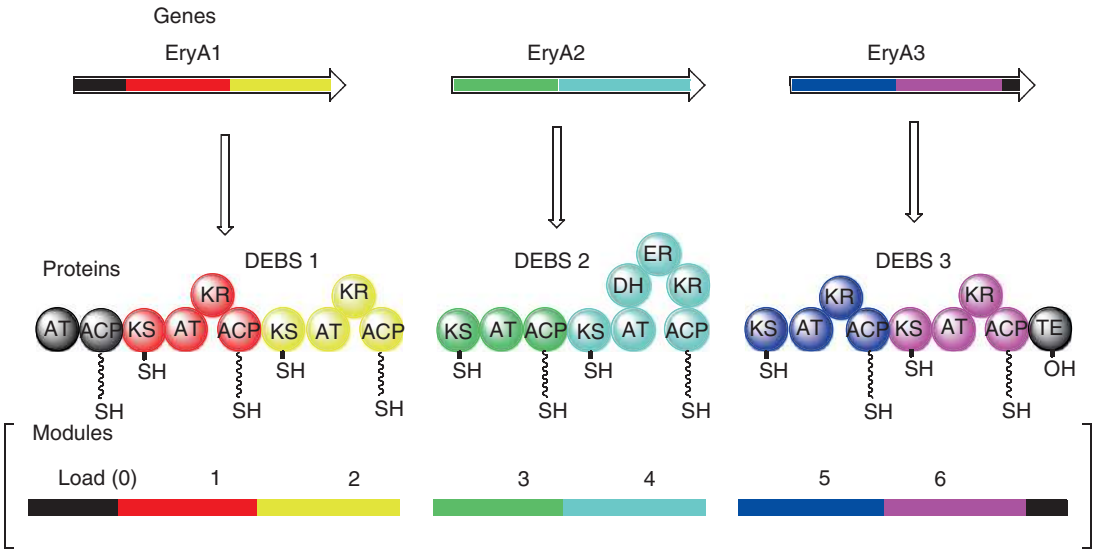
Although the enzymes for a variety of FAS systems had been isolated from various sources from the 1950s onward, the only PKS to be isolated was 6-MSA synthase from a penicillium mold.[31–34] This is a type I iterative multienzyme system which produces an aromatic product and so is not relevant to this chapter. In spite of enormous efforts in many laboratories, the search for the PKS proteins responsible for chain assembly on the ery pathway proved unsuccessful. For many years, workers in the field resorted to referring to the PKS as the black box, even showing it as such on their lecture slides. Fortunately, for the frustrated chemists and enzymologists, help was at hand from an enterprising group of geneticists and microbiologists.

The Hopwood group, based at the John Innes Laboratory in Norwich, UK, studied the genetics of the bacterium *Streptomyces coelicolor* which biosynthesizes an aromatic PK called actinorhodin.[35–37] Regions of the genome of the producer organism were sequenced to discover the location of the genes responsible for the PKS components (Chapters 1.03–1.07, 1.11). A region of the genome was identified which contained a collection of genes whose equivalent proteins showed strong homology (similarity of protein sequence, especially at the active site) with equivalent components of FAS systems. It turned out that the PKS consisted of separate enzymes and so was a type II system. The Hopwood group went on to develop DNA probes that could be used in other PK producers to identify regions of DNA coding for PKS components.[38] Many new PKS systems responsible for producing aromatic PKs were quickly identified as a result, but, disappointingly, the technology was not effective with the ery producer.[37] It turned out that the degree of homology between the DNA of aromatic and aliphatic PKS systems is not sufficiently strong for this probing strategy to be effective.

Fortunately, the Norwich group also investigated an alternative strategy based on identifying the gene for resistance to ery in the producer organism.[39] Without this gene, the producer could not survive biosynthesis of the antibiotic. Their hunch was that the gene for resistance would be located near the genes coding for the PKS, to increase the probability that the resistance genes and synthase genes would be expressed at the same time. It is beyond the scope of this chapter to describe this work in detail, but eventually the location of the resistance gene was determined.[37] The Hopwood group generously opened up this technology to the wider community. Two groups, the Leadlay group based in Cambridge and the Abbott group led by Katz, successfully exploited the opportunity to identify the genes responsible for the enzymes of the ery biosynthetic pathway.[40–43]

The search was far from straightforward, however, because nature had a major surprise in store. With the comparatively primitive technology of the time, sequencing DNA was an enormously challenging task, and it was only possible to sequence relatively short sections of DNA. This meant that disconnected pieces of evidence had to be fitted together based on overlapping sections of common sequence. As the lengths of DNA sequence gradually extended, it became clear that the ery PKS broke new grounds in terms of protein structure. In the first publication concerning these genes, the Leadlay group reported the existence of a giant gene coding for more than 3000 amino acids.[40] Shortly afterward, the Katz group reported two such genes,[41] one confirming the Cambridge data, the other corresponding to a second giant protein. Finally, a third giant gene was discovered.[43] These amazing discoveries, that the ery PKS consisted of three giant proteins, or megasynthases, excited wide interest both in academia and in the pharmaceutical industry. PKs moved to the forefront of biosynthetic research, and remain there to the present day.

Given the DNA sequences of the genes, the primary structure of all three giant proteins followed automatically. The challenge lay in predicting the secondary and especially tertiary folding of the protein chains. Fortunately, the precedents set by the enzymes of fatty acid biosynthesis provided a powerful lead. Examination of the primary sequence of the proteins revealed that each had multiple active-site motifs



Scheme 12 Relationship of the three giant genes to the three giant proteins; the modular organization of the megasynthases.

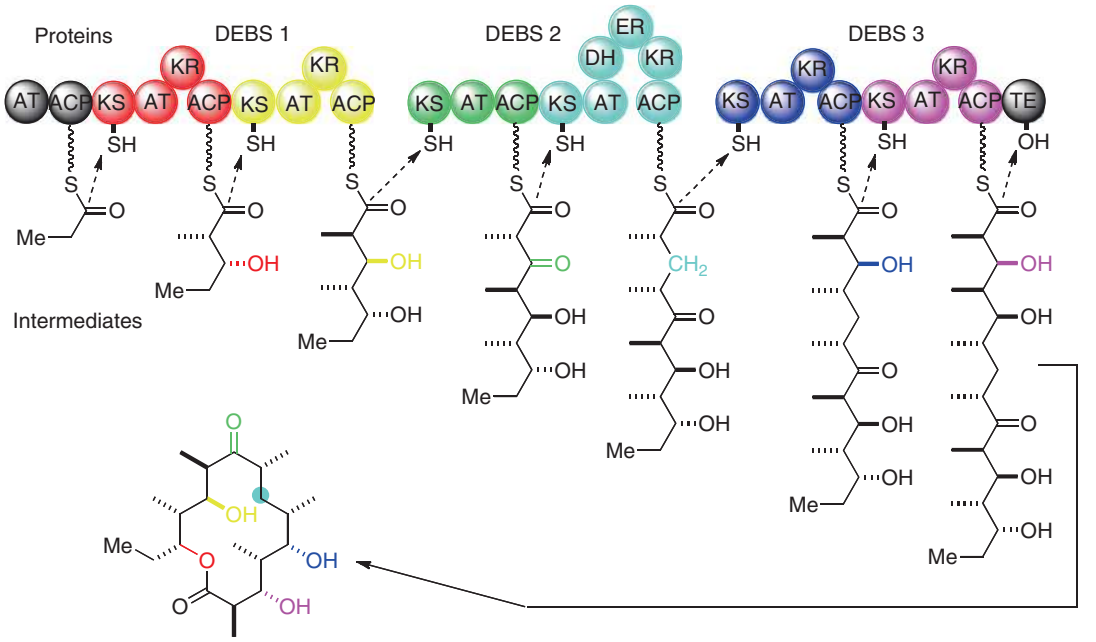
corresponding to the six catalytic activities required for the fatty acid chain extension.[40–44] It was argued that the protein around each motif would fold to create local sites of catalysis, called domains, as in the type I FAS systems. In a further parallel with the FAS, there were regions between the putative PKS domains that could serve as linkers.[15,24,42,45]

In all there are 28 domains spread across the three proteins, of which six are KS domains, which correspond to the six condensation steps required in the assembly of 6-dEB. Following each KS domain there is a set of domains leading to an ACP domain, in a pattern reminiscent of a type I FAS. This leads to the conclusion that there is a separate chain-extension cluster for each chain-extension cycle. These repeating sets of domains were designated modules. The complete picture of the distribution of domains and modules is shown in **Scheme 12**, with the domains in each module color coded.

The term modular PKS was coined to categorize this new class of megasynthase. An animal FAS is essentially a single module in which the domains are used repeatedly during each successive chain-extension cycle. It is therefore called an iterative system, as are the similar single-module PKS systems used in fungi to make both aliphatic and aromatic PKs.[7] Modular PKS systems seem to be the preserve of bacteria.

The order in which the domains appear in a module follows a set pattern. Module 4 has a full set of six domains, KS, AT, DH, ER, KR, and ACP, equivalent to those used for a chain-extension operation in the fatty acid synthetic cycle (see **Scheme 5**). Intriguingly, the order in which these domains occur is the same as that found in animal FAS systems (there are other parallels in the structure of the modular PKS and type I FAS systems, which will be explained later). In other modules, the number of keto group modification domains is reduced, but the order of the domains present always remains the same.

The next task was to speculate on the function of the three giant components of the PKS system, which came to be known as the DEBS (6-deoxyerythronolide B synthase). The key to this model for the DEBS is that the three giant proteins somehow dock together to form a gigantic molecular assembly line.[4] The giant proteins were designated DEBS 1, DEBS 2, and DEBS 3, in the putative order of operation in the synthetic operations, as shown in **Scheme 13**. At the front of DEBS 1, there is a didomain (AT_L–ACP_L), which was called the loading module (LM), whose function is to transfer a propionate unit from the external pool of propionyl-CoA to the KS of module 1. The domains of module 1 could operate as in the FAS cycle (as was illustrated in **Scheme 5**), except that the absence of the DH and ER domains ensures that the keto group modification can only go as far as the hydroxy stage. This hydroxy intermediate then transfers to the downstream KS situated in module 2, and the equivalent set of reactions would give a second hydroxy acyl intermediate.



Scheme 13 Intermediates passed forward at the end of each chain-extension cycle; the extent of keto group modification in each new C-3 unit matches the pattern of reductive domains in each module.

At this stage, the growing chain is passed from DEBS 1 to DEBS 2 and the third chain-extension cycle can commence. In this case, there is no active KR domain so the new keto ester product should move forward to the KS of module 4 without being modified. Here, there is a full set of FAS-type reductive domains so the next new β-keto group is converted into methylene. A second interprotein transfer delivers the growing chain to the KS of module 5. In the final two modules, two further rounds of chain extension, together with reduction of both the new keto groups to the hydroxyl stage, lead to the completed PK chain. At the end of module 6, there is a TE domain in place of the KS of a downstream module. The role of this domain is to carry out the essential process of removing the completed PK chain, as a macrolactone, rather than a free fatty acid. It is interesting to note that the sequence of intermediates predicted in **Scheme 13**, from the analysis of the composition of the modules, matches the sequence predicted in **Scheme 11**, based on retro-biosynthetic analysis. This degree of correspondence in the two strategies for analysis of structure and function was encouraging to the early investigators of the DEBS, and it continues to be important in current analyses of the putative role of newly discovered modular PKS structures.

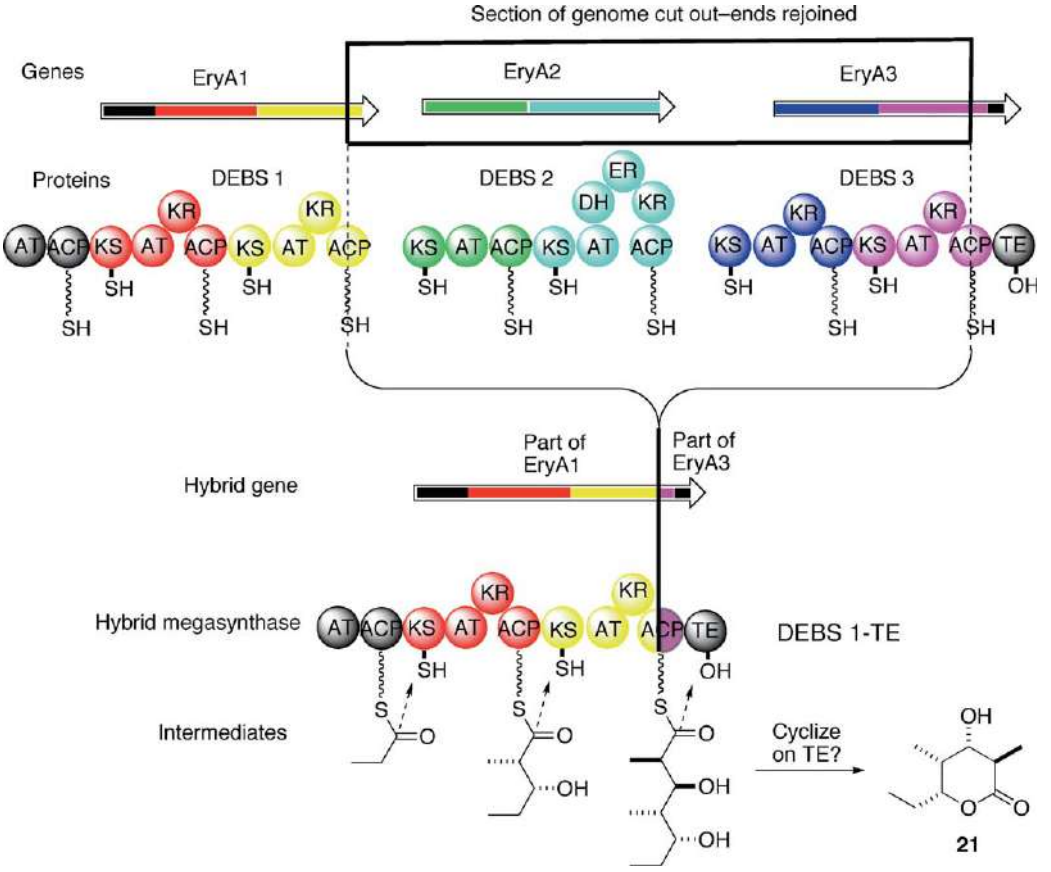
The conclusion from this analysis is that, given the sequence of domains in a modular PKS assembly line, and suitable clues as to the order in which multienzymes cooperate, it is possible to predict with a considerable degree of confidence what the product of the PKS might be. The small level of uncertainty arises from the occasional occurrence of silent domains, which retain the overall tertiary structure of an active component, but have a small structural defect that destroys the catalytic activity. Indeed, there is one such silent domain in the DEBS. In module 3 there is a defective KR domain located in the normal position.[42,43] This is not usually specified in the cartoon of **Scheme 13**, linking chemical structures to modules, because the silent domain contributes nothing directly to the chemical reactions. It may play an important indirect role, however, through its contribution to the quaternary structure of the megasynthase, as will be seen later.

Given the proven precedent of the FAS type I systems, the predictions for the DEBS look very persuasive. However, they needed to be subjected to experimental test, preferably with isolated proteins as had been the case with the type I FAS. A careful screening of a cell-free extract of proteins from *Sac. erythraea* using electrophoresis revealed the presence of three fine bands at high molecular weight consistent with the predicted masses of the DEBS.[46] Not surprisingly, given their unexpectedly high mass, these proteins had not been reported in previous studies. Further experiments confirmed that the giant proteins were indeed those predicted for the DEBS.[24–26]

1.10.3.4 Generation of Truncated Versions of the DEBS

The next problem was to find a protocol to study the catalytic behavior of the proteins *in vitro*. Two groups, one led by Leadlay and coworkers[47] and the other by Cane and coworkers,[48] addressed this challenge at about the same time, using very similar strategies to simplify the task. In both cases, a smaller model system was prepared in which the TE domain was transferred from the terminus of module 6 to the terminus of module 2. This created a mini-PKS consisting of just two chain-extension modules, 1 and 2, fronted by the normal LM, and terminated by a TE domain for product release. This truncated system was chosen to avoid the complications of multienzyme docking that is required for the functioning of the full DEBS. It was anticipated to produce (**21**), the δ-lactone form of the triketide intermediate. Formation of this product would confirm the synthetic role of DEBS 1, and its position in the putative DEBS assembly line.

Leadlay and Staunton named their construct DEBS 1-TE[47] (the dash signifying a covalent link or bond, not a minus sign). A DNA construct was created (**Scheme 14**) by deleting the section of the *Sac. erythraea* genome from a site corresponding to the active site serine (the point of attachment of the phosphopantetheine chain) of ACP2 to the corresponding site in the DNA for the equivalent residue of ACP6. On expression, this construct generates the required truncated protein with a hybrid ACP in module 2 in which the N-terminal portion was derived from ACP2 and the C-terminal portion was derived from ACP6. It was reasoned that the homology between the two ACP protein sequences was high enough for the resulting hybrid domain to fold into an active ACP species. The resulting mutant was grown in the usual way and it was found to produce the expected lactone product (**21**) in amounts equivalent to the yield of 6-dEB from the full DEBS system.



Scheme 14 Truncation of the erythromycin PKS to create a shorter PKS system as a model PKS for mechanistic and structural investigations.[47]

There are various possible explanations for this outcome. Obviously, the truncated construct with its hybrid ACP was able to carry out the normal synthetic operations of modules 1 and 2 to give the enzyme-bound triketide intermediate, but the mechanism of release was not clear. It was hoped that the TE domain would have a relaxed specificity toward the substrate and therefore that the mechanism would be enzyme catalyzed, involving transfer of the acyl chain to form an oxygen ester on the TE domain, followed by factorization. Alternatively, the product could have been released chemically because the triketide would be prone to form a δ-lactone by a spontaneous chemical process. The former mechanism would augur well for this strategy for genetic engineering-truncated PKS systems, whereas the latter mechanism could mean that this strategy for chain termination would be of limited applicability because of the lack of an efficient release mechanism. Not only would the yield be lowered, there would also be the serious risk that other reactive intermediates might off-load spontaneously leading to mixtures of products. Since this was the first pioneering effort to make an engineered PKS, it was important to resolve this ambiguity.

To directly address this issue, an alternative version of the truncated construct was made, in which the active-site serine residue of the TE was converted into an alanine, thereby rendering the TE domain inactive.[49] Incubation of this mutant again led to the formation of the triketide lactone (**21**), but the yield was greatly reduced to 10% compared with the productivity of the construct with an active TE domain. Clearly, this control experiment confirmed that chemical factorization does contribute to product release in the first experiment, but that the TE has a sufficiently broad specificity for it to be an effective agent for release of a PK chain much smaller than its normal substrate. There was, therefore, the promise that this strategy for producing engineered molecules might have wide applicability.

Interestingly, the American group used a significantly different splicing site for attaching the TE to the terminus of module 2.[48] The chosen site lay in the linker between the ACP and the TE, not within the region of either domain. The truncated PKS was given the name, DEBS 1+TE. Again, the expected triketide lactone product was observed. However, no control experiment was reported to determine the relative contributions of the spontaneous chemical- and enzyme-catalyzed mechanisms of product release, so there was an element of doubt concerning the mechanism of product release. Over the course of time, however, many further successful examples of the application of this strategy have been published,[50–56] so it would appear that the TE domain retains its ability to operate as a TE following both splicing strategies.

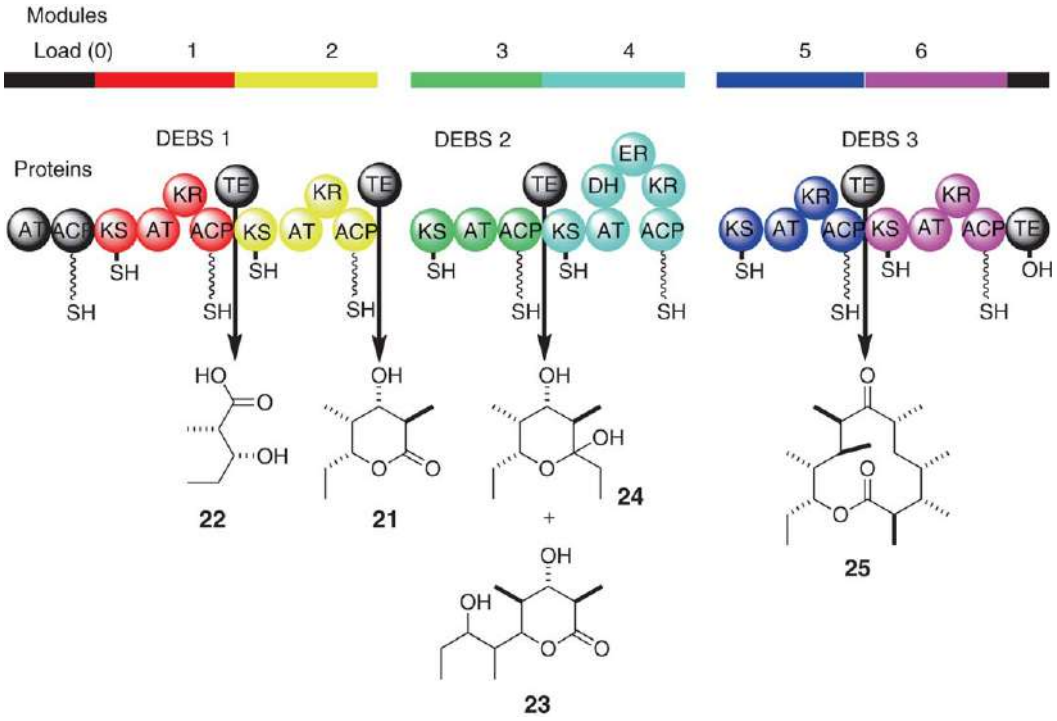
Following the encouraging success of these two pioneering studies,[47,48] many constructs have been made with TE domains as agents of product release. Of particular interest are those that are simple truncated forms of the DEBS (**Scheme 15**). Module 1 forms the diketide intermediate as a free acid (**22**).[50] The construct with the TE at the terminus of module 3 gave a lactone product (**23**) derived from the expected β-ketoacid, and a lactol (**24**) presumably formed by spontaneous decarboxylation of the ketoacid.[51] Of special interest is the formation of a 12-membered macrolactone (**25**) by the construct carrying the TE domain at the terminus of module 5.[52] This impressive example of the power of this type of genetic engineering to form novel natural products of considerable complexity will particularly interest synthetic chemists. The shorter synthases were more important for their contribution to studies of the mechanism and specificity of the synthetic operations.

Following these early successes, many other constructs have been made, which form novel PK products.[53–56] There has also been a lively investigation of expressing PKS systems, both natural and engineered, in foreign, heterologous hosts, such as *S. coelicolor* and *E. coli*.[57,58] There are formidable challenges in this type of experiment but impressive progress has been made. Some examples of this type of research will be given in later sections dealing with the potential for exploitation of the technology to produce novel natural products for commercial use.

1.10.4 Mechanism and Structural Specificity of Precursor Loading within the Modules of the DEBS

1.10.4.1 Specificity of Transfer of Acyl Group Building Blocks from External CoA Thioesters onto the DEBS Assembly Lines

The choice of building blocks in each module is an important issue in defining the carbon skeleton as the PK chain grows on a modular PKS. The wider range of possible building blocks is discussed later.



Scheme 15 Products released by truncated forms of DEBS with a TE domain at the indicated positions.[47,48,50–52]

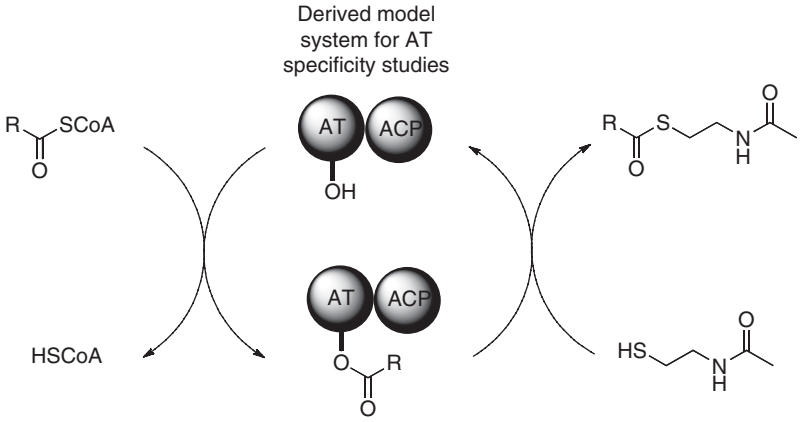
In the case of the DEBS, the starter acyl is propionyl. To ensure fidelity at this stage, the AT domain of the LM has to select propionyl-CoA as its substrate in preference to other potential starter units such as acetate, and forms the oxygen ester of its active-site hydroxyl group. From there, it is transferred onward to the ACP of the LM, and then on to the KS of module 1.

Similarly, the chain-extension unit added by all six chain-extension modules is methylmalonate rather than malonate, or one of the other substituted malonates used by modules of other PKSs. Again the AT domain in each module is presumed to be the prime agent of selectivity. Each AT domain transfers the methylmalonate to the ACP within the same module.

The extent of structural specificity under particular conditions depends on both the structure of the binding site, and the relative size of the pools of potential precursors. The importance of pool size for the LM is illustrated by experiments in which DEBS PKS constructs such as DEBS 1+TE were expressed in the heterologous host *S. coelicolor*.[59] In addition to the expected product, (21), with a propionate unit as starter, significant amounts of the analogue with acetate as starter were present. In contrast, DEBS 1−TE expressed in the native host, *Sac. erythraea*, incorporated only propionate into the starter unit.[47] This divergence probably reflects differences in the relative sizes of the pools of acetyl-CoA and propionyl-CoA in the two organisms.

1.10.4.2 Kinetic Studies of the Structural Specificity of AT Domains

There have been many kinetic studies of intrinsic AT domain specificity using various engineered constructs *in vitro*.[5,24] This approach presents unusual challenges because the AT domains of a PKS do not process their substrates, and then release a product to the medium. Instead, they transfer the bound acyl unit unmodified to a receptor in the form of a neighboring ACP. In the context of a working synthase, the substrate is then built into a PK intermediate and passes through many successive steps before the final product is released. The kinetics of such multistage operations could be formidably complex or simple, depending on the position of the rate-limiting step.



Scheme 16 A system for kinetic studies of AT domain specificity based on the loading module of DEBS.

The challenge for investigators who wish to use conventional kinetic methodologies for investigating the substrate specificity of an AT is to devise some simple form of artificial construct, which approximates to the context and chemical function of the AT domain in its natural environment. This is no easy task, and whatever solution may be adopted, one can never be sure whether the chosen artificial model is sufficiently close to the natural system for the measured kinetic data to have any significant relevance to the operations of an intact synthase.

These doubts have not deterred devotees of kinetic methodology from devising some ingenious systems that can be subjected to kinetic analysis.[24] An example is shown in **Scheme 16**, in which the experimental system is based on the structure of the LM of the ery PKS.[60] The system contains the AT domain under investigation, fused on its downstream side to the *apo*-form of the ACP domain of the DEBS LM, and on its upstream side to the N-terminal region (not shown) of the same module. Turnover was achieved by incubation of each test system with the various potential substrates as their CoA thioesters, in the presence of *N*-acetylcysteamine to act as off-loader in place of an ACP. The basis of the kinetic analysis was the rate of accumulation of the *N*-acetylcysteamine analogue.

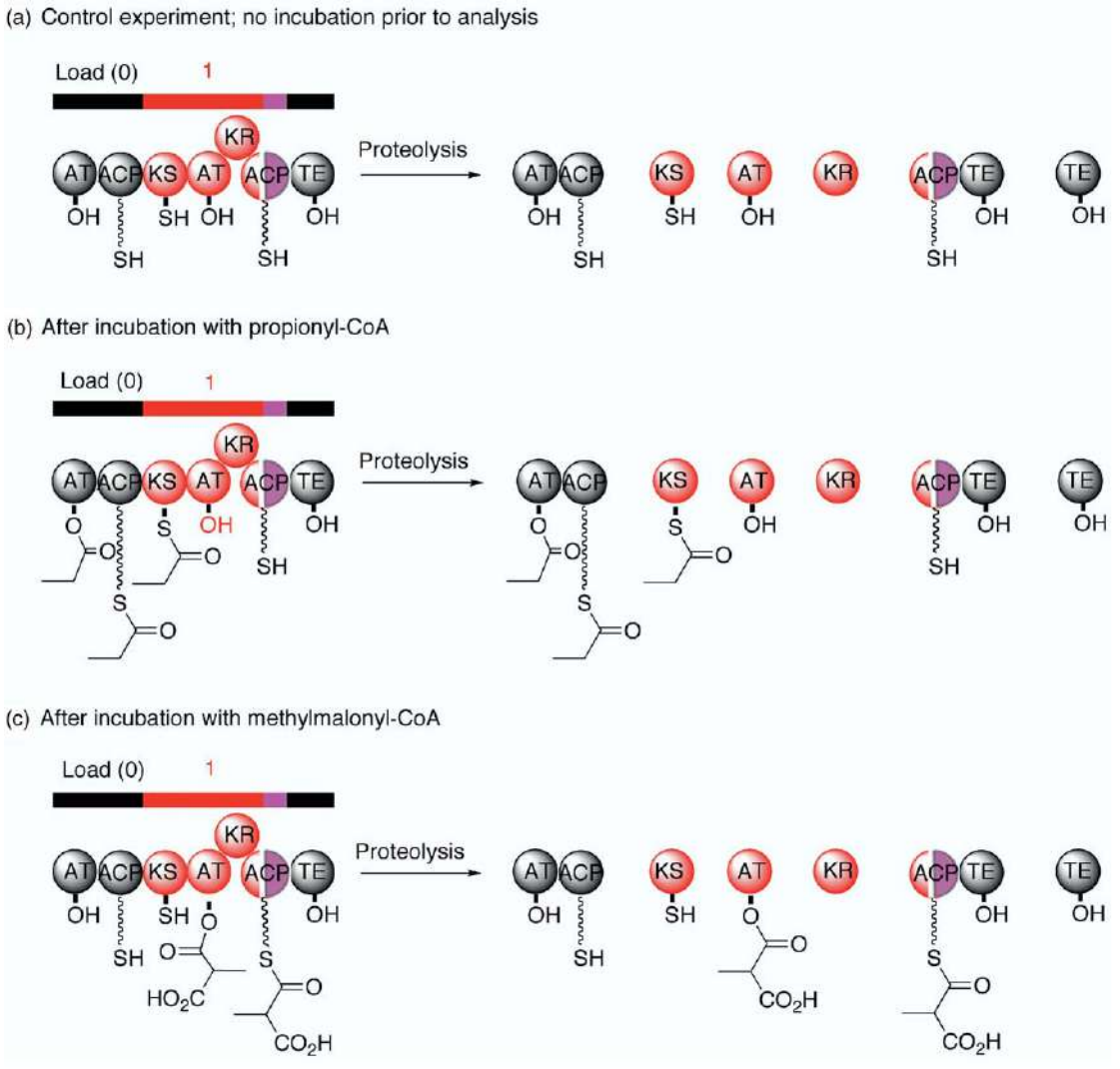
Using this protocol, it was shown that the natural AT of the DEBS LM had a relaxed specificity for saturated acyl groups from acetate up to butyrate. It was also possible to correct a claim in an earlier kinetic study using radiotracer technology that methylmalonyl-CoA could serve as a substitute for propionate as a source of the starter acid, by decarboxylation within the LM. Radioactivity measurements are notoriously at risk of error because of unsuspected radiochemical contamination in the commercial samples.

As a result of this work, a new series of kinetic parameters for turnover of fatty acyl-CoA thioesters by the AT of the loading domain were published.[60] The results show that all AT domains taken from chain-extension modules were highly specific for their natural substrate, either malonate or methylmalonate.

1.10.4.3 Mass Spectrometric Studies of the Structural Specificity of AT Domains

In an alternative innovative strategy to probe the specificity of AT domains,mass spectrometry was adopted as an analytical technique (H. Hong, Personal communication).[61] The test system was a diketide synthase prepared as explained earlier by positioning a TE domain at the terminus of module 1.[62] The diketide synthase could be analyzed in detail, domain by domain, using proteolysis followed by HPLC and mass spectrometry. Each protein fragment could be identified unambiguously by the measured mass. Several domains gave a single domain fragment, but the two ACP domains only appeared in didomain fragments, as shown in **Scheme 17** (H. Hong, Personal communication).[61] Fortunately, the mass of the ACP1 domain could be calculated from the values of its didomain fragment with the TE and that of the free TE.

After incubation with individual CoA thioesters of potential building blocks, the protein was cleaved into smaller fragments. Measurement of the mass of the various protein fragments, before and after incubation with substrates, gave the pattern of loading of substrates on all the possible sites.



Scheme 17 Studies of the specificity of loading building blocks onto an engineered model PKS, the diketide synthase.[61]

On incubation with propionyl-CoA, both the AT domain and the ACP domain of the LM, and the KS domain of module 1, became loaded with propionate; from the level of loading, it was apparent that all three sites were simultaneously loaded with starter unit in most molecules of the synthase. Other acceptable starter acids included acetate, butyrate, and valerate. Clearly, the substrate specificity of the AT of the load module is relatively relaxed, and equally significant, all three potential sites are simultaneously loaded (H. Hong, Personal communication).[61]

When the incubation was repeated with samples of methylmalonyl-CoA or malonyl-CoA, only the former substrate was incorporated. This shows that the substrate specificity of the AT1 domain is exceptionally high compared with the more relaxed specificity of the AT of the loading domain.

Mass spectrometry does have its limitations because, unlike radioactivity, it is not easy to obtain quantitative results, such as those determined in the kinetic strategy described in the previous section.[60,24] On the other hand, it has two important advantages. First, the structural identity of bound intermediates can be determined or confirmed directly, which minimizes the risk of spurious results produced by impurities in the compounds used as precursors. It also provides a convenient way to investigate intact modules in which domains are situated in their natural environments.

In view of these various reservations, it is reassuring to note that the two very different strategies for probing substrate specificity produced identical conclusions, at least in this particular set of studies.

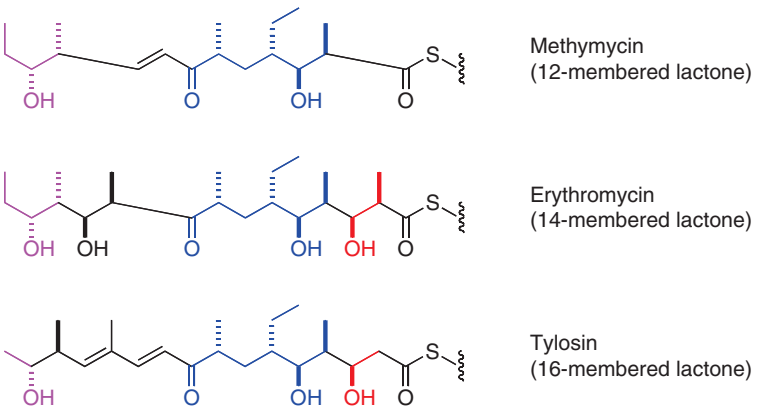
1.10.5 Experimental Investigation of Stereospecificity of the Reactions of the DEBS Chain-Extension Cycles

1.10.5.1 Celmer's Rules

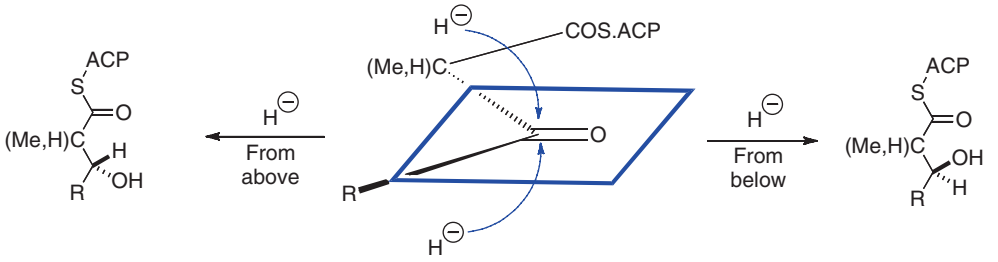
One of the most striking aspects of the aliphatic PKs is the high density of chiral centers. In the case of 6-dEB, for example, no fewer than 10 carbons of the 14-membered macrocyclic ring are centers of chirality corresponding to 1024 possible stereoisomers. The DEBS produces a single stereoisomer after six cycles of chain extension using apparently the same chain-extension reactions in each cycle. How is this diversity of stereochemistry achieved with such high fidelity using a very simple repetitive set of reactions?

In a comparison of members of the extended family of PK-derived macrolide antibiotics related to ery, Celmer noted the existence of intriguing common patterns of stereochemistry at structurally equivalent regions, which led him to formulate his so-called 'Celmer's rules'.[63,64] The molecular basis of these structural homologies, shown in **Scheme 18**, and the derived rules has long been a challenge to biosynthetic chemists.

From the various structures in **Scheme 18**, it is clear that there are two different types of centers to be explained. One is the methyl-branching center, and the second is the various hydroxylated centers. The basis of control of stereochemistry at the latter centers can be predicted to lie in the orientation of the intermediate ketoester in the KR active site. Based on the mechanism of reduction of the carbonyl groups of ketones and aldehydes by alcohol dehydrogenases, two orientations are possible, as indicated in **Scheme 19**. Reduction,



Scheme 18 The principles of Celmers's rules illustrated on the putative PKS products leading to various types of macrolide lactone.[63,64]



Scheme 19 Two directions of hydride addition are possible, leading to the alternative stereochemistries found at hydroxylated sites in polyketide chains.

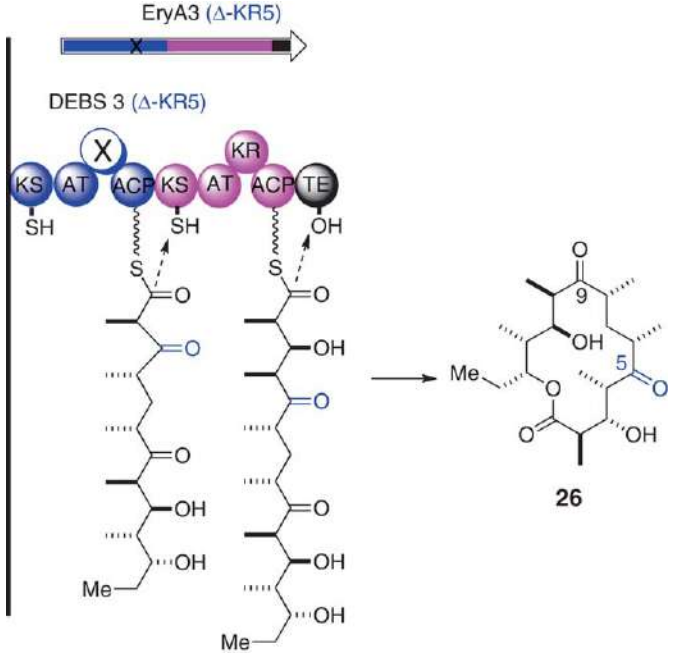
from one side or the other, leads to two alternate stereochemistries of the carbinol centers. It is known that the *pro*-4*S* hydride from the NADPH cofactor is delivered to the *β*-ketoacyl intermediate[65] and it is possible to predict which stereoisomer will be formed from the KR sequence.[66,67]

However, there is no such obvious precedent to explain the control of stereochemistry at the methyl centers. This is, therefore, the more challenging aspect of the problem, as we shall see.

1.10.5.2 Deleting Catalytic Activities

In the first successful experiment to produce a genetically modified version of the DEBS, the Katz group at Abbott deleted the activity of the KR group of module 5 by careful site-directed mutagenesis.[68] Deletion is indicated in the name of the construct by placing the symbol Δ in front of the affected domain. In the diagram, it is indicated by the symbol X. The synthase retained activity and, as shown in **Scheme 20**, it produced, as would be expected, a 5-keto analogue (**26**) of the normal product 6-dEB. This was a ground-breaking discovery because it confirmed that domains of the PKS downstream of the altered module could accept a structural analogue of the normal biosynthetic intermediate. The result encouraged researchers in the field to try other more ambitious alterations with a view to creating a wider range of structural analogues. It is for this exciting demonstration that alteration of a domain on a PKS assembly line can lead to production of a modified biosynthetic product, the paper is usually cited in reviews.

In the present context, it is the stereochemical implications of this development that are relevant. The configuration of the methyl group in the altered chain-extension unit is the same as that in the normal reduced product. This implies that the configuration at this center is determined in the condensation step and that it is unchanged in the normal reduction step. There is some uncertainty, however, in that the methyl center in the altered molecule is adjacent to a ketone function, and so could be subject to epimerization by enolization, not only on the enzyme, but also in the aqueous medium after release. The same uncertainty applies to the two chiral centers adjacent to the C-9 keto group in the natural product. It was, therefore, judged essential to devise an experimental protocol that would allow detection of enolization during the condensation and ketoreduction steps.



Scheme 20 Deletion of the activity of the KR domain in module 5 leads to the formation of the 5-keto analogue of 6-dEB.[68]

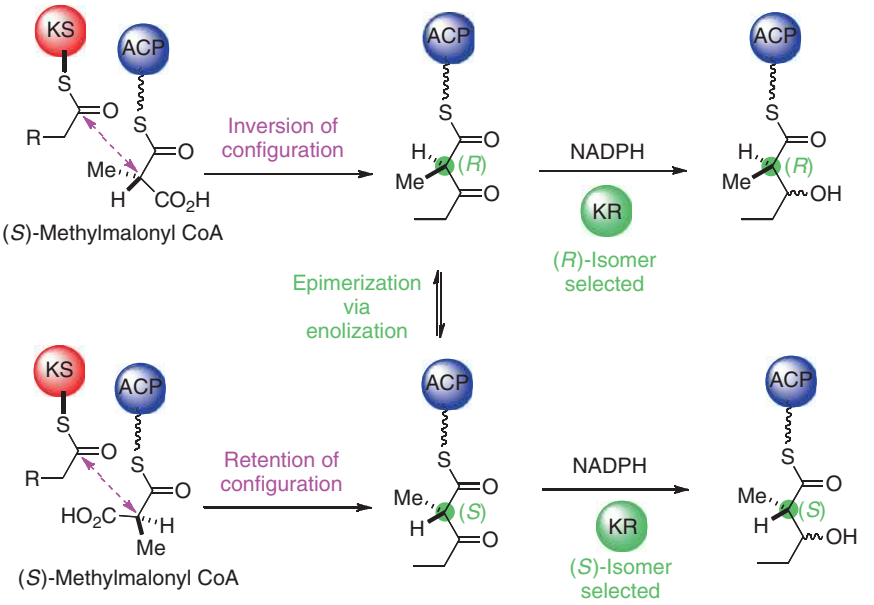
1.10.5.3 Determination of the Chirality of the Methylmalonate Precursors

The above experiments show that information concerning stereochemical mechanisms can be obtained as a by-product of genetic engineering experiments, but deeper insights can only come from experiments designed to probe the stereochemical events themselves. In the first such study, the three giant DEBS proteins were incubated separately *in vitro* with ^{14}C-labeled versions of the two stereoisomers of methylmalonyl-CoA.[69] For all three DEBS, radioactivity was incorporated from (*S*)-methylmalonyl-CoA, but not from the (*R*)-isomer. The extent of labeling steadily diminished on prolonged reaction, which was attributed to slow hydrolysis of the methylmalonate from the AT domains. This selective binding suggests that only one isomer, the (*S*), is incorporated by the AT domains of the chain-extension modules.

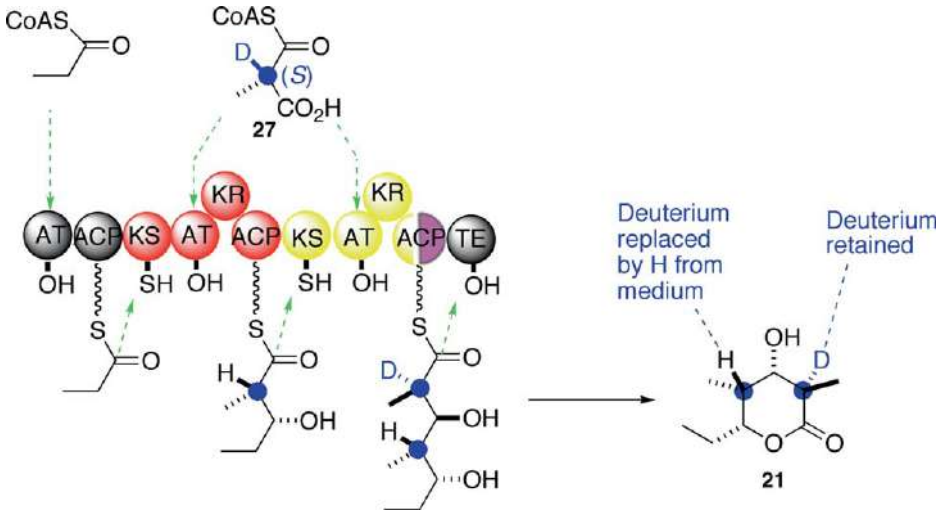
Scheme 21 shows a complex network of pathways that could provide routes to the two alternative epimeric forms at the methyl-branching centers from a single enantiomer of methylmalonyl-CoA. The alternative strategies involve either a different stereochemistry of carbon–carbon bond formation in the condensation step depending on the KS domain, or a single stereochemistry for condensation to give one epimer of the ketoester, which, if necessary, is then interconverted with the other epimer through enolization. In the case of enolization, the KR of the domain would have responsibility for ensuring that a particular isomer of the ketoester is reduced to the hydroxyester stage.

1.10.5.4 Studies of the Stereochemistry of Reactions using Intact Modules and Isotopically Labeled Precursors

Having established that only one epimer of the methylmalonate is used as a precursor in all modules of the DEBS,[69] the next task is to determine how different stereochemistries are created at the branching methyl centers in modules 1 and 2. One possibility is that the condensation reaction proceeds with retention in one module and with inversion in the other (**Scheme 21**). A second possibility is that the condensation reaction follows the same stereochemical course in both modules to give the same stereochemistry at the branching methyl of the ketoester. In this case, one of the centers would need to be epimerized. A logical process for this would be enolization, in which the hydrogen at the branching center is removed to give the planar enol form. Regeneration of the ketoester by addition of a proton to the opposite face would give the required epimeric form.



Scheme 21 Possible stereochemical variations at the branching methyl of the hydroxyester intermediate based on the use of a single enantiomer of methylmalonyl-CoA combined with reversible enolization of the intermediate ketoester.



Scheme 22 Incubation of DEBS 1-TE with deuterium-labeled methylmalonate to investigate the possible incidence of enolization at the ketoester stage.[70]

These proposals were investigated using the truncated model PKS, DEBS 1-TE as an experimental tool,[70] as indicated in **Scheme 22**. This system is an ideal candidate for investigation because, in one megasynthase, the two modules form different epimers at both the methyl-branching and the hydroxyl group centers. Obviously, different mechanisms of stereochemical operation would be required to achieve this outcome from a single (*S*)-stereoisomer of the chain-extension unit methylmalonyl-CoA.

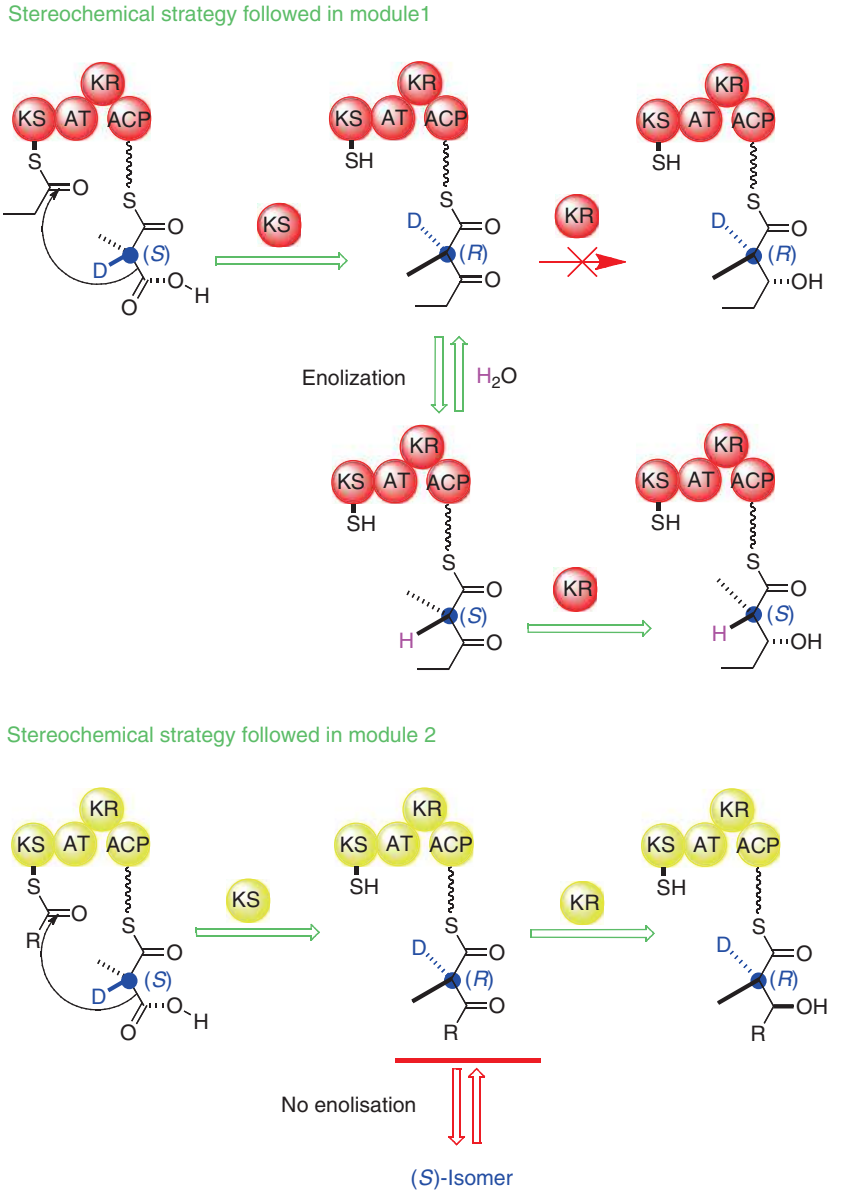
A purified sample of the synthase was incubated with (2*R*,*S*)-methylmalonyl-CoA in the presence of a supply of the starter unit, propionyl-CoA, and NADPH. Under these conditions, the triketide lactone product (**21**) was formed in sufficient amount for isolation and analysis by mass spectrometry. Following a hunch that an epimerization step based on enolization occurred in one module but not the other, the methylmalonyl unit was then labeled with deuterium at the chiral center marked by a blue spot in (**27**). On incubation with the enzyme in unlabeled water, a labeled version of product (**21**) was produced which retained deuterium at the methyl-branching position generated in module 2, but none was retained at the methyl-branching site generated by module 1.[70] After allowance for the results of control experiments, it was established that the isotopic label was completely lost in the chain-extension process carried out by module 1, whereas it was completely retained in the equivalent steps of module 2.

The explanation offered for these results in **Scheme 23** is that condensation of the (*S*)-methylmalonyl unit occurs with inversion of configuration in both modules to give a ketoester with the (*R*)-configuration at the new branching center. This is correct for product formation in module 2, so no epimerization is required before reduction by the KR. In module 1, epimerization is necessary, with resulting loss of the isotopic label, and its replacement with an unlabeled hydrogen species from the aqueous medium. The KR1 domain then selects the epimerized ketoester, the (*S*)-isomer, and reduces the keto group to a hydroxyl.

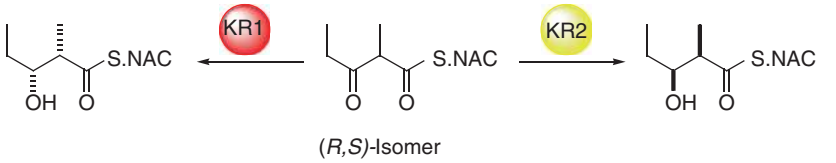
As explained above, it is not surprising that the two KR domains have differing stereospecificity for reduction of the keto group and so give two different configurations at the hydroxyl. What is significant is that in addition, KR1 is also indirectly responsible for control of stereochemistry at the methyl branch formed in module 1, by its proposed capacity to select only the epimerized isomer of the keto ester as its substrate.

This intriguing conclusion concerning the selectivity of the two KR domains has been tested by further experiments in which a synthetic analogue of the diketide ketoester intermediate, derivatized as its NAC thioester, was incubated *in vitro* with truncated versions of the DEBS, or with isolated KR domains.[71] The results, summarized in **Scheme 24**, were consistent with the proposals outlined in **Scheme 23**.

Another key early experiment was the study of the hybrid truncated PKS[47] shown in **Scheme 25**. This combined KS1 with the reductive KR2 of module 2. The sole product *in vivo* was a new diketide product, which matched the stereospecificity of module 2 at both the methyl and hydroxyl centers. This result supports the

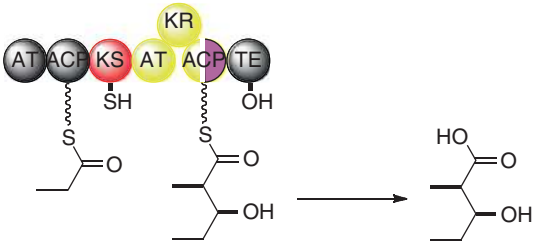


Scheme 23 Incubation of DEBS 1-TE with D-labeled methylmalonate; stereochemistry of the condensation step and keto group reduction steps in modules 1 and 2.



Scheme 24 Strategy for *in vitro* studies of the specificity of KR1 and KR2.[71]

proposal that KS1 initially makes the (*R*)-isomer of the ketoester and that this is subsequently epimerized by enolization to produce the (*S*)-isomer in the normal functioning of module 1. It is interesting that no product was obtained in which the methyl center was epimerized, presumably because KR2 efficiently discriminates against

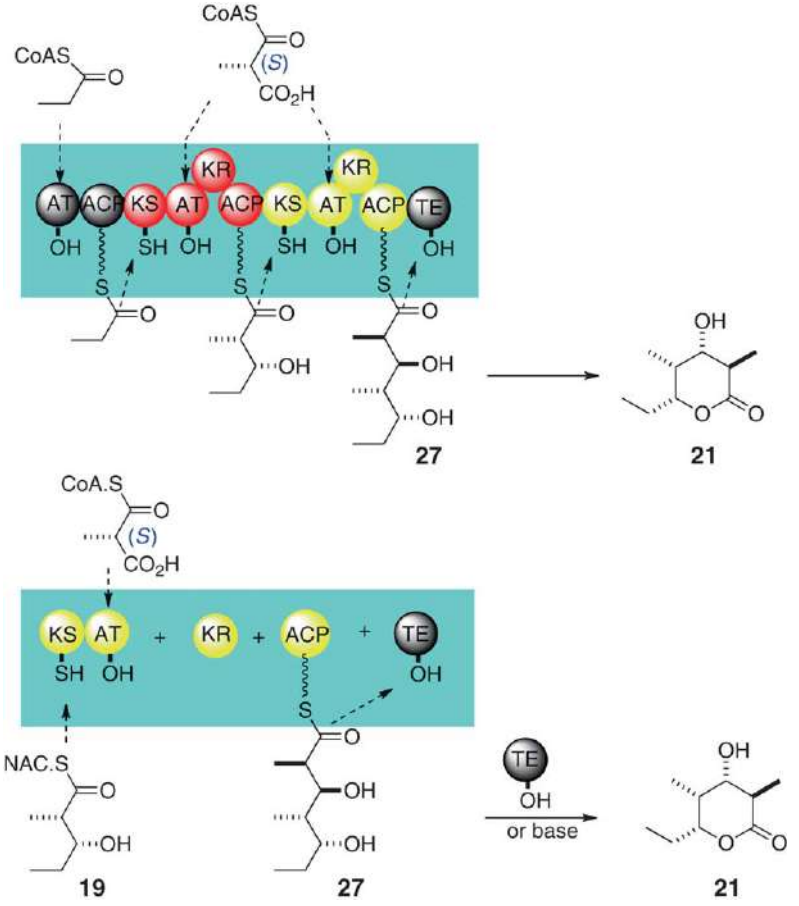


Scheme 25 Product formation on a hybrid diketide synthase combining the KS of module 1 with the KR of module 2.

this substrate. The seat of the epimerization remains to be established. There is a possibility that the epimerization observed in module 1 is a spontaneous chemical reaction rather than one that depends on catalysis by the protein.

1.10.5.5 Studies of the Stereochemistry of Reactions using Reconstructed DEBS Modules

Recently, Cane and Khosla published an alternative strategy for investigating stereochemical control in the DEBS modules using reconstructed modules comprising dissociated domains, rather than the normal intact multienzymes.[24,60,72–74] The strategy is illustrated in **Scheme 26**, which shows that component domains of a module can cooperate *in trans* to carry out a chain-extension cycle.[73] This strategy is convenient for rapid



Scheme 26 Comparison of the incubation strategies for *in vitro* experiments using intact synthases[70] (top) and combinations of isolated domains (bottom).[24,73]

screening of different combinations of domains. It avoids the considerable intellectual and practical demands of constructing large numbers of suitably tailored intact modules. Instead, only a small pool of individual domains needs to be prepared for mix-and-match experiments.

To initiate chain extension, it was necessary to use an unnatural process based on incubation of (**19**), a synthetic sample of a diketide synthase, derivatized as its NAC thioester. The diketide residue transfers on to the KS active thiol in a process widely used in mutasynthesis.[75] The products were analyzed by standard analytical techniques, especially mass spectrometry.

This strategy will be of particular interest to companies that aim to exploit genetic engineering strategies for the production of novel natural products.[76] It is less useful for fundamental studies of the structural basis of control of product formation in intact modules. However, there was one significant observation worthy of note. Attachment of a keto ester to an ACP seemed to reduce the rate of spontaneous chemical epimerization.[73]

This interesting paper poses some challenging questions concerning the operation of the DEBS. If the domains that make up the DEBS can operate effectively *in trans*, what advantage is gained from creating the three enormously complex megasynthase structures? Which is the more important factor in determining the specificity of the DEBS reactions, protein–protein interactions between individual domains, or the effects of positioning of domains relative to one another in space, under the constraints of the quaternary structure of the megasynthase?

1.10.5.6 Conclusion

In parallel with the above studies of working DEBS components *in vitro*,[24,60,72,73,76] there have been many equivalent studies based on the use of engineered constructs *in vivo*.[4,26,50,56,70,71] In general, these experiments have produced products or product mixtures consistent with the ideas deduced from the studies *in vitro*. There is a strong element of uncertainty in this approach, however, because the composition of the product mixtures produced *in vivo* can be seriously compromised by unpredictable effects such as selective degradation of some products, but not others, after release from the synthase. An additional complication, which militates against any attempt to draw trends from comparison of the results in these studies, is that sometimes the construct is expressed in the native host, whereas in others, a heterologous host such as *S. coelicolor* or *E. coli* is used. Heterologous expression will be covered in a later section reviewing commercial applications of genetic engineering technology.

It is apparent that some progress has been made in investigations of the stereochemistry of the reactions catalyzed by the DEBS but there are many important issues which remain unresolved. The evidence for an enolization step in module 1, based on the deuterium exchange experiment,[70] is very strong, but the mechanism and agent of the process remains unclear. The absence of enolization has only been firmly established for module 2; the absence of epimerization in the later modules does not necessarily rule out a nonproductive enolization process. There is much work to be done concerning the control of stereochemistry in the other modules of the DEBS megasynthase. Definitive information will only come from experiments with intact working modules that have the relevant quaternary structure, not from studies of reconstructed modules.

Finally, as we see later, there are many different types of PKS.[5–9] These may not follow the stereochemical precedents set by the DEBS. There is an enormous body of work waiting to be done more generally, possibly by new entrants to this field of study.

1.10.6 Studies of the Quaternary Structure of the Type I FAS

Since the discovery of the enormous size of the three DEBS components in 1990,[40,41] considerable effort has been devoted to studies of their structure. In part, the driving force is academic interest, but knowledge of the structure would greatly facilitate efforts to develop reliable protocols for rationale engineering of the synthase components to produce novel natural products. Studies of the DEBS have followed experimental procedures developed by investigators of the animal FAS megasynthases, and so it is helpful to begin with an overview of the techniques developed in that field and current proposals for the structure of animal FAS megasynthases.

1.10.6.1 Early Investigations of the Type I FAS Structure

In the 1950s, it was established that animal FAS systems from various sources are giant multifunctional type I systems, in contrast with the type II dissociable systems found in bacteria, as has been explained.[15] The protein sequences were determined by the protein sequencing technology of the day. Analysis showed that there was strong similarity between the individual proteins isolated from bacteria and sections of the animal type I system. This discovery led to the idea that an animal FAS consists of a string of catalytic domains held together in a set sequence by regions of relatively unfolded protein structures, considered to be linkers. Animal FAS systems from various sources were isolated and purified. Incubation studies with suitable precursors confirmed that purified versions of the isolated proteins were able to synthesize the predicted fatty acid products *in vitro*.

1.10.6.1.1 The homodimeric character and cross-linking studies

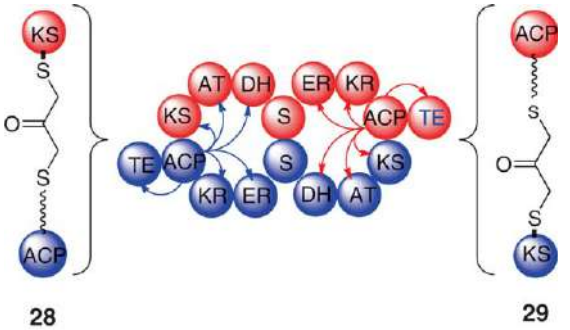
Surprisingly, given their already exceptional size, the megasynthases were found to be homodimeric.[15] A highly influential subsequent discovery was that on treatment with dibromopropanone, a chemical cross-linking that occurred between the active thiol of the ACP on one monomer chain, and the active cysteine thiol of the KS domain of the partner chain as indicated in structures **28** and **29**.[77] These two types of domains collaborate in forming the carbon–carbon bond in the condensation step, so this discovery was interpreted as an evidence that the KS of one chain is placed in proximity to the ACP of the partner chain.

1.10.6.1.2 The Wakil head-to-tail model for the type I FAS

To account for these results, Wakil proposed the homodimeric arrangement shown in **Scheme 27**.[78,79] The two chains are arranged side-by-side but running in opposite directions, with the KS of one chain (arbitrarily designated the head) placed alongside the ACP domain situated close to the tail end of the other. This alignment created two separate catalytic clusters, in both of which the ACP partners had three domains from its own chain (ER, KR, and TE), the three others (KS, AT, and DH) residing in the opposite chain. The primary site of binding to maintain the homodimeric structure was suggested to be the long linker region located between the DH and ER domains. This linker is large enough to be an active domain but it seems to have no catalytic function. The single letter, S, standing for spacer, is used to designate this region in the cartoon for the Wakil structure. This system of labeling is adopted generally in other cartoons, for relatively extended regions of primary structure that do not have a catalytic function, but are presumed to play a role in maintaining the quaternary structure required for effective synthetic operation. The alternative system of labeling such regions by a label to indicate their probable evolutionary origin (e.g., Ψ-KR) can be confusing for nonexpert readers.

1.10.6.2 Recent Studies of the Type I FAS Structure

The Wakil head-to-tail model for the animal FAS went unchallenged for several decades, but doubts began to emerge in the late 1990s as a consequence of some brilliant, innovative work carried out by the Smith group.[80–84] First, it was confirmed that inter-chain crosslinking does indeed occur with the dibromoacetone protocol,



28 **29**

Scheme 27 A cartoon showing the topology of the head-to-tail model proposed by Wakil for the animal FAS megasynthase and the structure proposed to account for head-to-tail chemical cross-linking after treatment with dibromopropanone.[78,79]

but that there is also intra-chain crosslinking.[84] Therefore, the assumption of a necessary head-to-tail arrangement was called into question.

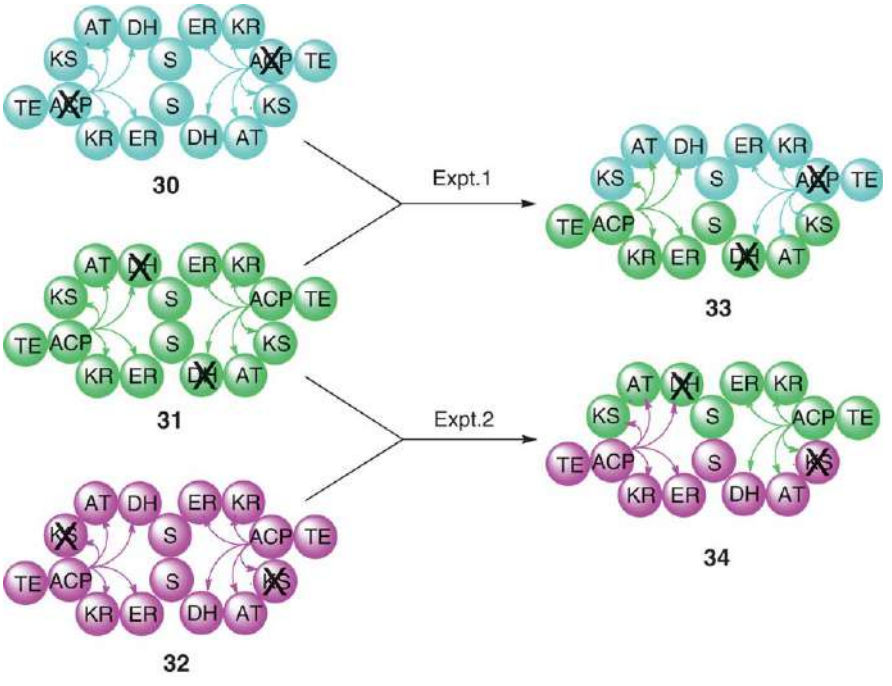
1.10.6.2.1 Mutant complementation studies, leading to the Smith structure for the FAS

A second, more significant, cause for doubt came from a novel strategy called mutant complementation developed by the Smith group.[85] In a normal FAS homodimer, the two chains are identical (even though they were differently colored in **Scheme 27** to keep track of the two chains contributing to the homodimeric structure). Targeted alteration through genetic engineering of the DNA will produce a new version of the FAS in which both strands of the protein carry the same change in structure. If the mutation destroys the activity of one of the catalytic domains, the resulting synthase will be completely inactivated in both channels of synthesis.

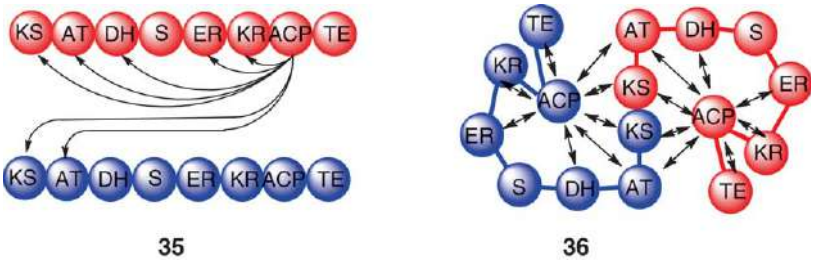
Mutant complement provides a method of redistributing the DNA strands of two different FAS systems so that a hybrid FAS system containing two different strands, one from each type of FAS, is formed. The protocol starts with treatment of both FAS systems to dissociate the strands to the monomeric state. The two sets of monomers are mixed and then subjected to conditions that promote dimerization. Statistically, a mixture of dimeric forms should be formed, which contains both original FAS systems (25% each), together with the new hybrid system in which one chain derives from the first FAS, the other from the second (50%). This reorganization opens an ingenious way to probe the pattern of cooperation between domains that occur in the FAS homodimer.

Two key discoveries are presented in **Scheme 28** to illustrate the principles of the technique.[81] In these experiments, three mutant FAS systems were generated in each of which the activity of one domain, the DH, the ACP or the KS, was inactivated by site-specific mutagenesis. Obviously, with one domain inactivated in both chains of each homodimer, none of the mutated FAS systems (**30**, **31**, or **32**) was able to synthesize fatty acids.

After mutant complementation in the first experiment,[81] the first pair of FAS systems (**30** and **31**), would be expected to produce a new hybrid homodimer with the distribution of defective domains shown in (**33**). According to the Wakil model, this new distribution of domains should have led to restoration of biosynthesis, albeit at a reduced rate of 50% compared with the original FAS, because only the left-hand reaction cluster would be competent. Surprisingly, there was no restoration of activity.



Scheme 28 Experimental strategy and results of mutant complementation studies with a FAS megasynthase.[81]



Scheme 29 Summary of mutant complementation experiments; allowed interactions of an ACP domain in the FAS megasynthase during a chain-extension cycle.

In the complementary experiment 2,[81] a different mutant FAS (**32**) with defective KS domains, was complemented with hybrid (**31**). In this case complementation results in a hybrid FAS (**34**) in which both reaction chambers of the Wakil model have one defective domain, and so it, like its two progenitors, should have been inactive. In fact, this hybrid mixture proved to be active. This was attributed to the fact, discovered in other mutant complementation studies, that each ACP can cooperate with both KS domains in the homodimer. On that basis, the right-hand cluster could be active.

The combined conclusion from these results is that each ACP can associate with the DH domain in its own chain in an intra-chain interaction at the dehydration step. More significantly, it cannot cooperate with the DH of the other chain, as would be predicted by the Wakil model. In this important respect, the Wakil model is found wanting.

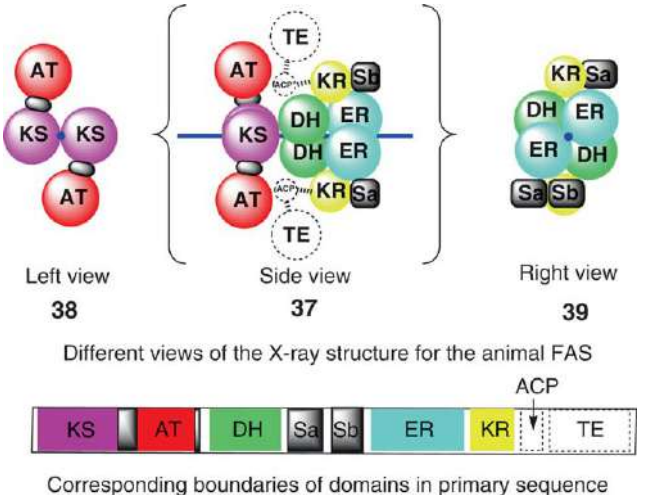
The Smith group performed a large set of these studies with strategically placed deletions in a range of mutant FAS candidates.[80–86] The general conclusions from the subsequent complementation studies are summarized in structure (**35**) of **Scheme 29**, where double-headed arrows connect one of the ACP domains with all the domains with which it can cooperate. Surprisingly, each ACP can interact with eight domains, the full set of six in its own chain, and in addition, the KS and AT of the cooperating chain. Even more intriguing is the pattern of noncooperation between an ACP and domains in the partner chain: whilst it can cooperate with the KS and AT domains of the partner chain, each ACP apparently cannot cooperate with the three reductive domains, DH, ER, and KR. This pattern of interaction versus noninteraction at different stages of the chain-extension cycle needs to be high on the list of factors to be rationalized in devising models for the FAS and PKS megasynthase systems, as we shall discuss later. A logical topology to account for these observations, proposed by Smith, is shown in the structural model (**36**).[86]

1.10.6.2.2 Direct observations of the complete type I FAS structure by X-ray crystallography: The Ban structure for the animal FAS

The internal tertiary structures of many isolated domains of FAS systems have been established by X-ray crystallography.[22,87–97] It is outside the scope of this review to give detailed pictures of the tertiary structures of individual domains. Instead, throughout this chapter, structures will be represented by simpler cartoons, in which spheres represent the various domains. For each type of domain, the volume of the sphere approximates to the relative molecular weight of the domain. These simplified representations have the advantage of focusing on the prime issue, the conformational shape of quaternary structures, based on the relative spatial positioning of domains. The complicated details of the internal tertiary structure of domains are, of course, vitally important in the complete understanding of the interplay between structure and function, but they can be set on one side in this introductory review.

In addition, all structures, real or postulated, are presented in a standard format with the axis of symmetry running horizontally rather than vertically. This standardization makes it easier to follow discussions of the relative merits of alternative proposed quaternary structures. It is also helpful in relating the structural cartoons to the standard cartoons in common use showing the positioning of intermediates on horizontal assembly lines.

The latest major breakthrough in the animal FAS field, as influential as the early proteolysis and cross-linking studies of Wakil and coworkers,[77–79] or the mutant complementation studies of Smith and coworkers,[80–86] came with the report by the Ban group of a high-resolution X-ray structure of the porcine FAS.[88] The image reveals



Scheme 30 The X-ray structure of the animal FAS with the position of the axis of symmetry shown in dark blue.

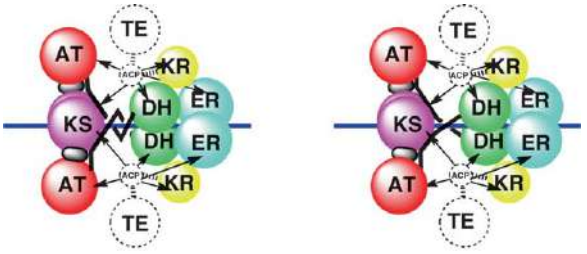
details of the tertiary structure, including linkers as well as domains. A cartoon showing just the quaternary arrangement of domains is shown in structure (**37**) (**Scheme 30**). As was explained above, a cartoon has been drawn with its axis horizontal, unlike the one presented in the original publication. Two additional views (**38** and **39**) are shown looking along the axis from different directions, to help establish the proposed 3-D structure. The various types of catalytic domains are colored differently to aid tracing the spatial arrangement in the quaternary structure, and also to relate the quaternary structure to the primary sequence shown in a linear form below. Unfortunately, the X-ray image did not reveal the location of the key ACP and TE domains even though the sample consisted of the complete homodimer. The missing domains are presumably relatively mobile or disordered in the crystal and so do not appear in the derived image.[88]

The structure is divided into two clusters of domains, one containing the KS domains plus the AT domains, the other the three domains involved in keto group modification, KR, DH, and ET, together with two extra globular domains designated Sa and Sb (these are called Ψ-Me and Ψ-KR, respectively, in the original paper),[88] which help maintain the quaternary structure. Linker regions join the two clusters through a narrow waist region in the middle of the overall structure. The arrangement of linkers at the waist is explained later.

The pairs of KS and AT domains lie in an approximately planar array, with the two KS domains in contact, and the two AT domains widely separated. An interesting feature of this grouping is the tight folding of the linker regions between the KS and AT domains to give small globular domains, each of which is indicated by a dark shape. The overall arrangement exhibits a twofold axis of symmetry. Another significant feature is the clockwise twist about the axis of the AT pair with respect to the KS pair, looking down the axis from the left.

The cluster of reductive domains has two pairs of domains, DH and KR, placed in a pseudo-dimeric arrangement on the axis, with the KR domains well separated from each other on the exterior of the cluster. In addition, this cluster has the two large blocks of globular structure, Sa and Sb, docked against the KR domains. This pair of contiguous domains corresponds to the equivalent structural region of the Smith and Wakil structures designated by an S. The positioning of the domains in this cluster is determined through a combination of the protein–protein interactions between the surfaces of the domains that make contact, with the tethering effects of the short lengths of linker between pairs of domains. The cluster is very well trussed and is clearly relatively stable and inflexible.

Although the ACP and TE domains did not appear in the X-ray image,[88] their positioning could be inferred from their position in the primary sequence, shown below the structural cartoon. The folded domains visible in the X-ray image are indicated on the primary sequence diagram by the rectangular blocks of color. The putative position and boundaries of the two missing domains, based on homology with known structures, are represented by rectangles bounded by dotted lines. Each ACP is tethered by a short linker to a particular site on the adjacent KR domain. This constraint places it in the nearby cleft of the structure. The locations of the TE



Scheme 31 Expanded view of **37** showing the route followed by the two AT-DH linkers, with alternative degrees of twist at the waist, and the consequential alternative patterns of cooperation of the two ACP domains.

domains, a short distance further out in the clefts, can in turn be hypothesized from the length of the tether between the TE and ACP domains. These proposed locations are indicated by dotted lines on the cartoon (**37**). It is interesting that one cleft is marginally wider than the other in the crystal structure, which means that the symmetry as a whole is not perfect, but this is ignored in the cartoon.

A further interesting feature of the structure can be found in the structure of the linker region between the AT and DH domains. These two linker regions join the two clusters through the waist region. Extended versions of the cartoon (**37**) are presented in **Scheme 31** to show the intriguing conformation of this part of the structure. In the Ban X-ray image,[88] represented by (**37a**), each of these linkers starts from the surface of its AT, and runs first to the site of the nearby KS–AT linker domain, where it makes a significant contact. It then runs across, and comes in contact with the surface of the KS domain until it reaches the waist. At this point, the two linkers leave the surface of the KS domains and twist together about the axis before joining to the appropriate surface residue of the downstream DH. In (**37a**), the extent of the twist amounts to a full turn; in (**37b**), the extent amounts to only half a turn. Apparently, there is sufficient ambiguity in the X-ray image of the region of proposed twisted linker for both the twisted and untwisted forms of the structure to coexist in the crystalline form of the megasynthase.

The pattern of cooperation between the two ACP domains and the catalytic domains in these quaternary conformations is also shown in structures (**37a**) and (**37b**), where the arrows indicate which domains each ACP can reach. Each ACP operates in a separate cleft lined with the full set of domains required for the chain-extension cycle. Considering structure (**37a**) first, it can be seen that all the cooperating domains reside in the same chain. This fits with the pattern for the reductive domains as determined in the Smith complementation experiments,[80–86] but it does not fit the observation that the ACP can reach both AT domains, one in its own chain, the other in the opposite chain.[98,99] The existence of a second, less twisted conformation, neatly accounts for this promiscuity, but it is important to recognize that the need for two conformations came from solution experiments, not the X-ray study.

1.10.6.3 Comparison of the Smith and Ban Proposals for the Type I FAS Structure

Interestingly, both of the revised models published recently, the Smith proposal[86] (**36**) in **Scheme 29**, and the Ban proposal[88] (**37**) in **Scheme 30**, place the two KS domains in contact with each other, rather than being separated by hundreds of angstroms, as would be predicted by the Wakil structure.[77] The models differ in their location of the reductive domains: in the Ban structure, the pairs of DH domains form a local homodimer, as does the pair of ER domains; in the Smith model, these domains are not dimerized but are looped out. Interestingly, the Ban structure[88] implies that the active sites of both the ER and DH domains are formed independently within each chain, and so both could be active in the monomeric form. To go from the waisted topology of the Ban model[88] to the looped out topology of the Smith model[86] would require a major reorganization of the quaternary structure. On the other hand, the Ban model[88] can only explain the complementation results if there is a major change of quaternary conformation, so, at this stage, there is a case to keep both models under consideration. This issue will be considered further after the review of PKS quaternary structures.

1.10.7 Studies of the Structure of the DEBS Modules

1.10.7.1 Isolation of the DEBS Multienzymes

The sequencing of the DEBS genes in 1990[40–43] pointed to the existence of a new form of megasynthase. The proteins were isolated shortly afterwards from the ery producer, *Sac. erythraea*, and their predicted giant size was confirmed. They soon became available in quantities suitable for structural studies.[100,101] Inevitably, the initial experiments were closely modeled on the early studies of the apparently related FAS megasynthases.

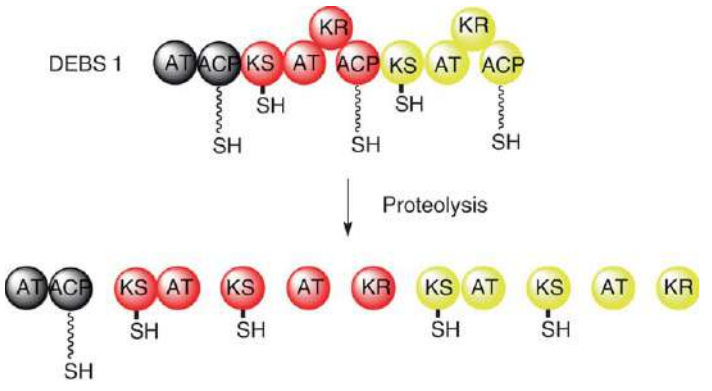
Obviously, the laborious task of determining the full primary sequence by conventional protein sequencing was avoided because the gene sequence was known. Remarkably, few errors in the sequence have subsequently come to light, and none is of any significance as far as structure and function are concerned. As was explained earlier, examination of the primary sequence revealed the active-site motifs for catalytic domains characteristic of those already identified in the earlier studies of the animal FAS.

1.10.7.2 Proteolysis Studies on the DEBS Multienzymes

Limited proteolysis experiments following protocols developed for the FAS, showed that the predicted domains are indeed present as reasonably stable globular entities.[69,102,103] The pioneering experiments relied on classical terminal sequencing technologies to determine sites of cleavage. In the more recent studies, mass spectrometry was used to identify fragments directly with extraordinary precision (H. Hong, personal communication).[61] The full set of fragments to have been isolated from DEBS 1 over the years is given in Scheme 32.

Depending on the conditions, the KS and AT domains can be isolated as separate domains, KS1, KS2, AT1, and AT2, as well as the corresponding didomains, KS1–AT1 and KS2–AT2. The two KR domains, KR1 and KR2, were also isolated as separate entities.[103] No identifiable fragment from modules 1 and 2 containing ACP1 or ACP2 was identified in the early work, but more recent studies using mass spectrometry to investigate fragments has shown that larger multidomain regions containing these domains can survive hydrolysis (H. Hong, personal communication).[61] The LM, AT_L–ACP_L, was also isolated from DEBS 1 as a didomain.[102] Cleavage sites resided in regions predicted to be linkers between domains, the actual site of cleavage in a particular linker being dependent on the proteolytic enzyme in use. In some linker regions, there are alternative sites of cleavage for a particular proteolytic enzyme, leading to more than one fragment for a particular domain.

Equivalent results were obtained from DEBS 2 and DEBS 3, which confirms that there is a common structural basis for all the chain-extension modules. The only additional result of significance was the identification of the TE domain at the end of module 6.[103] This was present in two fragments, a didomain fragment, ACP6-TE and as a single TE domain. One intriguing feature of these hydrolytic studies is the absence of any detectable quantity of an isolated ACP domain, apart from ACP5 derived from module 5.



Scheme 32 One- and two-domain fragments released by proteolysis experiments with DEBS 1.

1.10.7.3 The Homodimeric Character of the Multienzymes and Individual Domains

In the next phase of the investigation, the tendency of the DEBS components to dimerize was investigated. The results for the three intact DEBS caused a great stir because the giant proteins, individually already in excess of 300 kDa in size, were found to form strong homodimers of combined mass in the 600 kDa range. The discovery of the dimerization was also an important milestone because it established a further parallel in structural behavior between the DEBS and FAS megasynthase systems.[88]

Dimerization studies on the domain fragments of the DEBS gave further significant insights, which pointed to important differences between FAS and PKS megasystems. In solution studies, none of the isolated domains of the FAS appeared to exist as a strong homodimer. In contrast, two types of domain of the PKS do exist as very strong homodimers in solution.[69,102,103] These were fragments containing a KS domain, or the terminal TE domain. This was an exciting new discovery, which triggered thoughts on a radically new model for the topology of the modules of the DEBS, which came to be called the Cambridge Double Helical Model.

In another experiment following the precedents set by the earlier investigations of animal FAS systems, the DEBS modules were cross-linked by treatment with dibromopropanone.[103] The results of proteolytic cleavage of the cross-linked products are consistent with an interchain linkage between the KS of one partner chain and the ACP of the other chain.

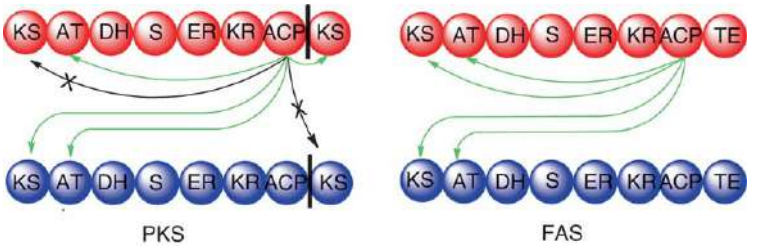
1.10.7.4 Mutant Complementation Studies

The first new information has come from mutant complementation studies by Cane and coworkers[104,105] on the truncated PKS, DEBS 1+TE, based on the protocols developed by Smith. The results of an impressively wide ranging and thoughtful set of experiments are summarized in **Scheme 33**, in the form used earlier for the equivalent FAS studies. The two chains are arranged in the head-to-head, tail-to-tail mode.

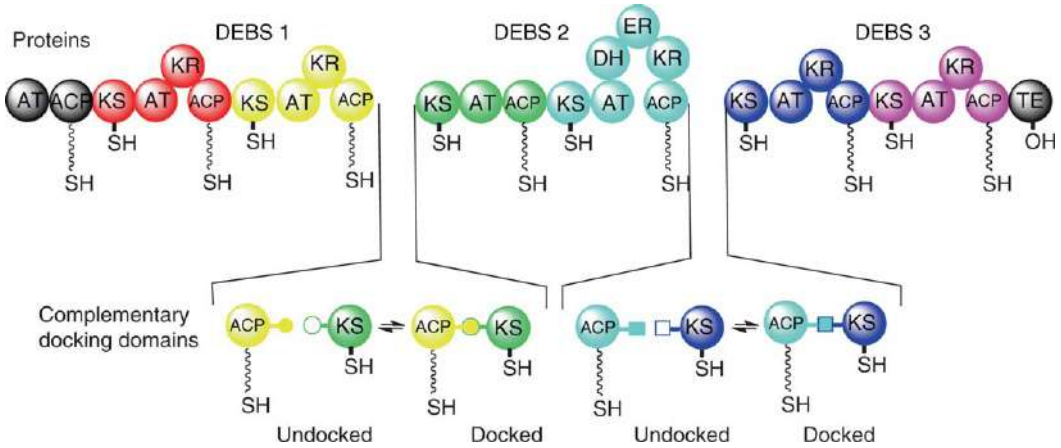
The aim of the experiments was to investigate the extent to which a given ACP cooperates with the two KS domains and two AT domains sited in the upstream region of its pair of modules, and the next pair of KS domains downstream to which it passes its completed chain-extension unit. The results show that the interactions with the KS domains are highly specific, in that a given ACP cooperates with the KS of its own chain downstream, and with the KS of the other chain upstream.[104] In contrast, the interactions with the AT domains are promiscuous, with a given ACP able to cooperate with either of the two AT domains upstream.[105] At present, there is no published information concerning the cooperation with the reductive domains.

1.10.7.5 NMR Studies of Docking Domains

Weissman and coworkers conducted a ground-breaking NMR study into the structure of the docking domains used in DEBS to bring separate multienzymes into collaboration.[106] A cartoon showing the ideas on which the study was based is shown in **Scheme 34**. The DEBS consists of three multienzymes that have to come together in some way to allow the passage of the growing chain from one multienzyme to the next. It was anticipated that there must be docking motifs to ensure that the multienzymes are held together, if not permanently, then sufficiently long to allow downstream transfer of a biosynthetic intermediate from the terminal ACP of one multienzyme to the KS domain at the start of the next multienzyme. There also needs to be a highly effective



Scheme 33 Summary of mutant complementation experiments on DEBS modules and comparison with the animal FAS; only interactions between an ACP domain and potential KS and AT partners are shown.[104,105]



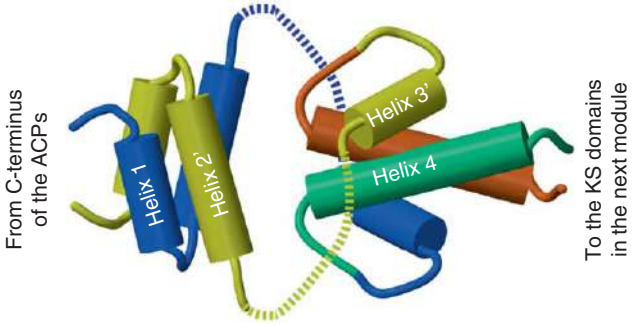
Scheme 34 Proposed mechanism of docking of the DEBS multienzymes.

form of discrimination in the docking process in order to ensure that only the correct pairs of multienzymes become associated. It was considered likely that the docking domains would lie in the C-terminal residues of the ACP partner and the N-terminal residues of the KS partner.

In initial exploratory experiments, covalent links were created between DEBS 1 and DEBS 2 and between DEBS 2 and DEBS 3 (**Scheme 35**).[107,108] These artificially joined systems retained an ability to participate in the biosynthesis of the normal macrolide metabolite, 6-dEB. In a follow-up study,[106] genetic engineering was used to prepare short protein fragments containing the C-terminal portion of the upstream ACP and the N-terminal end of the downstream KS domain, joined together by the artificial covalent linkage. The covalent linkage was anticipated to increase the chance that the isolated fragments would fold locally to form the domains and that these would associate in the manner responsible for docking.

The NMR study showed that this strategy had proved successful. The docking domain positioned at the N-terminal end of DEBS 3 (and on the basis of sequence comparison, the same is likely to be true of DEBS 2) takes the form of a rigid rod-like coiled-coil structure, consistent with earlier predictions from primary structure analysis. The section of the protein derived from the carboxyl terminus of the upstream ACP forms three helical regions, 1, 2, and 3. Helices 1 and 2 form a surprising structural motif consisting of an intertwined four α-helix bundle, which may serve as a dimerization motif for the tail end of the multienzyme. This is followed by a stretch of mobile linker region leading to the third helix, 3, which docks with the rod-like docking domain of the KS. This association between helices from different multienzymes can be considered the docking motif. The artificial join between the two original multienzymes is marked by the color change in the linker between helices 3 and 4.

This study broke new ground by providing the first images of regions of a PKS that lie outside the globular regions of the primary catalytic domains.[106] In follow-up papers, Weissman[109,110] went on to decode some aspects of the recognition processes between helices 3 and 4, which ensure that the correct pairs of



Scheme 35 Structure of the artificially linked N-terminal residues of DEBS 3 and the C-terminal residues of DEBS 2. One dimerization motif contains helices 1 and 2; the docking domains are helix 3 (ACP) and helix 4 (KS).

multienzymes associate. This is a vital aspect of fidelity control in the functioning of modular PKSs, but it is beyond the scope of this review to go into details of the experiments. The same is true of some important experiments by Cane and coworkers,[111–113] which showed that surface motifs on the docked ACP and KS domains may also contribute to the specificity, as well as the stability, of the docking interactions.

1.10.7.6 Characterization of a DEBS Didomain, KS–AT, by X-Ray Crystallography

As has been mentioned earlier, portions of the DEBS multienzymes have been expressed as isolated fragments containing certain catalytic domains (e.g., KS and TE) have been crystallized and analyzed by X-ray crystallography.[113–115] The parts of the structures responsible for their catalytic activity strongly resembled equivalent entities found in both type I and type II FAS systems. Interestingly, both the KS and TE domain exist as strong homodimers. In another very significant X-ray study, a fragment of DEBS 1 containing the KR domain plus its associated globular spacer domain (Ψ-KR or KR^s) was shown to have the folded domains side by side, presumably similar to their disposition in the intact module.

An even more fascinating insight into the possible quaternary structure came from the recent report from Cane and Khosla of an X-ray crystallographic investigation of the KS-AT didomain from module 5.[113] The high-resolution image gave an excellent 3-D picture of the homodimeric didomain. The two domains in each monomeric protein species are folded into globular residues whose internal structures follow well-established precedents set by equivalent domains from other sources. The quaternary arrangement of the dimer is shown in **Scheme 36**, using the now familiar illustrative device of representing domains by spheres (note the long rod-like docking element attached to the KS domains (first visualized in the NMR study by Weissman and coworkers[106]) is omitted from this diagram).

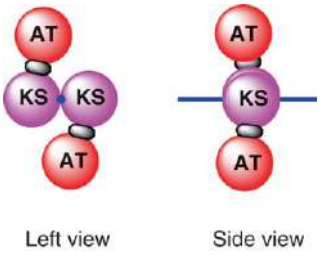
A striking feature of this X-ray structure,[113] which predates the high-resolution structure of the animal FAS,[88] is the positioning and orientation of the two globular domains. The KS domains form a strong homodimeric interface, with the AT domains placed well apart from each other in a manner very similar to the equivalent parts of the animal FAS. The quaternary arrangement of this KS-AT didomain, with its twofold axis of symmetry, shows an impressive similarity to that found for the equivalent region of the animal FAS.

Another fascinating similarity, which may give clues to the active quaternary structure of the complete megasynthase, is the positioning and structure of the linker regions. Of greatest interest is the tightly folded linker between the KS and AT domains, which is closely similar to the equivalent FAS structure. The other linker downstream of the AT is not tightly folded but lies in essentially an extended form across the surface of the AT domain. The significance of the positioning of this linker is uncertain because it lacks the protein structure normally found at its C-terminal end.

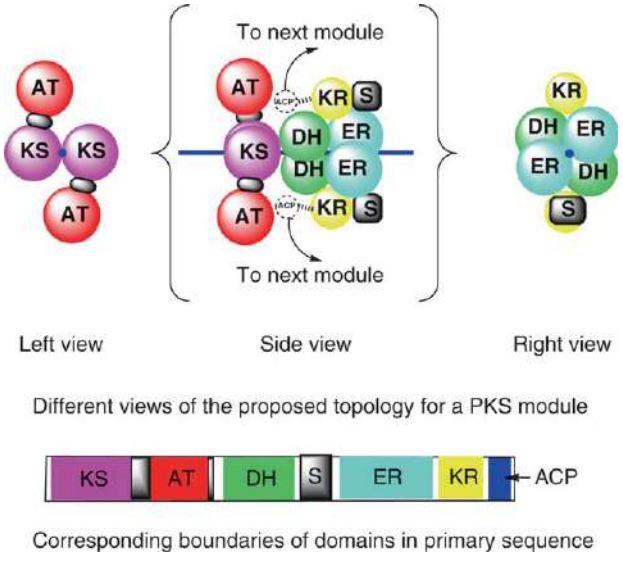
1.10.8 Current Proposals for the Topology of the DEBS Modules and Multienzymes

1.10.8.1 Structures Based on X-ray Images

Cane and Khosla proposed a topology for the full DEBS module 4, based on the X-ray data from their study of the KS-AT portion of module 5.[113] For the topology of the reductive domains downstream, they borrowed the topology proposed by Ban for the equivalent portion of the FAS megasynthase.[87] Inevitably, the overall



Scheme 36 Cartoon version of the X-ray image of the KS-AT didomain of DEBS module 5.



Scheme 37 A proposed topology for DEBS module 4[113] based on analogy with the animal FAS X-ray image.[87]

structure for the putative PKS module is essentially identical to the Ban structure for the FAS,[88] so the same cartoon, shown for the animal FAS in **Scheme 30**, is repeated in **Scheme 37**, with the significant replacement of the TE domain by a proposed downstream link to the KS of the next module.

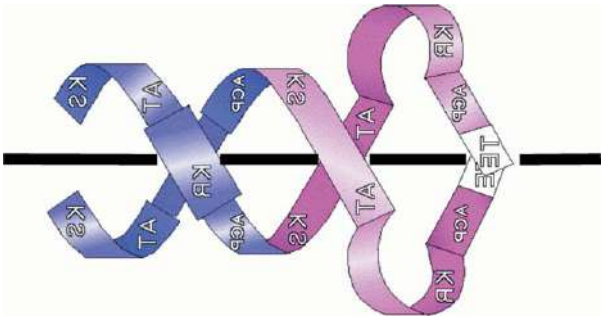
In an interesting exercise in molecular modeling, Keating-Clay and Stroud[114] started from their X-ray structure for the reductive loop region of module 1 of the DEBS to a produce a quaternary arrangement for a full DEBS module. Remarkably, this also turned out to be sufficiently close to the Cane and Khosla arrangement for the cartoon shown in **Scheme 37** to be an adequate representation. Finally, when they published their high-resolution image of the animal FAS,the Ban group also noted that this topology could serve as the basis of a general topology for PKS modules.[88]

In recent comments on the new FAS structure, Weissman gave a generally positive evaluation of the significance of the high-resolution X-ray image of the animal FAS.[45] She expressed concern, however, in relation to the direct extension of the new FAS topology to inform models for PKS architecture, as indicated in **Scheme 37**.[45,116] One source of worry is the relative shortness (about 80 Å) of the linker joining the ACP of one module to the KS of the next, in pairs of modules that are covalently linked. Can this linker stretch the considerable distance from the pair of dimerized KS domains in the downstream module to an ACP domain situated in the appropriate reaction cleft of the upstream module? She also noted that the length of the linkers (about 40 Å) connecting the ACP domains to their neighboring KR domains on the upstream side might be too short for the ACP to make a necessary contact with its downstream KS partner, so that the chain can be passed on.

It will be interesting to see whether these issues can be resolved without invoking significant movement of the catalytic domains from their fixed positions in the quaternary constellation revealed in the X-ray study of the animal FAS.

1.10.8.2 The Cambridge Topology for the PKS Module

The earliest topology proposed for the PKS modules in DEBS is illustrated in **Scheme 38**.[103] The cartoon uses a ribbon style cartoon to show the topology proposed for the complete multienzyme DEBS 3. The image shown is the mirror image of the original to take into account the direction of twist revealed the X-ray image for the KS–AT didomain of module 5,[113] as will be explained below. The black line marks the position of the twofold axis of symmetry in this topology. Given the question marks that hang over the topology based on X-ray structures, this topology deserves more detailed consideration.



Scheme 38 The ribbon cartoon used to illustrate the principles of the Cambridge topology for DEBS 3. The illustration is the mirror image of the original model (including, to emphasize the point, the letters in the captions for the domains).

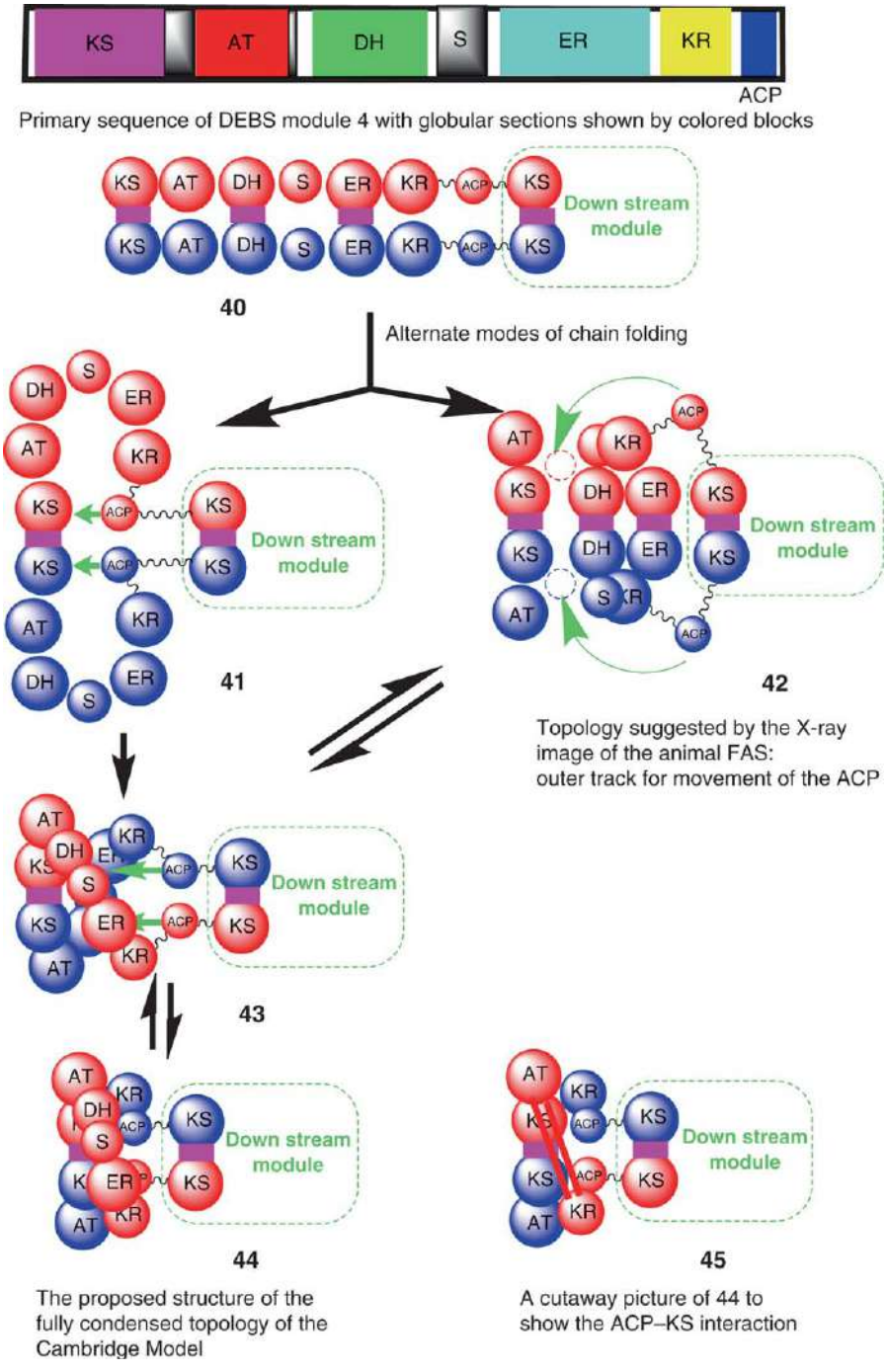
To aid comparison, the logic underlying this topology will be developed using cartoons based on spheres to represent globular domains. At the outset it should be emphasized that the topology is derived by logical deduction, not by inspection of any image generated by a physical technique (there were none available in the early 1990s). The starting point is the primary sequence, and the proposed division into globular sections and linker sections. Next to be considered is the evidence that the KS and TE domains of the DEBS are strongly homodimeric in isolation from the multienzyme,[102,103] which made it logical to place them in contact inside the multienzyme. This led inevitably to a head-to-head arrangement of the chain domains and the starting point for the dimerization process shown in the cartoon (**40**) of **Scheme 39**. The primary sites of local dimer formation are the KS domains. The DH and ER domains can now be added as potential dimerization sites based on recent evidence from X-ray studies.[88]

The unfolded dimer can fold in two different ways to give quaternary conformations leading to either (**41**), in which all the domains between the ACP and upstream KS are looped out, or (**42**), in which the DH and ER domains are left associated on the axis, and only the KR and spacer domain are looped out. Structure (**42**) is the topology suggested in **Scheme 37** to conform with the X-ray evidence from various systems.[88,113] The key feature distinguishing topologies (**42**)[88,113] and (**44**)[103] from the operational standpoint is the ability of the ACP to reach its KS partner in the condensation step. In terms of linker lengths, this is an easy operation in (**41**), but is problematic in (**42**), as Weissman and Müller noted.[45]

The problem with proposed intermediate (**41**) is that it predicts the wrong pairing between ACP and KS domains for the condensation step. This is simply resolved by twisting the downstream module through a half-turn to give a double helical arrangement in (**43**).[103] Now when the ACP domains move to make contact with the KS domains in structure (**44**), the correct pairing between domains in opposite chains is achieved. This is shown in (**45**), the cut-away version of (**44**), in which the DH, S, and ER domains have been removed to reveal the inner working of the topology.

A final feature of the interchange of conformations in **Scheme 39** is the suggested interconversion of the open structure of the Cambridge Model (**41**) with the closed topology in conformation (**42**). This can be best understood by referring back to the equivalent FAS structures shown in **Schemes 30** and **31**.[88] These structures have a twist built into their topology but localized to the so-called waist region. Because of this feature, it is possible to envisage conformer (**42**) opening up, origami-like, simply by unzipping the interfaces between the DH and ER domains and allowing them to move outward. This transformation would have a transition state to overcome, but the significant energetic consideration is the relative stability of the two forms. This cannot be judged in the absence of knowledge concerning the interdomain interactions in (**44**), which would replace those visible in (**42**).

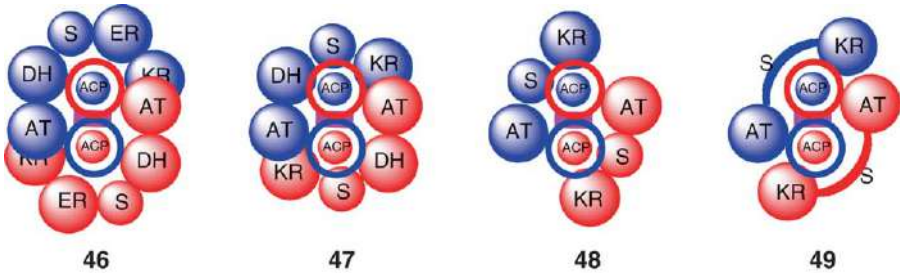
All the views in **Scheme 39** are from the side. In **Scheme 40**, an alternative set of views of the Cambridge Model looking down the axis help establish further features of the proposed conformation. The KS domains at the head of the modules are shown as see-through circles to reveal the position of the two ACP domains. The relative orientation of the KS pair of domains to the AT pair is based on the X-ray structure provided by Cane and Khosla.[113] These views make the point that the outer domains, AT to KR, form a ring of protein structure around the pair of ACP domains at the core. This makes it easy to see that the individual ACP domains can



Scheme 39 Suggested sequence of steps to convert the initial unfolded dimer (**40**) of a PKS module with a full set of reductive domains to the folded topologies of the two current models for DEBS modules, and also to allow them to interconvert.

reach a full set of domains, and that the pattern of favored interactions is consistent with the results of mutant complementation studies.[104,105]

The reductive loops in the first two modules (**46** and **47**) form a complete circle that could derive added stability by vertical interactions between the AT domains and the KR domains of the opposite chain. By the



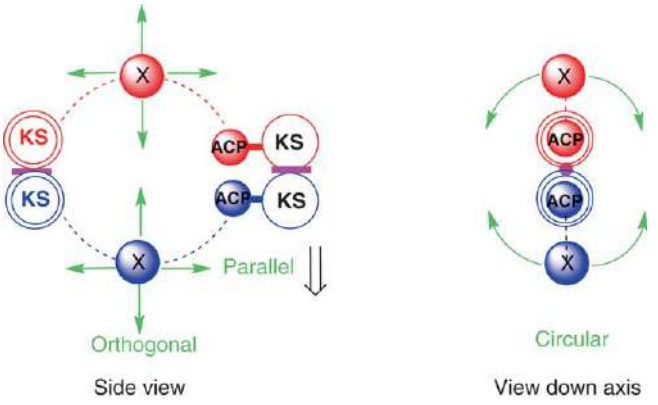
Scheme 40 Views down the axis of the modules with differing complements of reductive domains.

time the size of the loop has been cut down to three domains (AT, KR, and a spacer) in structure (**48**), the potential connectivity is no longer available, without blocking the passage of the ACP, unless the spacer adopts a more extended conformation as shown in (**49**). The spacer can now be seen as a potentially crucial contributor to the stability of the topology. Significantly, both the spacer and the dead KR are normally retained rather than being deleted in modules that have no reductive activity, whereas dead DH and ER domains tend to be deleted, consistent with the need to retain a sufficient size of reductive loop for effective operation of the module.[24]

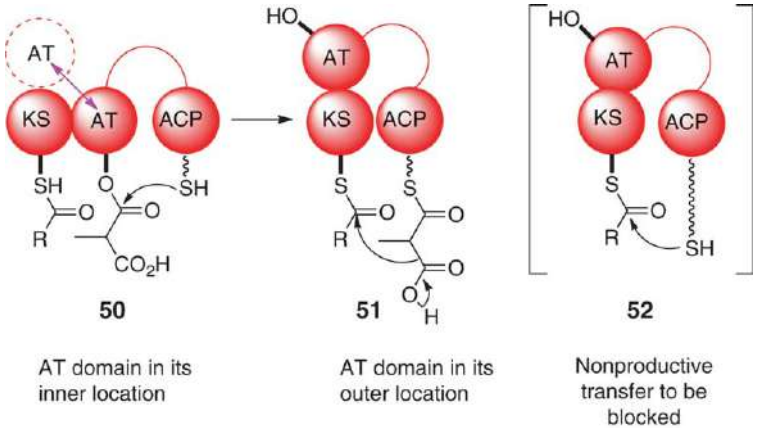
Another consideration that guided thinking about the Cambridge topology was the ready proteolysis of both upstream and downstream linkers to the AT domains.[69,102,103] Both linkers must be capable of existing in extended form in solution, allowing them to be accessed by a relatively large proteolysis enzyme, even though the recent X-ray image of the KS–AT domain shows the KS–AT linkage tightly folded.[113] Extension of both AT linkers implies that the AT domains may be capable of moving with respect to their neighboring domains. The diagrams in **Scheme 41** show that there are several modes of movement that can take place without compromising the fundamental twofold axis of symmetry.

A possible advantage of such mobility of the AT domains is illustrated in **Scheme 42**, which shows two active quaternary conformations. The first (**50**) would be adopted for interaction of the ACP with the AT domain to facilitate the addition of methylmalonate to the ACP; the second conformation (**51**) would be adopted when the ACP interacts with the KS domain in the condensation step.

In conformation (**50**), the AT lies in an inner position close to the axis, where it blocks access of the free SH group of the ACP to the KS. This would prevent a potentially damaging acyl transfer from the KS to the ACP, shown in (**52**), which could lead to the synthetic operations of that module being skipped. Once the ACP has been loaded with methylmalonate in conformation (**50**), the AT domain can safely move out of the way allowing the loaded ACP to approach the KS in conformation (**51**). This latter conformation is the one shown in cartoons (**44**) and (**45**) for the Cambridge topology.



Scheme 41 Allowed complementary movements of domains in a PKS module with preservation of the twofold axis of symmetry.



Scheme 42 Proposed movements of the AT domain in a DEBS module between inner and outer locations. The simple curved line represents a general reductive loop.

In effect, the proposed movement of the AT domains allows them to act as a gating mechanism to maintain fidelity of operations.

With this potential gating effect in mind, the Cambridge group adopted the conformation (**50**) in drawing the original ribbon cartoon for the Cambridge topology,[103] and in designing the cartoons now widely used to represent PKS assembly lines.

In conclusion, the Cambridge double helical topology is consistent with all the available experimental evidence for the DEBS modules, including the evidence for dimerization of KS and TE domains in solution,[103] the mutant complementation studies,[104,105] and the NMR investigations of the docking domains.[106,109,110] It is also consistent with the X-ray structure for the KS–AT didomain,[113] apart from one key feature: the original form of the helix was arbitrarily shown twisting one way around the axis, whereas the X-ray image shows the predicted twist around the axis between the KS pair and AT pair, but in the opposite direction. That is the reason for the important revision of the chirality of the proposed double helix.

As things stand, even with the benefit of X-ray images of fragments of modules we are forced to speculate on the working structure of the PKS module. Increasingly, experts in the field are coming to the view that the structure may be highly mobile so that multiple images will be necessary to capture the many different conformations adopted by a module in the course of a synthetic cycle.[15,45,98,99] The ACP domain certainly makes considerable changes of location. It is possible that various catalytic domains, including the DH and ER domains as well as the ACP, also undergo movements between inner and outer locations. Significant mobility of domains to explain solution behavior was also mooted by Ban and coworkers in their latest proposals for the solution behavior of the animal FAS.[88]

Since the completion of the manuscript, a single-particle EM study[117] has been published demonstrating that the substrate-loading and condensation domains of rat FAS make huge movements from side to side and swivel (±80–100°) in order to access substrates within each reaction chamber. Smith has shown that a heterodimer comprising a wild-type subunit partnered with a mutant lacking all seven functionalities is functionally active[15] and this would require a full 180° swivel during each catalytic cycle. Clearly, these structures are highly mobile with many conformations. The challenge of explaining how they move to achieve necessary contacts is even greater with modular PKS.

1.10.9 Other Polyketide Synthases

Although the ery PKS (polyketide synthase) is the most studied type I modular PKS, many other clusters have been identified, some with NRPS modules (Chapter 5.19), others lacking integral AT domains (the 'AT-less').[4–9] A small set of examples has been selected to illustrate some of the special variations that are possible on the DEBS paradigm.

1.10.9.1 Variation in the Packaging of Modules into Multienzymes

An outline of the packaging of modules into multienzymes for four PKS assembly lines is given in **Scheme 43**, where modules are represented by simple boxes. The proportions of these boxes correspond to the prediction of the Cambridge double helical model. Each box is a side view of a cylinder with a diameter and length of approximately 100 Å. The small bulge at the front of each cylinder represents the upstream face of the KS domain. A very similar picture could be produced for the topology proposed based on the X-ray image of the animal FAS, but with greater variation in length from module to module.

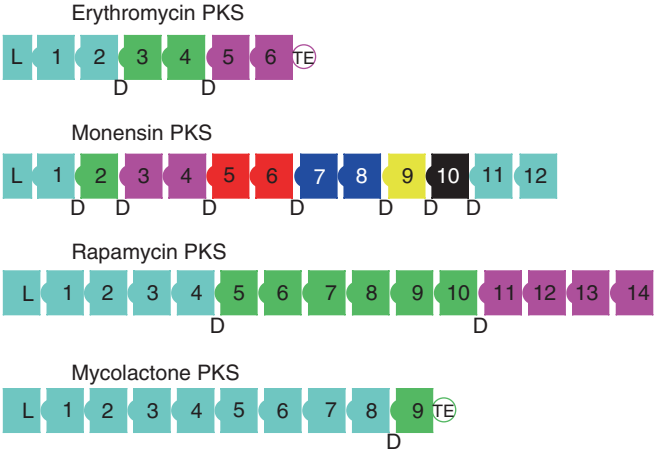
The housing of the modules into multienzymes is indicated by coloring of the modules and by a letter D to show positions of docking. Instead of the simple 'two-by-two' arrangement of modules in multienzymes found in the DEBS,[40–43] there is an extraordinary diversity of packaging found in these different PKS systems. At one extreme is the monensin (mon) PKS,[118,119] there are multienzymes containing single modules, as well as others with two. At the other extreme, the rapamycin (rap) PKS[120,121] has very large multienzymes, containing sets of four, and six linked modules.

The current record for length of multienzyme goes to the mycolactone PKS,[122] which has one with eight modules.

The length of the DEBS multienzymes is about 200–330 Å, which would mean that the full DEBS would be around 700 Å, as long as the multienzymes are permanently docked during synthesis. The largest of the rap multienzymes, would be 600 Å, and the total docked length about 1500 Å. These are indeed megasynthases.

These observations raise an interesting question: why does the length of the multienzymes vary so widely? The answer may lie in the extent to which these systems have evolved from their original evolutionary precursors. There is a general correlation between the size of the multienzymes and the degree of homology within particular types of domain in different modules of each PKS.[123] In rap,[120,121] the degree of homology is relatively high, suggesting that it has evolved to a relatively small extent from its original structure; in contrast, the degree of homology within the mon PKS[118,119] modules is relatively low. Astonishingly, the degree of homology for the KS domains of mycolactone is more than 99%. This seems to be a very primitive system.

It is tempting to speculate that splitting of larger primitive multienzymes into smaller ones is an important aspect of evolution. It has been suggested that the degree of homology between similar domains in the mycolactone PKS[122] is so high (99%), the explanation for the maintenance of specificity and fidelity of the synthetic operations in the varied modules cannot lie in kinetic effects determined by the tertiary structure of the domains. The alternative explanation is that the exit process is very slow, which means that each set of chain-extension domains has ample time to complete all the keto-group processing steps, before passing on the intermediate to the next module. We suggest the term, residence-time control, for this effect, to distinguish it from occupancy-level control, which is discussed in the later sections.



Scheme 43 Scale diagram of the erythromycin,[40–43] monensin,[118,119] rapamycin,[120,121] and mycolactone[122] PKS systems. Modules are numbered, and their housing in multienzymes is shown in color. Docking interfaces are marked by a D.

The general trend for PKS systems to show increased productivity as the system gets broken up into smaller multienzymes, is consistent with this suggestion.[124] Similar to domains housed in different modules evolve, they will be better able to control fidelity through kinetic effects rather than residence-time control.

Another feature of **Scheme 43** is the absence of an integral TE domain at the terminus of the rap and mon PKS systems. This and other important variations on the DEBS paradigm will be explored with specific examples in the following sections.

1.10.9.2 The Monensin PKS

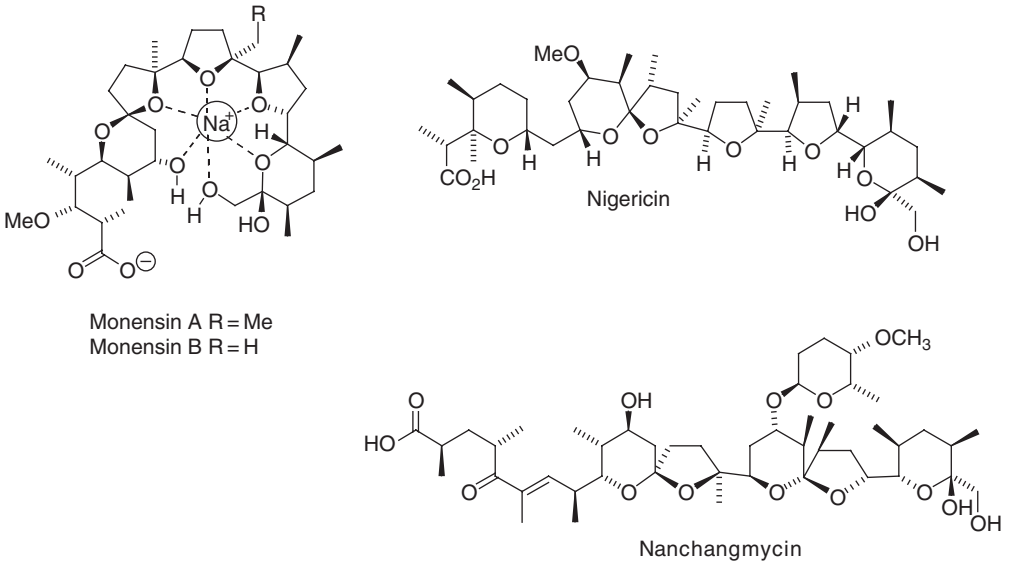
Mon is the archetypal example of a major class of PKs, the polyethers, that use a type I noniterative multifunctional PKSs for their biosynthesis.[125] These PKs are noted for their ability to transport ions across biological membranes[126] and their structures are exemplified by numerous ether rings, for example, mon, nigericin, nanchangmycin (**Scheme 44**). Gene clusters for several polyethers have been reported;[127–132] only the gene cluster for mon will be discussed in detail here.

1.10.9.2.1 Origin of the core structure of monensin

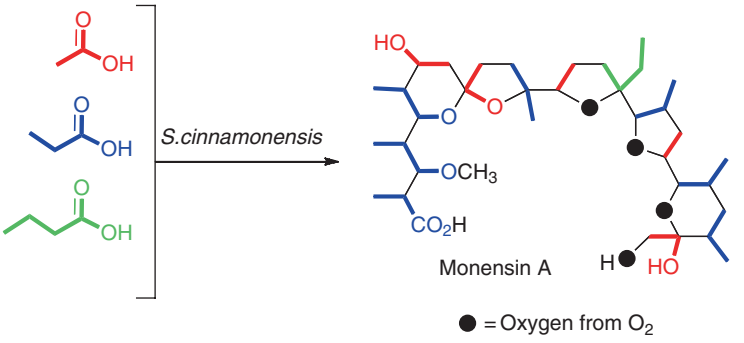
Streptomyces cinnamonensis produces mon A and mon B that differ only in the side chain at C-16 (ethyl/methyl) as indicated in **Scheme 42**. Classical feeding experiments to elucidate its biosynthesis showed that mon A is derived from five acetate, seven propionate, and one butyrate with four of the nine oxygen atoms derived from molecular oxygen (**Scheme 45**).[133,134] Mon B incorporates an extra propionate unit in place of the butyrate unit using ethylmalonyl CoA as the extender unit.

On the basis of these results, Cane *et al.*[135] proposed that a linear *E,E,E*-triene precursor might be produced by the mon PKS. According to their proposal, subsequent epoxidation and cyclization reactions would yield the final polyether structure. The publication of the mon gene cluster (**Scheme 46**) by Leadlay and coworkers[118,119] almost 20 years later revealed a type I PKS comprising 12 modules in eight contiguous reading frames is consistent with the proposed production of a linear triene (structure (**53**) in **Scheme 47**).

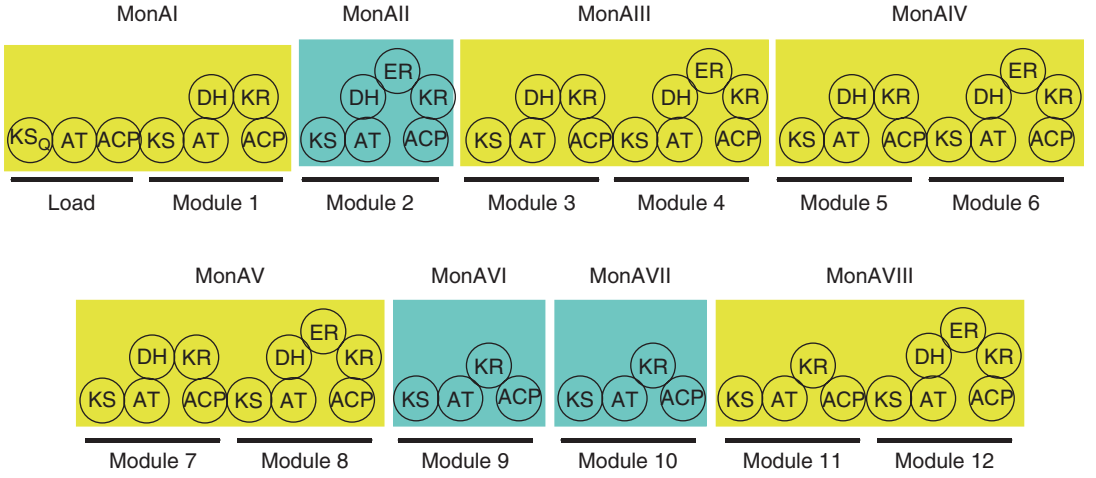
Unlike the ery PKS, there is no TE located at the C-terminus of the mon PKS. An unusual KS-like protein MonKSX transfers the triene from the PKS to a discrete ACP, MonACPX, where it is tethered, while the tailoring enzymes complete their operations.[127] Hence, instead of a sequence of enzyme-free intermediates envisaged by Cane and coworkers,[135] it seems that all the epoxidation and cyclization steps occur with the acyl chain attached to ACPX, before release of the final product by a discrete type II TE,[136] as indicated in **Scheme 47**.



Scheme 44 Polyether polyketides.



Scheme 45 Origin of the carbon and oxygen atoms in monensin A.[133,134]

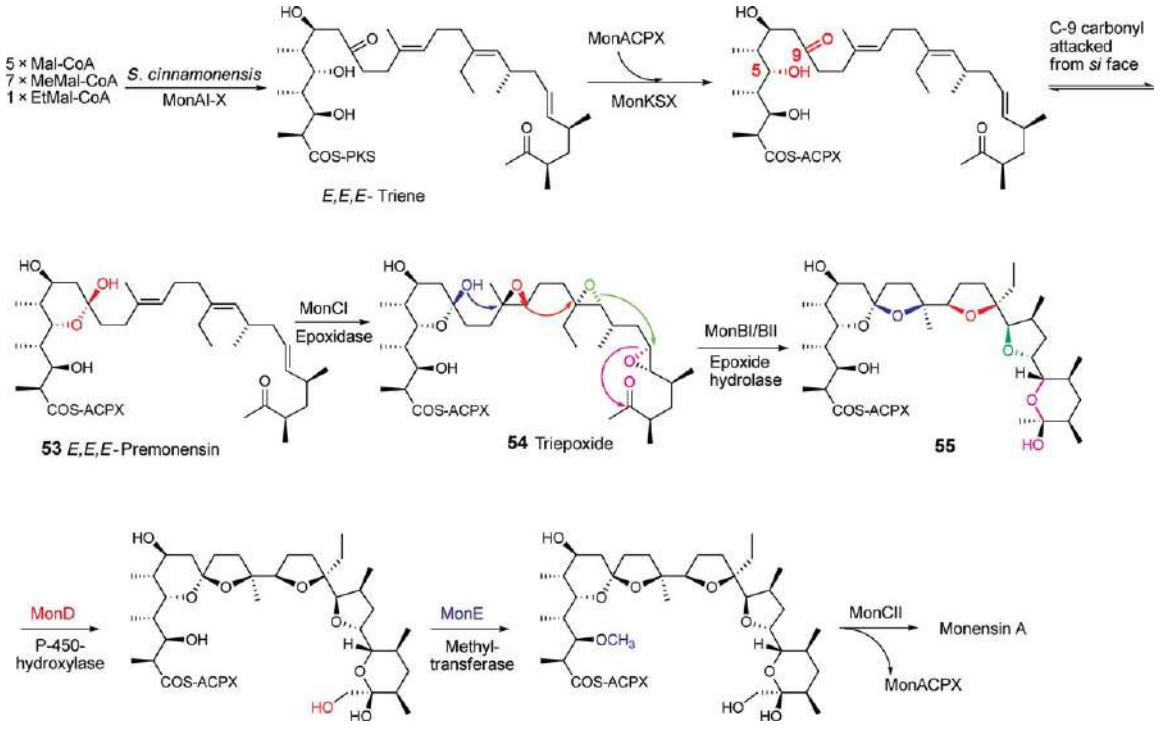


Scheme 46 Organization of the monensin PKS multienzymes; modules colored yellow contain two modules, those colored blue contain only one.

According to the original proposal,[135] the conversion of the triepoxide into the cyclized polyether product should take place by a concerted series of nucleophilic displacements, which involved two S_N2 reactions at tertiary centers. In a more thoughtful analysis of the mechanistic operations, Spencer advanced a multistep sequence of reactions, which recognized the mechanistic challenges of this multistep operation.

The spiroketal ring is formed first from the nucleophilic attack of the C-5 hydroxyl from the *si* face of the C-9 carbonyl group. Epoxidation of the *E,E,E*-premonensin (**53**)[137] catalyzed by MonCI gives the triepoxide intermediate (**54**) attached to MonACPX. The ring closure is catalyzed by MonBI/BII:[138] these enzymes are responsible for the stepwise opening of the epoxide groups and subsequent polyether ring formation. MonBI/BII accelerate, but do not change, the stereochemical outcome of this process.

Cyclization of the triepoxide (**54**) commences with the nucleophilic attack of the hemiacetyl hydroxyl group on the first epoxide to form the B ring; attack is at the more hindered carbon and goes with inversion of stereochemistry at this center. An S_N2 mechanism is not possible in this case due to the sterically hindered nature of the epoxide and an S_N1 mechanism does not explain the observed stereochemical outcome. Instead, a 'loose S_N2' mechanism, reminiscent of bromonium ion chemistry, has been proposed. Acid catalysis facilitates the opening of the epoxide in the transition state thereby permitting the developing carbocation to be attacked by the incoming hydroxyl nucleophile. Attack at the less hindered carbon would give the 6-*endo-tet* cyclized product that is kinetically disfavored using Baldwin's rules.[139] The same principles can be applied to the formation of the C-ring, and closure of the D-ring, which involves a favorable *exo-tet* S_N2 reaction. The hemiacetal ring closure is fully reversible and the natural configuration is the thermodynamically favored



Scheme 47 Late stages of monensin biosynthesis.

product (**55**). The final steps in mon biosynthesis are hydroxylation, O-methylation,[140] and cleavage of the ACPX by the TE MonCII.[138]

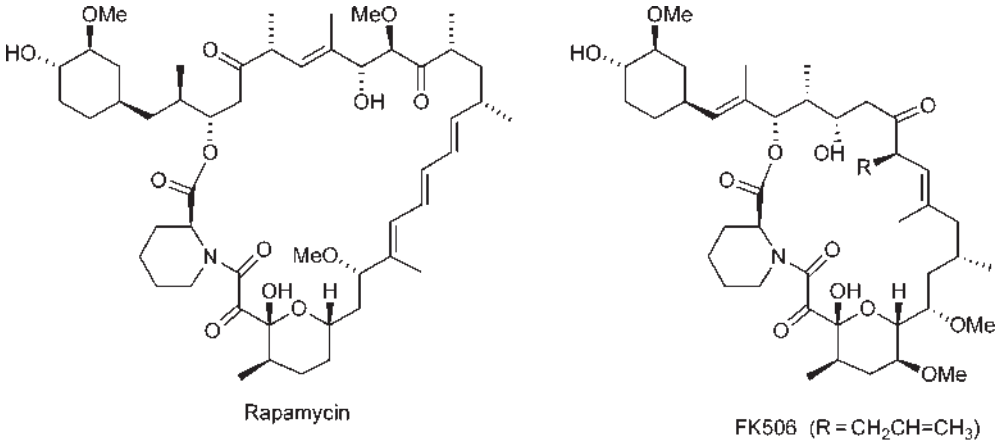
Another interesting feature of the mon PKS, shown in **Scheme 46**, is the organization of the loading domain. Compared with the DEBS, this contains an additional domain, KS_Q, which is a KS domain with the active site cysteine mutated to a glutamine.[141,142] This change destroys the ability of the KS_Q domain to bind an acyl chain as a thioester, but it leaves its ability to decarboxylate a malonate unit to the enolate, followed by protonation to give the corresponding saturated acyl unit bound to the ACP of the loading domain. This alternative process for providing the starter unit acyl group is much more selective, because the malonate-specific AT domain in this LM has a much higher selectivity toward its substrate than the acyl-specific AT used by the ery LM. LMs based on the KS_Q strategy are commonly found in PKS systems.[142]

In conclusion, the mon PKS follows closely on the ery PKS in its chain-extension steps, but it uses very different strategies to initiate chain formation and in the release of the completed chain at the end of chain assembly.

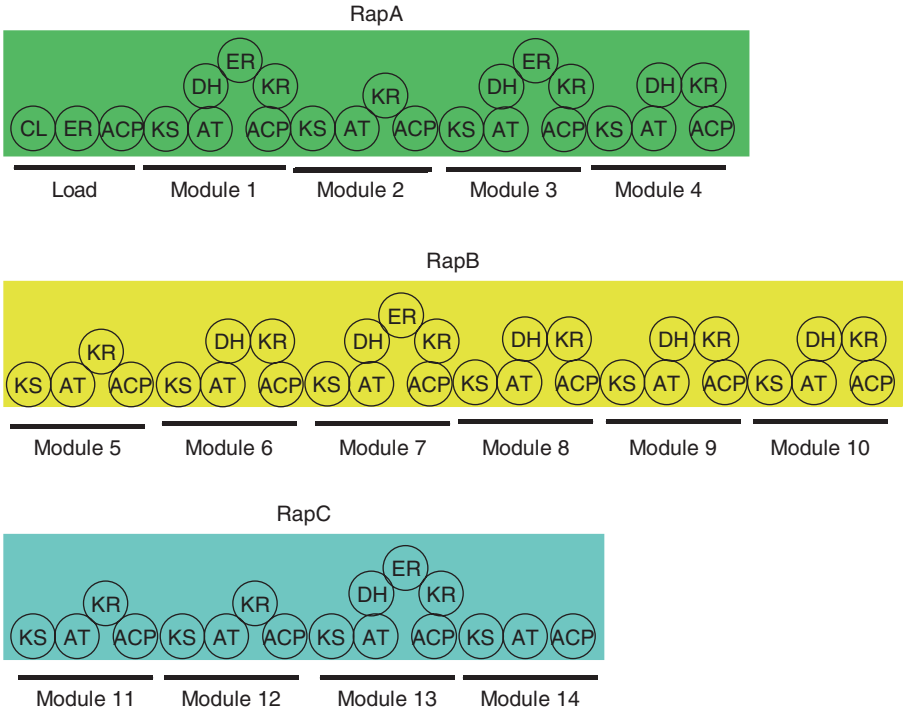
1.10.9.3 The Rapamycin PKS

Rap and the structurally related compounds FK506 and FK520 (**Scheme 48**) are PKs that exhibit antitumor, antifungal, and immunosuppressant activities. All three compounds have a common structural motif, which is responsible for their specific binding to FK506-binding proteins (FKBPs). Their biological effects arise from their subsequent binding of the PK/protein complex to a downstream protein target, mTOR (mammalian target of rapamycin) or calcineurin (FK506, FK520).[143,144]

Rap is biosynthesized on a mixed type I PKS/NRPS.[120,121] Three proteins, RapA (~900 kDa, 4 modules), RapB (1.07 MDa, 6 modules), and RapC (660 kDa, four modules) (**Scheme 49**) are responsible for loading the starter unit (3,4-dihydroxycyclohexanecarboxylic acid (DHCHC))[145] and elongating it with eight malonyl- and six methylmalonyl-CoA units[146] to give the putative tetradecaketide (**56**) (**Scheme 50**). In modules 3 and 6, there are possibly active but unused KR, DH, and ER domains.[120,121] This small discrepancy between the sequence of domains and the required set of reactions is a common feature in modular PKS systems.[4]



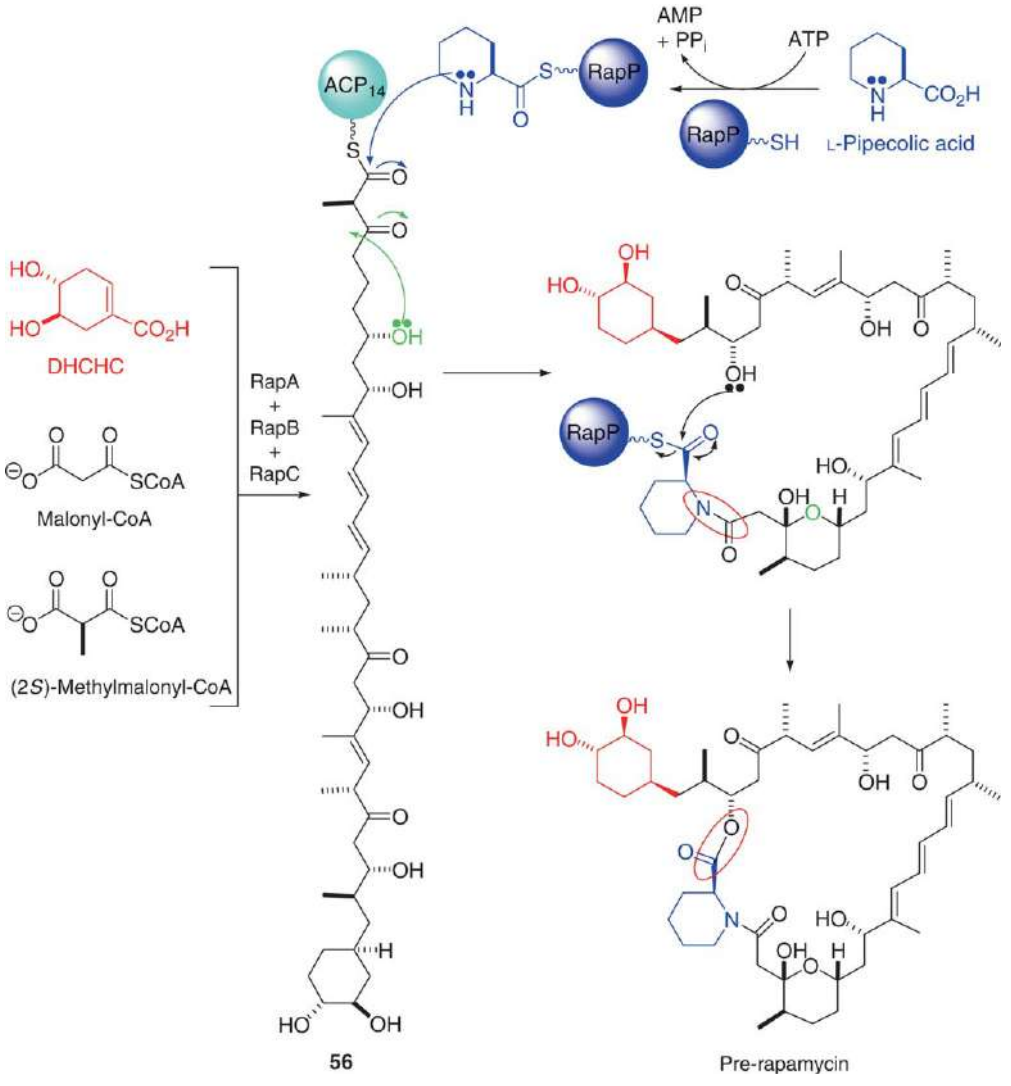
Scheme 48 Structures of rapamycin and some related compounds.



Scheme 49 Organization of the rapamycin PKS.

There is no TE activity in the rap PKS. Instead, a nonribosomal peptide synthase (NRPS) RapP[147,148] incorporates an L-lysine derived L-pipecolic acid moiety into the chain:[149,150] two new bonds are made, an ester and an amide bond, to give the first enzyme-free intermediate pre-rap (**Scheme 50**).[151] It is thought that attack of the RapP-bound pipecolate amine moiety onto the thioester bond of (**56**) releases the PK chain from the rap PKS. The terminal hydroxyl group of the PK chain attacks the acyl residue of the pipecolate residue, derivatized as a thioester on the multienzyme, RapP, to release the macrolide ring as the first enzyme-free intermediate.[147,148]

Finally, five post-PKS modifications are required,[148] three O-methylations and two oxidations. It has been shown that these proceed in the sequence:[152] methylation of the C-39 hydroxyl by RapI; oxidation of C-9

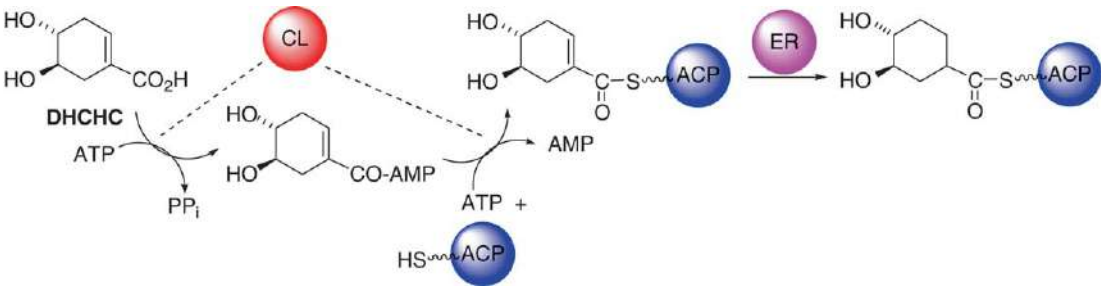


Scheme 50 Rapamycin biosynthesis.

which, unusually, is catalyzed by only one enzyme, RapJ; methylation of the C-16 hydroxyl group; oxidation at C-27 catalyzed by RapN, followed by methylation of the resulting hydroxyl moiety by RapQ.

The rap loading domain is very different from that found in DEBS. The CoA ligase (CL) domain activates the starter acid, DHCHC,[145] by attaching an adenosyl moiety (**Scheme 51**). The adenylated starter is then transferred to the ACP of the loading domain and is reduced by the adjacent ER domain to give the (1*R*, 3*R*, 4*R*)-DHCHC starter unit, ready for transfer to the KS1 domain of RapA, to initiate chain elongation. Biosynthesis of the DHCHC starter unit is dependent on RapK, which is thought to be involved in the regulation of its production and/or biosynthesis from shikimate.[153]

In conclusion, the rap PKS exhibits interesting variations from the DEBS. The much larger size of the multienzymes has already been noted. More significantly, it uses highly specialized strategies for both chain initiation and chain termination. Another important distinguishing feature is the structure of module 14, which cooperates with the specialized enzyme used in the off-loading process. Unlike DEBS module 3, which also forms a ketoester, this module has a simple direct link from the AT domain to the ACP, with no apparent intermediate domain.[120] It therefore must have a special quaternary structure.



Scheme 51 Proposed mechanism for the activation and loading of the DHCHC starter unit to the rapamycin PKS.[145]

1.10.9.4 The Rifamycin PKS

The ansamycins are a family of PK antibiotics comprising a naphthalene/benzene or napthaquinone/benzoquinone ring system bridged at nonadjacent positions by an aliphatic chain.[154,155] Examples include rifamycin (rif) and ansamitocin (asm) (**Scheme 52**).

Rif B is the major component of the rif mixture produced by *Amycolatopsis mediterranei*. Although it has a modest antibacterial activity, rif can be converted into the more potent and clinically used rif SV.

Floss and coworkers[156] and Schupp *et al.*[157] independently reported the *rif* gene cluster which contains the genes coding a type I modular PKS (**Scheme 53**) comprising five ORFs, *rifA–E*, with rifA fronted by an NRPS adenylation/thiolation loading didomain.

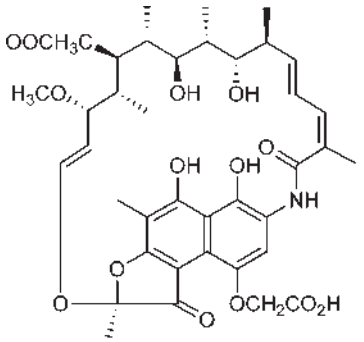
The set of genes in the first multienzyme, RifA, is appropriate for the synthesis of the tetraketide intermediate, P8/1-OG, by normal PKS chain-extension processes (**Scheme 54**).[158] At this point, the intermediates follow a very unusual course, catalyzed by an accessory protein, named Orf19.[159] An oxidative cyclization onto the ring of the starter acyl group generates a naphthalene residue, in intermediate (**57**). This is then transferred to module 4 of RifB, and further standard rounds of chain extension lead ultimately to the undecaketide product (**58**) bound to the ACP of module 10.

There is no TE function at the terminus of module 10; instead, another external protein, *RifF*, which encodes an amide synthase activity,[160,161] catalyzes the cyclization to the macrolactam triggering release of the undecaketide, proansamycin X.

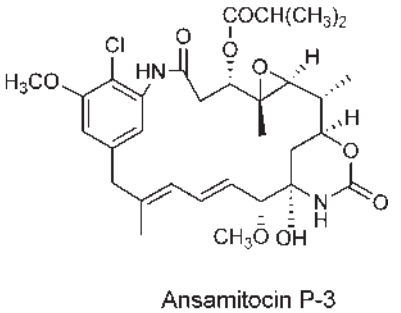
The remaining steps involve post-PKS oxidations and cleavage reactions to give rif B.[155]

1.10.9.5 The Mupirocin PKS, an AT-less System with Special Mechanisms for Generating C-1 Branch Points

Pseudomonas fluorescens produces the PK antibiotic mupirocin (mup) which is active against Gram-positive bacteria including methicillin-resistant *Staphylococcus aureus*. It is a mixture of pseudomonic acids, each of which

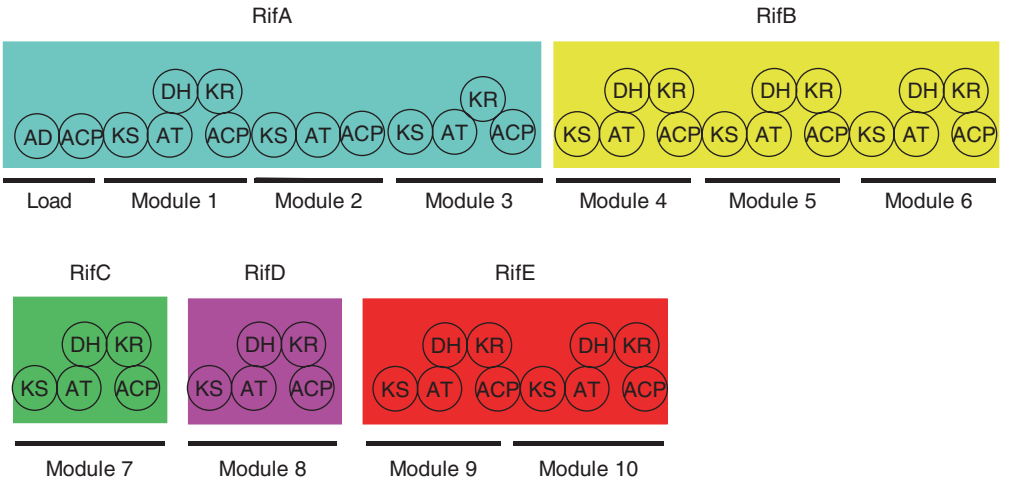


Rifamycin B



Ansamitocin P-3

Scheme 52 Representative ansamycin structures.



Scheme 53 Organization of the rifamycin PKS.

comprise a C_{17} monic acid (MA) and a C_9 9-hydroxynonanoic acid (9-HN) joined by an ester linkage. The principle component of the mixture is pseudomonic acid A, **Scheme 55**. The PKS responsible for synthesizing the MA component uses a different mechanism from the DEBS for generating C-1 centers on the PK chain.[162] In the scheme, C_1 units derived from SAM by methyl group transfer, are shown in blue. Other C_1 units derived from the methyl group of acetate, are shown in pink. Two other natural products, which rely on the same strategy, are pederin[163,164] and bacillaene.[165,166]

The multienzymes that cooperate in the assembly of the basic carbon skeleton of the MA component of mup are shown in **Scheme 56**.[162] The longer one, MmpD, contains four modules with a distribution of reductive domains consistent with the formation of the putative pentaketide intermediate. The second, MmpA, has three modules consistent with the next three cycles of chain extension leading to a putative octaketide.

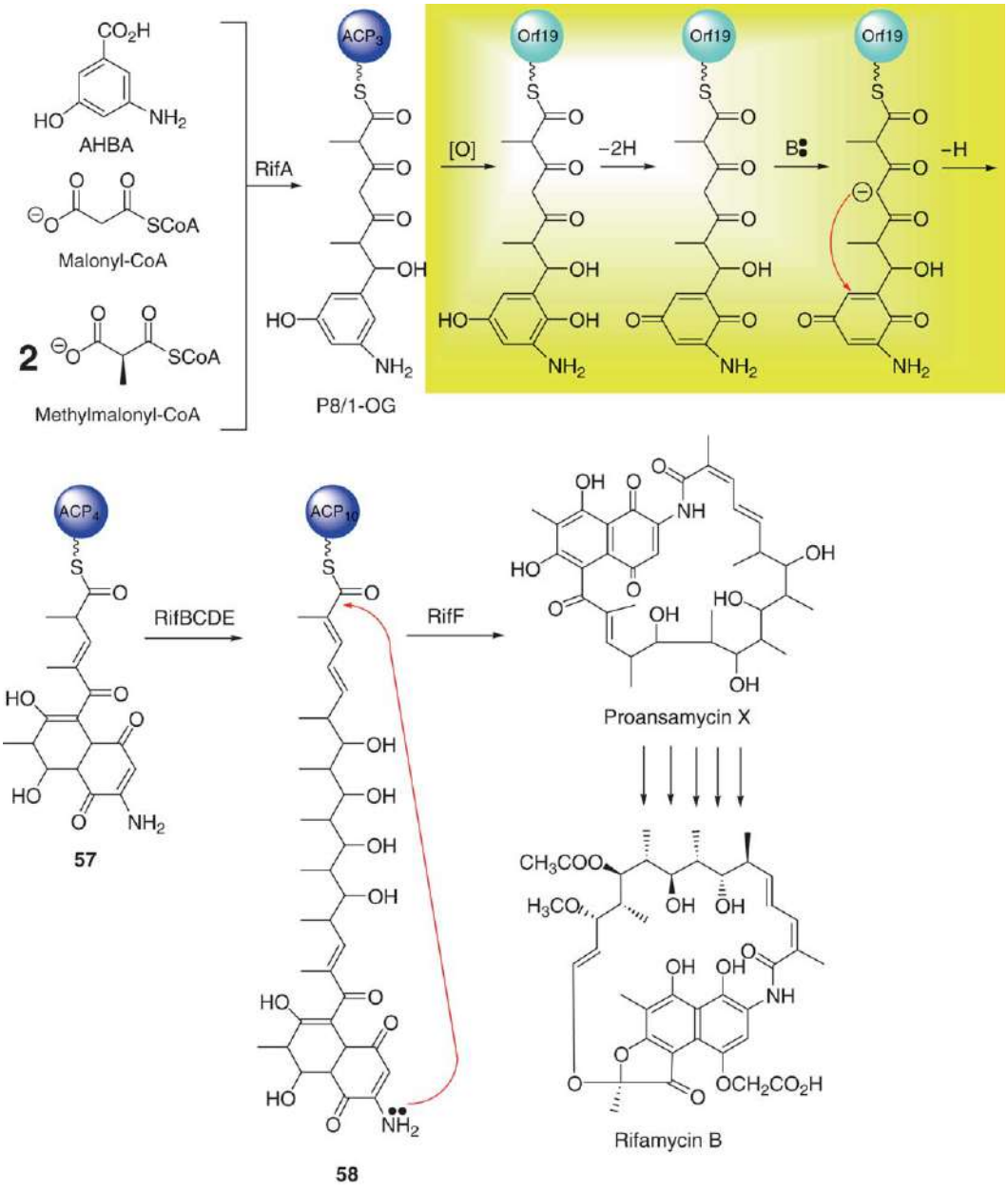
However, closer inspection reveals striking differences between the mup and ery PKS modules. First, the DEBS modules have inbuilt AT domains, whereas the mup modules are AT-less. Second, modules 1 and 3 contain an MT (methyl transferase) domain associated with the reductive domains. Finally, module 6 is unusual in having two ACP adjacent domains[167] and no TE domain.

The necessary AT activity is provided by a separate didomain, MmpC, consisting of a pair of AT domains.[162] By working *in trans* with each module, MmpC can deliver a unit of acetate or malonate to the two ACP domains of the presumed homodimeric modules by the usual chemical operations. It is interesting to note that this mode of operation is reminiscent of that proposed earlier for the DEBS modules, even though the DEBS AT domains are covalently bound to the module. In both the DEBS and the AT-less systems, the pair of AT domains move in and out with respect to the core of the module at different stages of the chain-extension cycle.

Malonate is the only chain extender used for all the chain extensions. This explains the presence of the MT domains in modules 1 and 3. These catalyze the transfer of a methyl from SAM onto the methylene of the β-ketoester intermediate. This provides an alternative to the use of methylmalonate as the chain extender for branched units in the DEBS. It is the standard method in fungal PKSs.[168]

Mup biosynthesis (**Scheme 57**) begins with an acetyl-CoA starter unit[169] being loaded onto the first module located on MmpD. Chain extension with four malonyl-CoA extender units loaded onto the PKS by the AT didomain gives the pentaketide (**59**). This is transferred onto module 5 on MmpA so that another two cycles of chain extension can take place. The 'loading/transfer domain' preceding module 5 may facilitate this chain transfer process.[162]

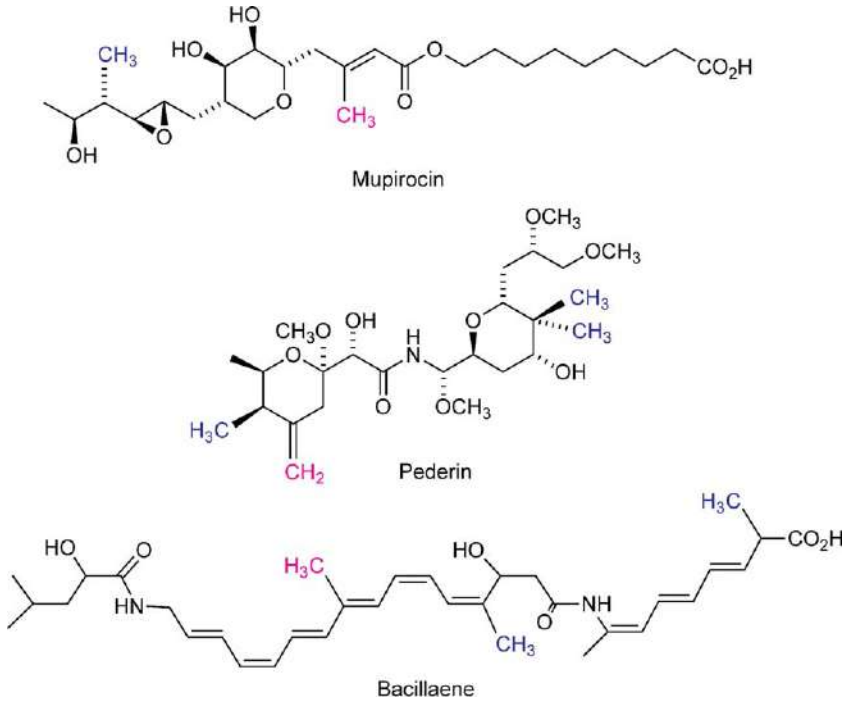
Once the heptaketide (**60**) has been made, the ketoester undergoes a further condensation to add an acetate unit but not by the usual process in which the terminal acyl thioester is attacked. Instead, the condensation, which recalls a similar step in mevalonic acid biosynthesis, takes place at the ketone group to give (**61**). The process is carried out by a separate suite of enzymes, called the HMGS (3-hydroxy-3-methylglutaryl-CoA synthase) cassette,[6] to emphasize the relationship to the equivalent step in the mevalonic acid biosynthetic



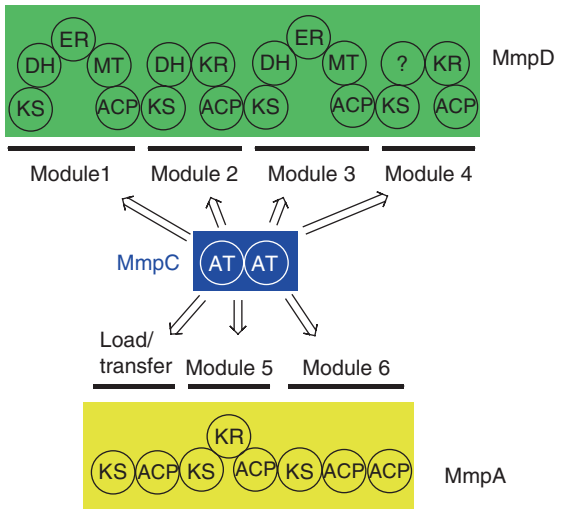
Scheme 54 Rifamycin biosynthesis.[152] Aromatic ring closure is catalyzed by Orf19 and does not occur on the PKS; the reactions are shown in the yellow box. The remaining steps involve post-PKS oxidations and cleavage reactions to give rifamycin B.[155]

pathway. The process starts with the generation of acetyl-ACP by an AT-catalyzed loading of Mal CoA (possibly by the AT didomain, MmpC) onto a discrete ACP followed by decarboxylation by a KS. Subsequent condensation onto the β-ketone catalyzed by an HMGS and dehydration and decarboxylation by enoyl CoA hydratase-like (ECH) domains gives an α-methyl branch as shown in (**62**).[162]

To transform this intermediate into the final MA requires a series of posttranslational modifications including reduction of the 8,9-alkene, hydroxylation at C6, epoxidation and esterification with the 9-HN residue.[162] Some catalytic functions have been identified by mutational studies carried out by Simpson and coworkers,[170–174] but the details of this sequence of late-stage reactions are unclear. One of the difficulties lies in the interpretation of

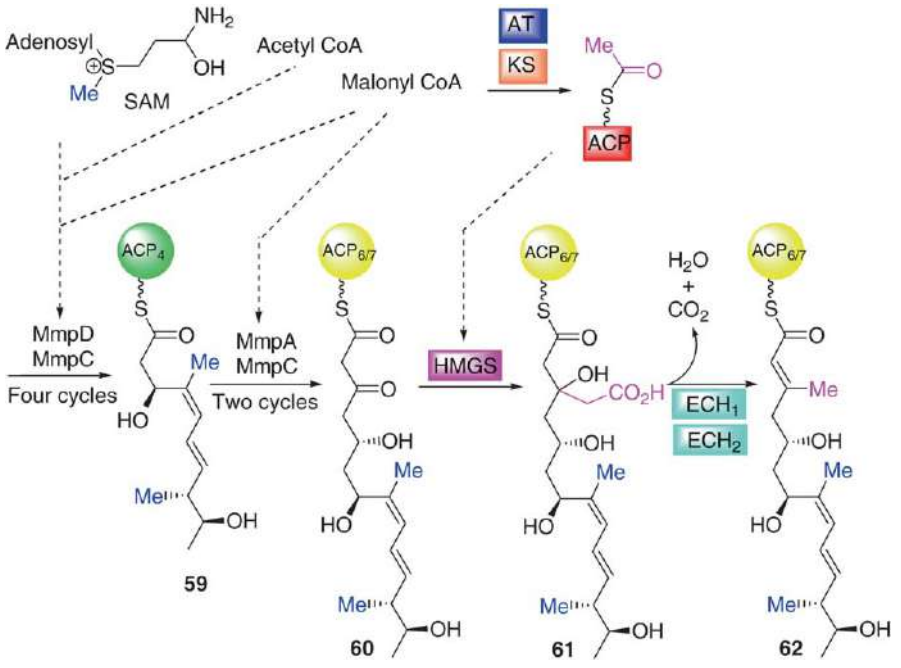


Scheme 55 Representative AT-less polyketides. Mupirocin is a mixture of pseudomonic acids: the structure of pseudomonic acid A is shown. The methyl groups inserted by embedded methyltransferase domains are shown in blue and the C_1 units added by the HMGS are pink.



Scheme 56 Organization of the AT-less modules in the PKS responsible for biosynthesizing the monic acid portion of pseudomonic acid A.[162]

the results of experiments in which key late-stage enzymes were deleted. Wherever the deletion was made, the same stage intermediates accumulated. Simpson suggested an interesting explanation, illustrated by the concept of a leaky hose.[174] When the outlet of a hose is blocked, pressure within increases, and water escapes at leaky points wherever they may be. It was suggested that with a late-stage blockage, the PKS assembly line becomes loaded with intermediates throughout the earlier steps, so the intermediates with more labile links to the synthase can be released into the medium. Simpson also noted, however, that some of the released intermediates



Scheme 57 Key intermediates in the chain-extension reactions carried out by the mupirocin PKS. The HMGS cassette enzymes are boxed.

may not accumulate, if they are susceptible to degradation by the very active fatty acid degradation enzymes, present in the producer cells. It can be unsafe, therefore, to draw conclusions about the mechanism or specificity of constituent enzymes from the pattern of metabolites isolated from incubations of mutant organisms.

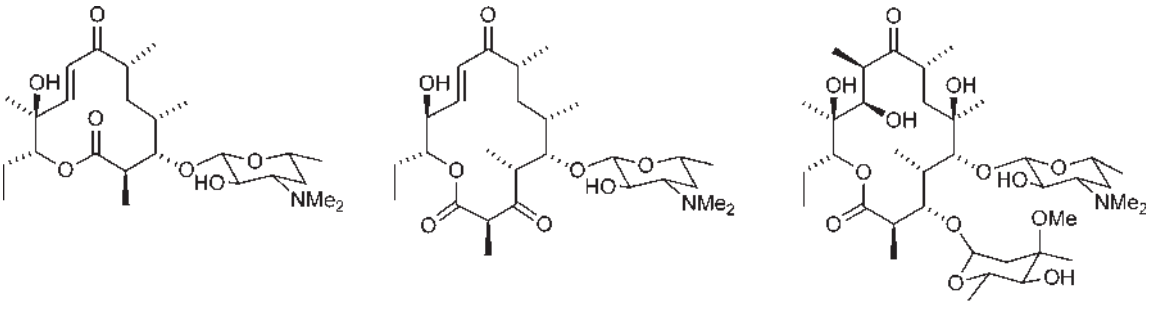
In conclusion, there are several special features of the chain-extension processes operating in the mup PKS. The special arrangement for loading methylmalonate units and branching methyls has already been mentioned. The use of the HMGS strategy for adding an extra branching structure has been found in 11 pathways to date.[6] It introduces another methyl branch after the modification steps, but note that this branch is at a site that originates from the carboxyl group of the original acetate building block, not the methyl. This branch is therefore out of step with the usual pattern of methyl substitution in the PK chains. This pathway also illustrates the extraordinary versatility of synthetic strategy that can be incorporated into the basic fatty acid chain-extension chemistry employed by simpler PKS assembly lines, such as the DEBS.

1.10.9.6 Methymycin and Pikromycin

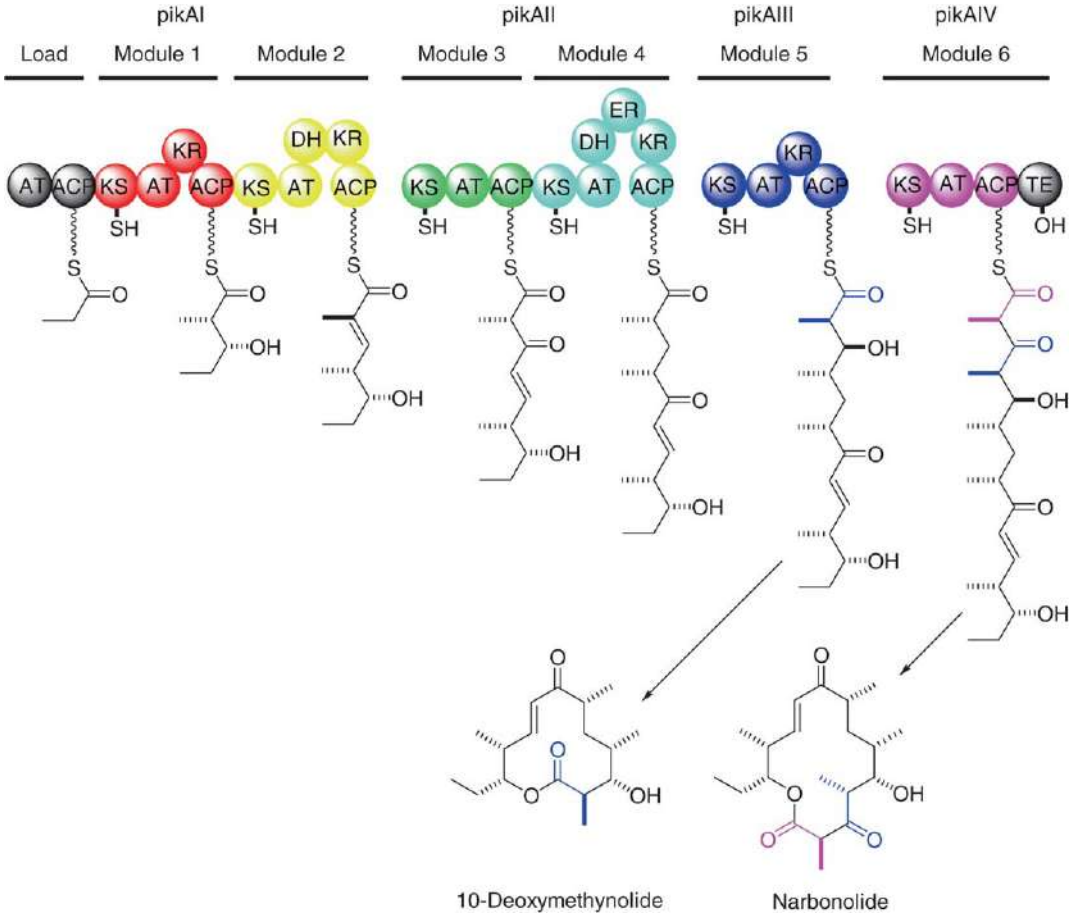
Methymycin and pikromycin (pik) (**Scheme 58**) are two macrolide antibiotics related to ery. The 12- and 14-membered macrolide rings of methymycin and pik are both produced by the *pik* gene cluster in *Streptomyces venezuelae* ATCC15439;[175] the TE domain is able to catalyze the cyclization of both the linear hexaketides and heptaketides.[176]

The modular PKS responsible for chain assembly of the 14-membered macrolide core of pik closely resembles the DEBS (**Scheme 59**).[175] The differences lie in module 2, which forms a double-bond intermediate, and modules 5 and 6, which exist as separate multienzymes, pikAIII and pikAIV.

The formation of narbonolide is unexceptional and needs no comment, but the release of a 12-membered lactone ring by the same organism is very puzzling. Sherman and coworkers explored the reasonable idea that one of the external type II TE enzymes found in the pik cluster might interrupt the chain-extension process at the terminus of module 5 (pikAIII) by releasing the hexaketide intermediate as 10-deoxymethynolide.[177] There was no support for this explanation, however, in experiments to delete possible candidates for the role of off-loader. Perhaps this outcome is not a surprise given that the predicted role of type II TE enzymes is to edit aberrant intermediates from PKS assembly lines.[178] For them to be effective in this rôle, they should ideally act



Scheme 58 Structures of macrolide antibiotics.



Scheme 59 The pikromycin PKS and putative intermediates leading to the release of the 12-membered macrolide 10-deoxymethynolide, the precursor of methymycin, and the 14-membered macrolide, narbonolide, the precursor of pikromycin.

as hydrolases not cyclases, so that released products are linear, and are therefore easily digested by degradation enzymes. In retrospect, it is a pity that the lactone-forming domains were not given a distinguishing title, such as lactone cyclases (LC), rather than TE, to emphasize the special nature of their role.

Sherman then made the very perceptive observation that the gene for module 6 had a potential alternative start site for expression that would lead to the formation of a truncated form of module 6 with a section of

N-terminal residues missing.[177] It was suggested that the resulting defective KS would be able to accept the product of module 5, but that it would not be able to carry out the normal condensation step. The way forward is then to bypass the operations of the terminal module, a process known as skipping, by transferring the hexaketide directly from the defective KS to the ACP and then on to the TE for normal release.[179] An alternative suggestion was the direct ACP_5–ACP_6 transfer which would result in skipping the KS_6 altogether.[180] It has been subsequently shown *in vitro* that full-length PikAIV can produce 10-deoxymethynolide when paired with PikAIII,[181] which is not consistent with the truncation model.

The latest *in vitro* experiments have provided some insight into how both 12- and 14-membered rings can be made by the same PKS.[182] Point mutations were made in each of the catalytic domains of PikAIV and all mutants except the TE mutant were able to produce 10-deoxymethynolide suggesting that only the TE domain is required for the lactonization step. Next, the intermodular docking domains of PikAIII and PikAIV were mutated, complemented with their respective wild-type module and incubated with a synthetic pentaketide NAC thioester substrate. Neither mutant could make narbonolide but both could make 10-deoxymethynolide albeit 30–50 times less efficiently than the wild type. It was also shown that recombinant TE could cyclize the PikAII-bound hexaketide intermediate although less efficiently than with a PikAIII–TE fusion protein.[181] Taken together these results suggest that 10-deoxymethynolide production requires PikAIV when docked to PikAIII to assume a conformation that permits the TE to interact with ACP_5 of PikAIII. The requirement for PKS modules to be flexible to enable ACP interaction with all necessary catalytic domains has already been discussed and so it does not seem unreasonable that the required conformation for 12-membered ring catalysis can be achieved to allow skipping of the final module.

The real puzzle is why the PKS adopts this conformation in the first place rather than chain extending the hexaketide intermediate and cyclizing the resultant heptaketide. Sherman has suggested that if the AT_6 domain is insufficiently active to load ACP_6 before PikAIV docks with PikAIII, then PikAIV adopts the conformation required for the biosynthesis of the 12-membered ring.[182] Evidence to support this notion comes from Cane and coworkers, who showed that in the absence of methylmalonyl-CoA, the PikAIV catalyzed macrolactonization of hexaketide intermediate was approximately twofold higher.[176] Sherman points out that the *in vitro* data are not always consistent with previous *in vivo* results,[177] and suggests that other factors such as proteases, translation start sites, and regulators that operate *in vivo* may lead to a disparity in macrolide product formation.[182]

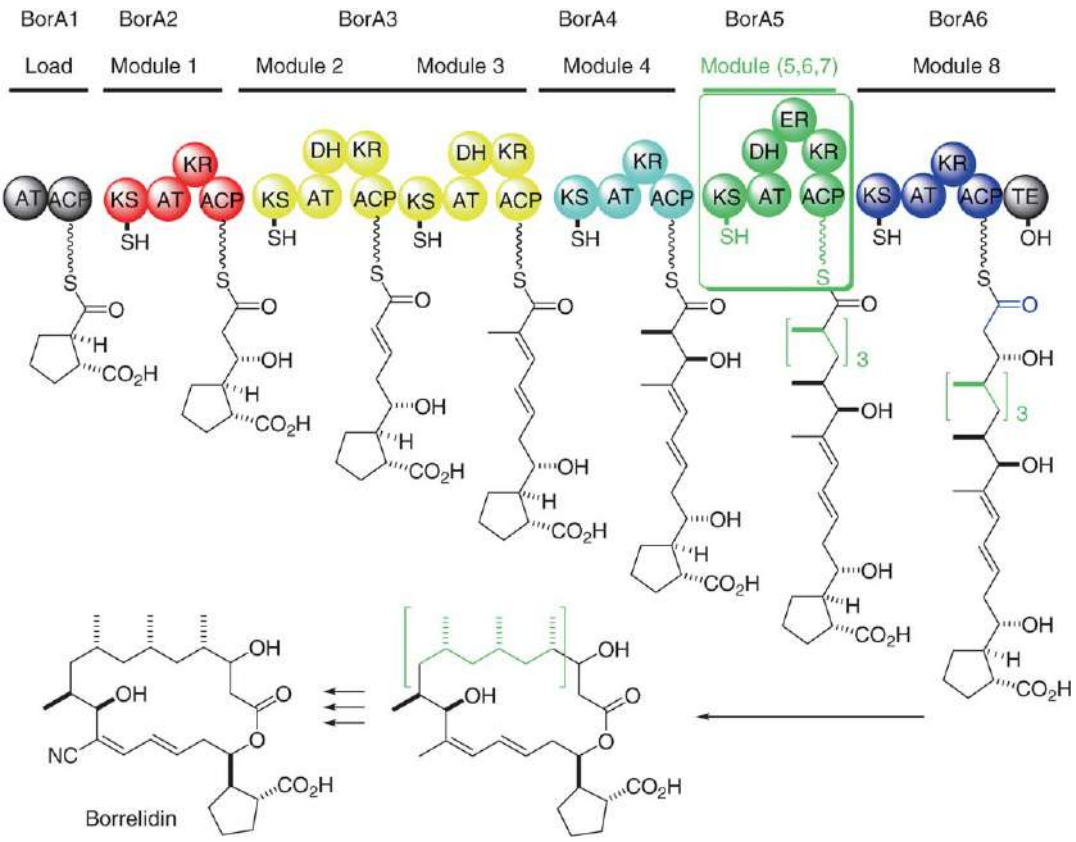
Whatever the explanation for the premature truncation of the chain-extension processes in this module, the process called skipping[183] provides yet another interesting way for increasing the potential range of chemical products of a PKS.

1.10.9.7 Borrelidin

Borrelidin (bor) is an 18-membered macrolide characterized by its unusual cyclopentane dicarboxylic starter acid and the nitrile functional group. It exhibits antibiotic and antitumor activity and is produced by a number of streptomycete species. The bor gene cluster has been cloned from *Streptomyces parvulus* Tü4055 and surprisingly only six chain-extension modules, instead of eight, was found in addition to the LM (**Scheme 60**).[184] To account for this discrepancy, it was suggested that the fifth chain-extension module (BorA5) undergoes three successive rounds of chain extension to introduce three identical chain-extension units before passing on the chain to the next module.[185]

This repetitive mode of operation, iteration, is the standard mode of operation for an animal FAS,[15] and for fungal PKS systems,[168] but it is only rarely observed in modular PKS assembly lines.[186] That brings us to one of the major puzzles in modular PKS operation: how is the iterative mode almost invariably prevented in PKS modules, and how is it promoted in the few modules that operate in that manner? Other examples include stigmatellin,[187] aureothin,[188,189] and to a very small extent, module 4 of the DEBS.[190]

For a module to iterate the upstream KS must be kept unoccupied, so that it can accept back transfer of the growing chain when reductive processing steps are complete. Hence, one possible explanation for the suppression of iteration is occupancy level of the KS domains. If, after the condensation step, every KS domain is rapidly reloaded with the product of the previous module, iteration will not be possible, when the reductive



Scheme 60 Organization of the borrelidin PKS.

processing of its ketoester intermediate is complete. There is only one way for the intermediate to go, and that is downstream when the downstream KS next becomes vacant.

This mode of operation can be described as occupancy-level control, and its operation in suppressing iteration can be compared with a dispenser of a beverage can. When a can is removed from the bottom position, each can above in succession falls down by one position, until the vacancy reaches the top and there is space to add a new can. In the context of a PKS assembly line, this implies that the control over the overall rate lies with the release of the final product by the exit mechanism, usually a TE domain. In this mode of operation, a much slower rate of throughput is the price to pay for prevention of iteration. The mechanism would work equally well with highly developed or primitive assembly lines.

Incidentally, occupancy-level control is clearly related to residence-time control, which was discussed earlier in relation to the maintenance of fidelity in the operation of reductive loops of primitive PKS modules. Both effects are achieved by slowing the rate of release of the final intermediate on the assembly line. The two effects have a fundamental difference, however, which justifies treating them as separate effects. Occupancy-level control is a spatial effect that reaches an effectively maximum extent once the KS domains are fully occupied (or fully unoccupied as we shall see below). Residence-time control is a temporal effect that will continue to increase after the occupancy level of the KS and ACP domains has reached 100%.

An alternative strategy for prevention of iteration, based on the internal structure of the KS domains, can also be considered. Weissman and Müller[45] recently proposed that there are two access channels to the KS domains of a PKS, one on the upstream face for downstream passage of the completed intermediate of the previous cycle, the other on the downstream face to facilitate access by the malonate building block, and release of the ketoester product of the condensation reaction. It could be that the KS domains of modules in a PKS assembly line are in some way structured to restrict acyl group transfer to the upstream access channel.

The next question is: How is iteration promoted in the case of iteration such as the bor module 5, 6, 7? Two obvious conditions have to be met: the KS in the iterating module has to be kept free and the downstream transfer needs to be restricted. The latter condition can be met by occupancy-level control, which keeps the downstream KS fully occupied. The former condition could be met if the cooperation with the upstream module is controlled by a docking mechanism, as is the case in the relevant bor module. As long as this docking operation has a low efficiency, passage of the upstream product to the iterating module can be restricted so that its KS domain is kept temporarily unoccupied for the iteration process. In effect therefore, occupancy-level control can operate in two senses. In the case of iteration both senses are helpful at different sites.

For bor biosynthesis, the number of cycles of iteration has also to be controlled to three. This number will presumably be determined in part by the substrate specificity of the iterating KS, as is the case in the animal FAS. It can accept the product of two iterations but does not accept the product of three. Iteration is a fascinating variation on module behavior. Even though it is rarely observed in a modular PKS, it deserves to be investigated more thoroughly.

In passing it should be explained that the PKS does not participate in another interesting feature of the bor pathway, the conversion of a branching methyl group into a nitrile.[191] The transformation is carried out by a fascinating sequence of post-PKS oxidative operations, but they are not relevant to the present topic.

1.10.9.8 Conclusion

This survey of PKS assembly lines has revealed two contrasting aspects of their organization. All the PKS systems contain modules that conform closely to the DEBS paradigm: they operate noniteratively, and the structure of the chain-extension unit each module produces corresponds to the pattern of reductive domains.

Much greater structural diversity can be achieved by superimposing special chemical operations on the standard modular pattern. Most importantly are the various oxidation and cyclization reactions, which can take place while the chain is attached to an ACP. These can occur at the end of the modular operations, as in the mon PKS, or can be inserted at an intermediate stage of chain development, as in the rif PKS.

Variation of the modular steps by skipping (pikromycin PKS), or iteration (bor PKS), is another potential source of structural and operational diversity. Indeed, on taking stock, it can be seen as a very fortunate circumstance that the DEBS was the first modular PKS to be discovered, ensuring that it became the focus of subsequent investigations. It is the simplest in its operation and is conveniently small as an experimental target. Some of the PKS systems reviewed in this section would have presented a considerably greater challenge to the early pioneers in the field, and early progress would also most certainly have been much slower.

1.10.10 Commercial Applications of Genetic Engineering of Modular Polyketide Synthases

The annual sales of PK-derived medicines has reached the staggering figure of $20 billion.[76] It is no surprise that the rapid growth of academic research into PK biosynthesis since 1990 has been paralleled by a major effort to engineer PKS systems to produce novel natural products.[5,75,76,192–197] The steadily falling rate of discovery of new natural PKs, at least in the terrestrial environment, has provided an added incentive to invest in this new strategy.[198] Both type I and type II systems have been the subject of intensive investigation. Considerable progress has been achieved on all fronts, as will be seen from the following brief survey of major areas of activity. All the projects selected for review in this section featured academic research groups working in association with industrial partners.

1.10.10.1 Development of New Versatile Super Hosts and Combinatorial Biosynthesis of Aromatic Compounds by Type II PKS Pathways

Early pioneering studies led by Hopwood and Khosla, in both their academic laboratories and in a biotechnology company, KOSAN, produced some early demonstrations of the power of genetic engineering to produce

novel metabolites.[10,37] The target PKS systems were type II systems producing aromatic PKs, which are not directly relevant here, but the advances in microbiology generated in these studies is more widely applicable.

The specific aim in this study of type II systems was to set up methods to allow rapid screening of large numbers of different combinations of PKS components capable of generating libraries of novel natural products. At that time, the generation of libraries of compounds had become a major focus of chemical synthesis in major pharmaceutical laboratories. The primary aim of this research effort, in keeping with the thrust of contemporary work on combinatorial synthesis, was the generation of extensive unfocussed libraries of novel PK structures.

A large number of relevant type II PKS systems were sequenced to provide the potential for wide variation in PKS composition.[10,37] To facilitate throughput, special strains of type II PK producers, *S. coelicolor*[199] and *S. lividans*[200] were developed, from which the natural PKS genes had been deleted. This ensured that there was no background of natural production to interfere with the operation of unnatural combinations of genes.[10,201]

It was then possible using these super-host strains to explore many unnatural combinations of genes, which produced aromatic PKs. The products had diverse chain lengths and cyclization patterns, with the carbocyclic aromatic rings fused in various ways to pyrone rings.

In terms of size and chemical diversity, however, the resulting library proved to be a disappointment. Using standard synthetic methodologies, synthetic chemists would be able to create equivalent libraries of PK aromatics with ease. The new host organism technology developed in this research effort is potentially very important, however, as an avenue for novel compound generation with engineered type I modular PKS systems, as will be seen below.[202–205] Also the super-host organisms have proved very useful for overexpression of type I PKS proteins for *in vitro* studies.[206]

1.10.10.2 Strategies for Engineering the Erythromycin PKS

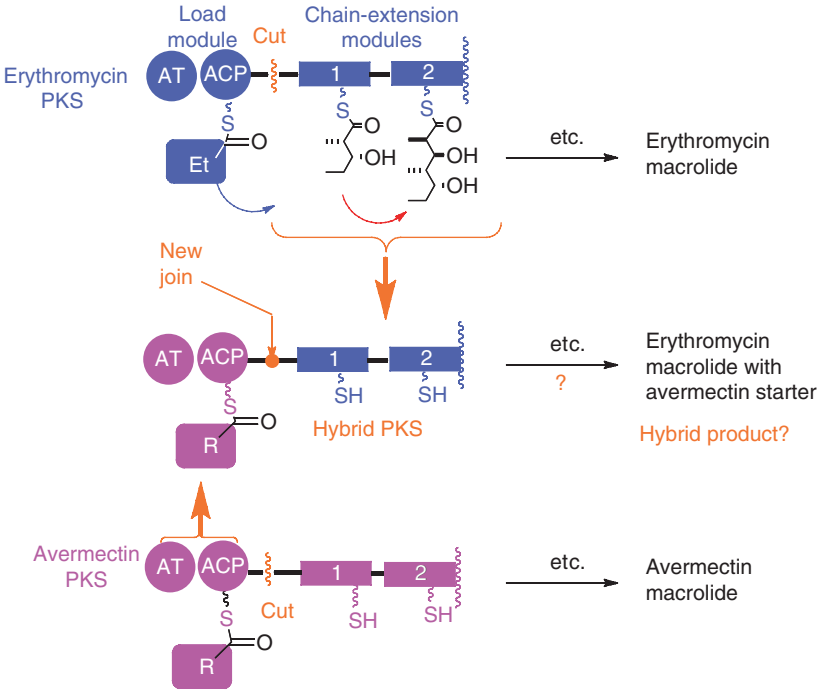
Erythromycin A has been in clinical use for more than 50 years. Semisynthetic derivatives, such as azithromycin,[207] produced by chemical modification of the natural product, have also achieved commercial success. Given the complex structure of the natural product, total synthesis of ery analogues is possible,[208] but not economical, so it is no surprise that there is no totally synthetic variant in commercial use. This limitation of chemical synthesis makes engineering of the biosynthesis of the macrolide core a very attractive prospect for producing new variations on the ery structure.

1.10.10.2.1 Creation of hybrid versions of the DEBS

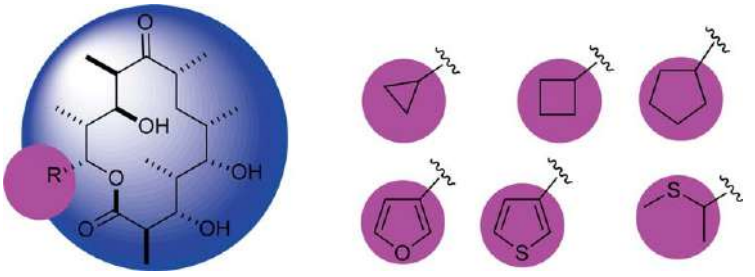
The first efforts to engineer the DEBS were inspired by earlier successful experiments by Pfizer to produce novel avermectins, by feeding unnatural starter acids to the producer organism.[209,210] The normal starter acids for avermectin are branched chain fatty acids *iso*-butyrate and *sec*-butanoate, derived by degradation of amino acids valine and leucine.[209] Several hundred carboxylic acids were fed to the producer organism, and it was found that more than 60 alternative starter acids could replace the natural ones.[210] The cyclohexyl analogue of the avermectin, doramectin, proved to have superior properties to the two normal natural products, and it came to dominate the commercial avermectin market.

Leadlay and coworkers reasoned that the LM of the avermectin PKS must have very broad specificity, and that this property might be retained if it could be transplanted to replace the natural more specific LM of the DEBS.[211] A collaboration was set up between Pfizer and Biotica, a biotechnology company founded by the Cambridge group to explore commercialization of genetic engineering of modular PKS systems. Biotica had the DEBS genes[40] and Pfizer the avermectin genes.[212] Genetic engineering in the Biotica laboratory created a hybrid modular PKS in which the first multienzyme consisted of DEBS 1 with the avermectin LM in place of the natural LM (**Scheme 61**).[211] This covalent combination of heterologous structural components was, of course, a much more ambitious and risky operation than simply mixing separate type II enzymes, which normally act *in trans* with their cooperating partners.

Incubation of the mutant containing the construct without added starter acids gave the natural product, erythromycin A with a propionate starter, together with three structural analogues having as starter acids, acetate, iso-butyrate, and sec-pentanoate. Clearly, the hybrid PKS multienzyme was active, and the avermectin LM had retained its normal broad specificity in the foreign environment of the DEBS multienzyme. In subsequent experiments, carried out by Pfizer, other potential starter acids were added to the incubation



Scheme 61 Strategy for generating a hybrid PKS incorporating the avermectin loading module in place of the natural DEBS loading module.

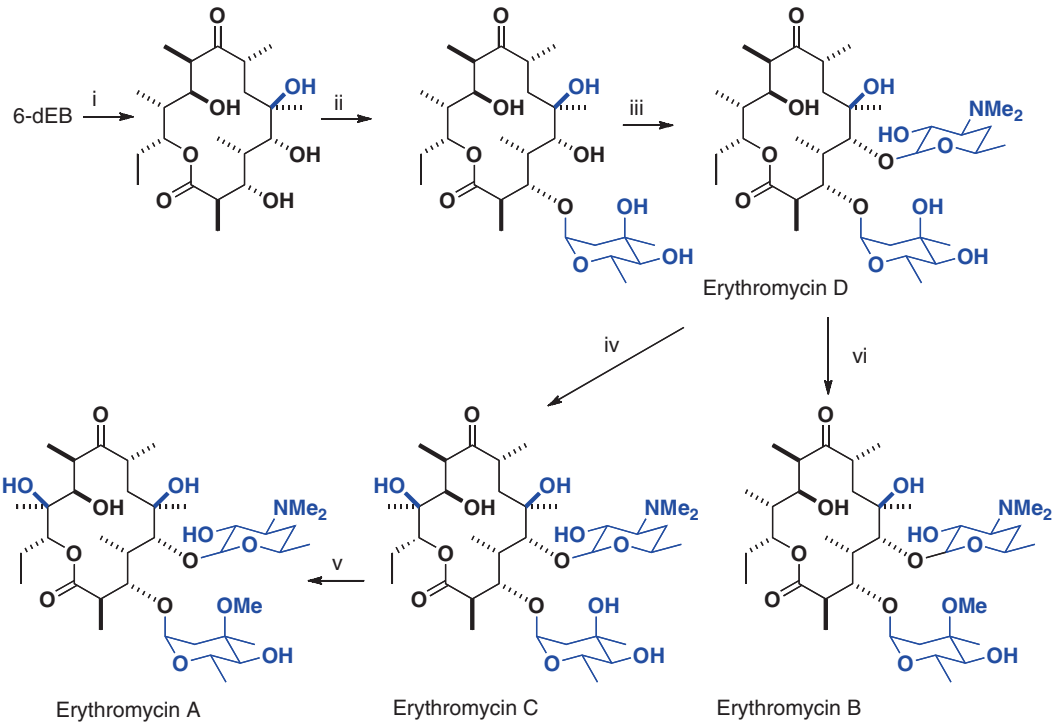


Scheme 62 Examples of the unnatural starter acid structures incorporated into the macrolide core by the hybrid PKS shown in **Scheme 61**.

medium and several were incorporated to produce the corresponding macrolide analogues.[213–215] **Scheme 62** shows examples of the novel macrolide products released as intermediates by heterologous PKS assembly line.

The hybrid PKS was operating in its normal host, *Sac. erythraea*, with its normal complement of post-PKS tailoring enzymes. Most of the initial PKS products were processed with remarkable efficiency through the post-PKS steps[26] shown in **Scheme 63** to give analogues of both erythromycins A and B, the normal final metabolites of the ery pathway. The major product with smaller starter acids is the A analogue, but with larger groups there was an increased proportion of the B analogue, presumably because increased size of substituent at C-17 inhibits the hydroxylation reaction at C-16.[211] Both the A and B analogues of ery have strong antibiotic activity, and this activity was retained in the fully processed biosynthetic analogues.

This was a pivotal development, which convinced synthetic chemists based in major pharmaceutical companies that genetic engineering of modular PKS systems could provide a powerful new strategy for generating novel complex natural products with promising pharmaceutical activity, as an economical alternative to total synthesis. The fact that the PKS products were fully converted into analogues of the normal



(i) C-6 hydroxylation; (ii,iii) glycosylation; (iv) C-12 hydroxylation; (v,vi) O-methylation

Scheme 63 Post-PKS steps in erythromycin biosynthesis.

natural product was an important advantage. The resulting library was relatively small, but the fact that most of its members showed useful pharmaceutical activity makes this focused strategy very promising for pharmaceutical chemists. As the first example of the creation of a heterologous modular PKS, the initial exploratory experiment also had major academic significance in that the success of the transplantation to form a heterologous hybrid proved that there was a useful degree of structural compatibility between multienzymes from different modular PKS megasynthases.

1.10.10.2.2 Mutasynthesis experiments with a DEBS 1 containing an inactivated KS1

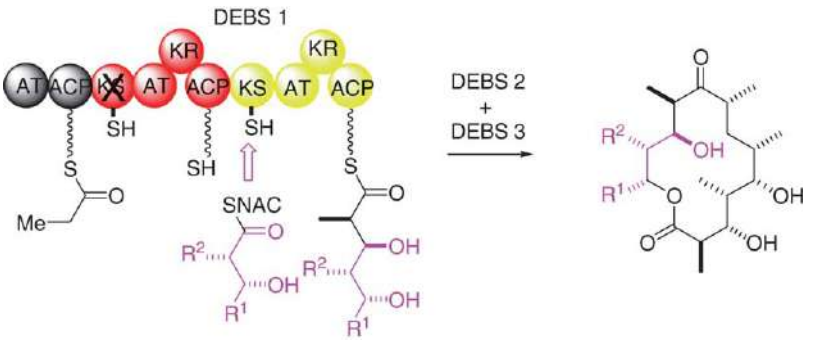
An alternative semisynthetic strategy, shown in **Scheme 64**, was developed by Cane and Khosla to produce analogues of 6-dEB containing analogues of the normal starter acid.[216] The strategy was based on inactivation of the KS domain of module 1 to prevent normal biosynthesis. Feeding synthetic analogues of the normal diketide to the mutant organism restored the synthetic operation to produce a range of structural analogues of 6-dEB.

These biosynthetic products were generated in a super host,[199] which lacked the normal post-PKS enzymes, so analogues of erythromycins A and B were not formed. The products were potentially transformable in various ways to produce extensive, but unfocussed libraries of compounds.

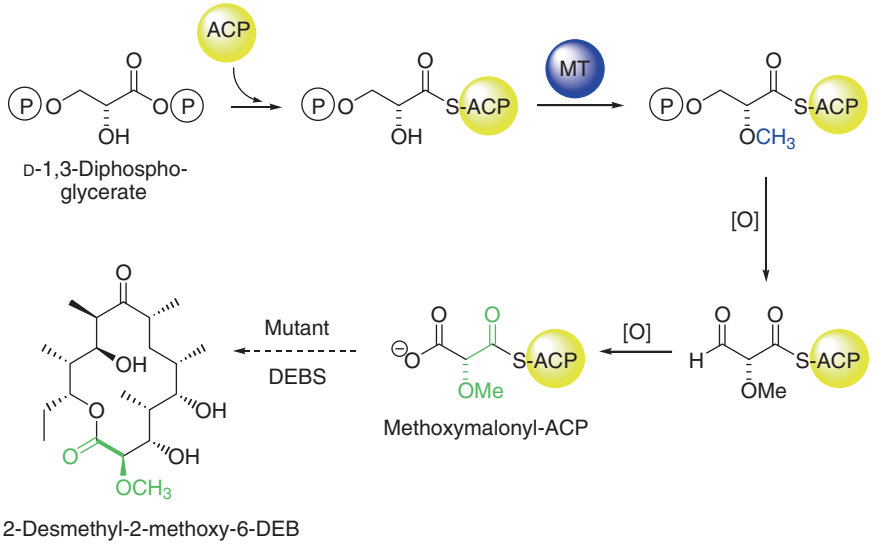
1.10.10.2.3 Genetic engineering of the chain-extension modules of the DEBS

McDaniel and coworkers at KOSAN carried out a number of important mutations in the chain-extension modules to replace the natural methylmalonate-specific AT domains, with foreign AT domains specific for malonate.[217] This gave access to analogues of 6-DEB lacking branching methyl groups at key sites. Once again, the initial small library of compounds could be extended in various ways to form larger unfocussed libraries.

In a related, more ambitious study, Floss and coworkers replaced AT6 in DEBS 3 with an AT domain specific for incorporation of methoxymalonate.[218] A methoxymalonyl unit is used for the biosynthesis of a



Scheme 64 Mutasynthesis using synthetic diketide analogues and an engineered DEBS with the activity of KS1 deleted.



Scheme 65 Incorporation of a novel methoxymalonate extender unit into 6 dEB.[218]

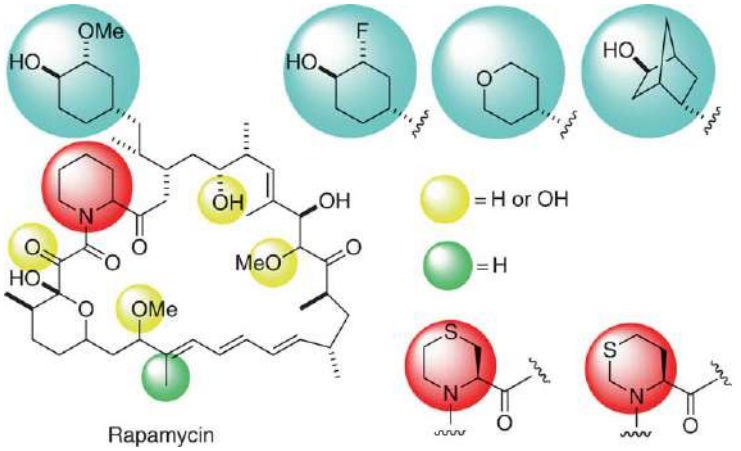
number of metabolites including FK520 and asm. Both asm and FK520 have five genes coding for the proteins responsible for the generation of this extender unit.[219,220] The set is comprised of an ACP, two oxidoreductases, an MT and a protein thought to act as an acyl transferase (AT) responsible for the loading of 1,3-diphosphoglycerate onto the ACP. The sequence of steps from D-1,3-diphosphoglycerate to methoxymalonate[221] is shown in **Scheme 65**.

Floss and coworkers heterologously expressed both the modified DEBS along with the five enzymes for methylmalonate production in one of the KOSAN super hosts, *S. lividans*, to generate the target analogue, 2-desmethyl-2-methoxymethyl-6 dEB.[218] There are two other novel ACP extenders known, hydroxymalonyl- and aminomalonyl-ACP, both of which are required for the biosynthesis of the antibiotic zwittermicin A from *Bacillus cereus*.[222] All three ACP extender units enable novel β-branches to be introduced into PK chains.

The three strategies presented in this section provide an impressive demonstration of the power of genetic engineering of modular PKS systems to produce libraries of complex metabolites not efficiently accessible by conventional total synthesis.

1.10.10.3 Generation of Analogues of Rapamycin

As an immunosuppressant, rap has been the subject of intensive efforts to generate analogues by both chemical and biosynthetic strategies.[76,153,193,223,224] Specific targets for alteration have been both the starter acid and the pipecolate part of the structure.



Scheme 66 Some of the variations of the rapamycin structure.

The starter acid can be replaced by simple feeding of analogues, as was first demonstrated by Leadlay and coworkers.[150,220] The efficiency of the strategy was greatly increased by Biotica through disruption of the pathway by which the normal starter acid is synthesized.[153,224] Some of the starter acid analogues produced by Biotica are shown in **Scheme 66**, along with alterations to the PK core shown in yellow and green. In total, around 70 analogues have been reported. The experiments were performed in the natural host for rap production, so the PKS products were fully processed by the available post-PKS enzymes to produce active rap analogues. A high proportion of the analogues showed rap-like activity.

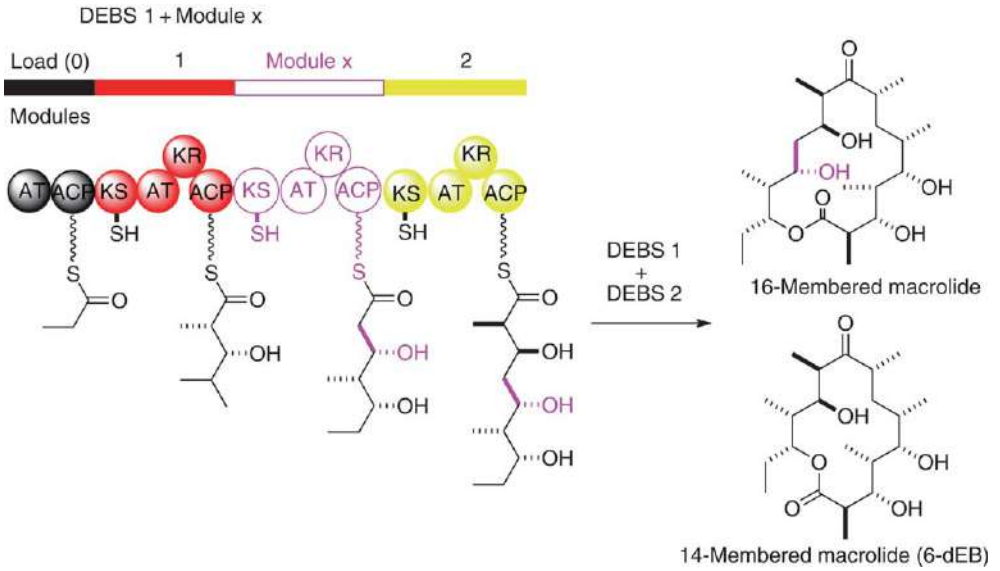
Several investigating groups have also produced structural variants around the pipecolate portion of the structure, a selection of which are shown.[225,226]

1.10.10.4 Insertion of a Complete Module into the DEBS Assembly Line

Leadlay and coworkers at Cambridge and Buss and coworkers at Glaxo, collaborated on a project which had as its goal the insertion of a complete module from the rap PKS between modules 1 and 2 of DEBS 1.[227] The aim was to push the boundaries of genetic engineering to new areas rather than immediately generating potentially commercial products. The expected result was the conversion of the DEBS assembly line into synthesis of a 16-membered macrolide, in place of its normal 14-membered product, 6-dEB, as shown in **Scheme 67**. Based on its function in the rap PKS, the inserted module was anticipated to incorporate a malonate unit and be converted into a hydroxyester intermediate, as indicated.

The experiment proved a success in that the anticipated 16-membered macrolide product was formed, but only in low yield. Surprisingly, the major product of the engineered PKS was the normal product of the DEBS, the 14-membered macrolide, 6-dEB. The explanation offered for this result was that the inserted module was subject to skipping.[227,183] There are many reasons why this could happen. The inserted module may not fold correctly, perhaps because of a worrying structural incompatibility of the foreign module with the flanking DEBS modules. An alternative simpler explanation is that the supply of malonate in the ery host is too low for the inserted module to maintain an adequate level of loading on its ACP domains. This would be expected to promote skipping, as was explained earlier.

Before closing, however, it is important to refer once again to a recurrent theme in this review: in an *in vivo* experiment the relative amounts of accumulated products do not necessarily represent the relative amounts of corresponding biosynthetic intermediates produced by the PKS. The normal role of the erythromycin TE is to form a 14-membered lactone and so the active site must be structured to ensure that the terminal hydroxyl is placed in the optimum position for nucleophilic attack on the enzyme-bound oxygen ester intermediate.[228] This orientation may not be so easily achieved with a larger macrocyclic ring, which probably explains the formation of a minor by-product from the extended chain in which the macrolide ring is 14-membered in size. It is feasible that the main mechanism of release of the extended chain is not macrolactonization but hydrolysis,



Scheme 67 Insertion of a complete foreign module into the DEBS assembly line.[227]

or ester transfer to a small alcohol acceptor such as methanol or glycerol. This would leave the chain vulnerable to fatty acid degradation, in which case the measured amounts of accumulated products would lead to an overestimation of the extent of skipping.

1.10.10.5 Prospects for Future Commercial Exploitation

It is apparent even in this brief survey that there is considerable potential for redirecting the biosynthetic operations of modular PKSs to produce libraries of novel natural products.

Two strategies have been developed. In one, the modular PKS is moved to a super host, stripped of its normal PKS genes, which acts as a heterologous expression system with a blank PKS background. As a result, the post-PKS tailoring steps are eliminated, and usually the product of the PKS accumulates. This opens the way to further steps, in which the initial product can be modified by biotransformation with suitable host organisms, or by chemical modification. The unfocussed strategy opens up a wide exploration of structural space, but only after further rounds of processing.

The alternative strategy is to express the modified PKS in its normal host organism. Now the post-PKS enzymes are present, as well as pathways to generate building blocks, opening the possibility that any modified PKS intermediates may be converted through into corresponding analogues of the normal natural product. The result is a smaller, more focused, library of compounds, but with a much higher prospect of finding useful pharmaceutical activity.

Many active analogues have been discovered as a result of genetic engineering. It can only be a matter of time before a commercial product is developed.

1.10.11 Future Perspectives

The aim in this closing summary is to look to possible future directions of research in this area, specifically from the point of view of a young chemist interested in joining the field at the start of a career.

Since the first modular type I PKS was reported in 1990,[40] there has been rapid progress in understanding how PKs are made. At that time, the Actinomycetes provided almost all known examples of reduced PKs. In the intervening years, PK metabolites have been isolated from other sources including myxobacteria,[9] insects,[229] and marine organisms.[230,231] Symbiotic bacteria have been discovered to be responsible for the production of

pederin in *Paederus* and *Paederidus* beetles,[163,164] rhizoxin in the fungus *Rhizopus microsporus*[232] and onnamide from a marine sponge.[233] Genome sequencing projects have enabled the identification of new PK gene clusters which has led to the subsequent identification of new PK metabolites such as bacillaene.[165,166] Many type I modular mixed nonribosomal-PK gene clusters, such as epothilone,[234,235] have been identified.[5,45,206,236,237] The NRPS modules allow novel starter units and amino acids to be incorporated into the PK chain. Since the first report of an AT-less PKS in 2002,[238,239] a further 10 AT-less type I modular gene clusters have been identified.[6] Three new ACP-extender units have been discovered in recent years,[221,222,240] expanding the genetic tool box for incorporating different β-side chains. The HMGS cassette has provided yet another way to attach α-side chains.[6] Structural biology[82,111–113] and NMR studies[106] have provided invaluable information about the structure of the megasynthases and demonstrated the importance of the linker regions in protein–protein interactions.

A striking feature of the activities in the field up to the present is the importance of multidisciplinary collaboration. Most of the leading groups have consisted of chemists and biologists working together toward a common goal, sometimes based in nearby laboratories in the same institution, sometimes based in laboratories long distances apart. The enormous breadth of technologies, which are required to establish a leading position, has dictated this pattern of operation. The required disciplines in biology include genetics, enzymology, and microbiology. In chemistry, there is a need for expertise in organic synthesis, reaction mechanisms, and isotopic labeling. There is also a need for expertise in spectroscopy and other forms of structural analysis in both the major disciplines. One of the pleasures but also the challenges for individuals working in this area is having to explaining ones own discipline to outsiders, while at the same time making the necessary complementary effort to understand, at least on a basic level, other very different disciplines.

Chemists have made important contributions to progress by studies of the mechanism and specificity of the component enzymes of the modular PKS megasynthases. This is a major area for investigation. For example, there is a major effort to investigate specificity by kinetic studies of isolated components of modules, and by site-specific mutagenesis of domains. Nevertheless, there has been relatively much less investigation of specificity effects in intact modules. For example, very little is known about the extent of loading of modules by substrates at multiple sites and the impact this would have as a consequence of occupancy level. This could have a profound impact on specificity as was noted earlier. Nothing is known about the extent of possible linkage between the two channels of synthesis on modular PKS. Do the channels operate entirely independently, or do they possibly cooperate in terms of the timing of the individual synthetic steps? In the cooperative mode, there could be a major impact on specificity and fidelity of the synthetic operations.

The study of products of engineered modules is another area that calls for a more rigorous approach. Usually these investigations are carried out *in vivo* for reasons of experimental convenience, but there is a price to pay in terms of the reliability of the results. The point has been made earlier that the cells of producer organisms, with their digestive enzymes, are hostile environments for small organic molecules that have linear chains terminated with a carboxylic acid or its derivative. Undoubtedly, many *in vivo* studies have been compromised because significant amounts of some of the products of PKS systems have been removed before analysis.

Answers to these sorts of questions and concerns can come from studies of intact modules *in vitro*. It will be necessary to have a major input from chemists, especially spectroscopists, able to identify organic intermediates both bound to the enzymes and free in solution. Expertise in organic synthesis would also be required to produce structural analogues to probe specificity.

Another very challenging field, which will certainly require major input from chemists, is the investigation of the quaternary structure of the PKS modules. The necessary movements of domains to facilitate the synthetic operations, whether it is just the ACP domains or others in addition, imply that there is a set of different quaternary conformations adopted by the megasynthase in the course of a synthetic cycle. This proposed behavior has a strong parallel in the familiar interconversion of conformational isomers found in simpler organic structures. The various conformers interchange in a dynamic equilibrium, the relative proportions being dependent on the relative stabilities of the components. In the megasynthases, domains replace atoms and linkers replace bonds. The behavior of a megasynthase is much more fluid than a simple organic molecule, but the energetic principles are the same.

The binding of substrates to the active sites of successive domains in a chain-extension cycle must play a role, possibly major, in determining which conformer is adopted at each stage. There is a need, therefore, to generate

modules, suitably derivatized by appropriate ligands, for study by X-ray crystallography, NMR, electron microscopy or whatever. To ensure the stability of any binding process, it may be necessary to generate covalent links through suicide inhibitor technology. All of these operations will rely heavily on the expertise and guidance of synthetic and mechanistic chemists. Chemical analysis will also be required to validate the structure and purity of protein samples before the structural investigation, and to check that decomposition, possibly by trace amounts of protease contaminant, has not altered the structure of the sample under study.

These studies will take decades, not months or years. The field offers a very exciting career path for young adventurous chemists.

Abbreviations

ACP	acyl carrier protein
ACP0	acyl carrier protein of a loading module
AcCoA	acetylCoA
AD	adenylation domain
AHBA	3-amino-5-hydroxybenzoic acid
asm	ansamitocin
AT	acyl CoA ACP transacylase (acyl transferase)
AT0	acyl transferase of the starter unit
B.	*Bacillus*
bor	borrelidin
bp	base pair (in DNA)
CL	carboxylic acid ligase
CoA	coenzyme A
D	docking interface
6-dEB	6-deoxyerythronolide B
DEBS	6-deoxyerythronolide B synthase
DH	dehydratase
DHCHC	3,4-dihydroxycyclohexanecarboxylic acid
E.	*Escherichia*
ery	erythromycin
HMGS	3-hydroxy-3-methylglutaryl synthase
9-HN	9-hydroxynonanoic acid
KR	3-oxoacyl ACP reductase (3-ketoacyl ACP reductase)
KR1	ketoreductase in module 1
KS	3-oxoacyl ACP synthase (3-ketoacyl ACP synthase), also called KAS or KSα
KSα	ketosynthase (aromatic polyketides)
KSβ	also called chain length factor
KS_Q	a ketosynthase with the active site cysteine replaced with glutamine
LM	loading module
MalCoA	malonyl CoA
MA	monic acid
MAT	malonyl acyl transferase
MT	methyl transferase
MMCoA	methylmalonyl CoA
Mod	module
mon	monensin
mup	mupirocin
6-MSA	6-methylsalicylic acid
NAC	*N*-acetyl cysteamine (*N*-acetyl-2-aminoethanethiol)
NRPS	nonribosomal peptide synthase

P-450	cytochrome P-450 mono-oxygenase
pik	pikromycin
PK	polyketide
PKS	polyketide synthase
rap	rapamycin
rif	rifamycin
S	spacer
SAM	*S*-adenosyl methionine
TE	thioesterase

Nomenclature

Da	dalton or relative molecular mass (RMM)
kb	kilobase
kDa	kilodalton

References

1. J. N. Collie; W. S. Myers, *J. Chem. Soc.* **1893**, *63*, 122.
2. J. N. Collie, *J. Chem. Soc.* **1907**, *91*, 1806.
3. A. J. Birch; P. A. Massy-Westropp; C. J. Moye, *Aus. J. Chem.* **1955**, *8*, 539.
4. J. Staunton; K. J. Weissman, *Nat. Prod. Rep.* **2001**, *18*, 380.
5. A. M. Hill, *Nat. Prod. Rep.* **2006**, *23*, 139.
6. T. J. Buchholz; J. D. Kittendorf; D. H. Sherman, Polyketide Biosynthesis, Modular Polyketide Synthases. In *Wiley Encyclopedia of Chemical Biology*; T. P. Begly, Ed.; John Wiley and Sons, 2008, doi: 10.1002/9780470048672.wecb459.
7. B. Shen, *Curr. Opin. Chem. Biol.* **2003**, *7*, 285.
8. S. G. Van Lanen; B. Shen, *Curr. Top. Med. Chem.* **2008**, *8*, 448.
9. S. C. Wenzel; R. Müller, *Nat. Prod. Rep.* **2007**, *24*, 1211.
10. C. Hertweck; A. Luzhetsky; Y. Rebets; A. Bechthold, *Nat. Prod. Rep.* **2007**, *24*, 162.
11. K. Watanabe; A. P. Praseuth; C. C. Wang, *Curr. Opin. Chem. Biol.* **2007**, 279.
12. R. O. Brady, *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 993.
13. S. J. Wakil, *J. Am. Chem. Soc.* **1958**, *880*, 6465.
14. S. J. Wakil, *J. Lipid Res.* **1961**, *2*, 1.
15. S. Smith; S.-C. Tsai, *Nat. Prod. Rep.* **2007**, *24*, 1041.
16. Y. Tang; C.-Y. Kim; I. I. Matthews; D. E. Cane; C. Khosla, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11124.
17. M. A. Reed; M. Schweizer; A. E. Szafranska; C. Arthur; T. P. Nicholson; R. J. Cox; J. Crosby; M. P. Crump; T. J. Simpson, *Org. Biol. Chem.* **2003**, *1*, 463.
18. Y. Y. Tang; A. Y. Chen; C. Y. Kim; D. E. Cane; C. Khosla, *Chem. Biol.* **2007**, *14*, 931.
19. S. I. Chang; C. G. Hammes, *Acc. Chem. Res.* **1990**, *23*, 363.
20. A. Witkowski; A. Joshi; K. S. Smith, *Biochemistry* **1997**, *36*, 16338.
21. C. Y. Lin; S. Smith, *J. Biol. Chem.* **1978**, *253*, 1954.
22. M. Leibundgut; S. Jenni; C. Frick; N. Ban, *Science* **2007**, *316*, 288.
23. J. R. Martin; W. Rosenbrook, *Biochemistry* **1967**, *6*, 435.
24. C. Khosla; Y. Tang; A. Y. Chen; N. A. Schnarr; D. E. Cane, *Annu. Rev. Biochem.* **2007**, *76*, 195.
25. J. Staunton; B. Wilkinson, *Chem. Rev.* **1997**, *97*, 2611.
26. B. J. Rawlings, *Nat. Prod. Rep.* **2001**, *18*, 190.
27. D. E. Cane; H. Hasler; P. B. Taylor; T. C. Liang, *Tetrahedron* **1983**, *39*, 3449.
28. S. M. Friedman; J. W. Corcoran; T. Kaneda, *J. Biol. Chem.* **1964**, *239*, 2386.
29. T. Kaneda; J. C. Butte; S. B. Taubman; J. W. Corcoran, *J. Biol. Chem.* **1962**, *237*, 322.
30. D. E. Cane; C. C. Yang, *J. Am. Chem. Soc.* **1987**, *109*, 1255.
31. P. Dimroth; H. Walter; F. Lynen, *Eur. J. Biochem.* **1970**, *13*, 98.
32. P. M. Schoolingin-Jordan; J. B. Spencer, *Biochem. Soc. Trans.* **1993**, *21*, 222.
33. J. B. Spencer; P. M. Schoolingin-Jordan, *Biochem. J.* **1992**, *288*, 839.
34. C. J. Child; J. B. Spencer; P. Bhogal; P. M. Schoolingin-Jordan, *Biochemistry* **1996**, *35*, 12267.

35. F. H. Malpartida; D. A. Hopwood, *Nature* **1984**, *309*, 462.
36. D. A. Hopwood; D. H. Sherman, *Annu. Rev. Genet.* **1990**, *24*, 67.
37. D. A. Hopwood, *Chem. Rev.* **1997**, *97*, 2465.
38. F. Malpartida; S. E. Halla; H. M. Kieser; H. Motamedi; C. R. Hutchinson; M. J. Butler; D. A. Sugden; M. Warren; C. McKillop; C. R. Bailey; G. O. Humphreys; D. A. Hopwood, *Nature* **1987**, *325*, 818.
39. C. J. Thompson; J. M. Ward; D. A. Hopwood, *J. Bacteriol.* **1982**, *151*, 668.
40. J. Cortés; S. F. Haydock; G. A. Roberts; D. J. Bevitt; P. F. Leadlay, *Nature* **1990**, *348*, 176.
41. S. Donadio; M. J. Staver; J. B. McAlpine; S. J. Swanson; L. Katz, *Science* **1991**, *252*, 675.
42. D. J. Bevitt; J. Cortés; S. F. Haydock; P. F. Leadlay, *Eur. J. Biochem.* **1992**, *111*, 51.
43. S. Donadio; L. Katz, *Gene* **1992**, *111*, 51.
44. J. S. Tuan; J. M. Weber; M. J. Staver; J. O. Leung; S. Donadio; L. Katz, *Gene* **1990**, *90*, 21.
45. K. J. Weissman; R. Müller, *ChemBioChem* **2008**, *9*, 826.
46. P. Caffrey; D. J. Bevitt; J. Staunton; P. F. Leadlay, *FEBS Lett.* **1992**, *304*, 225.
47. J. Cortés; K. E. H. Weismann; G. A. Roberts; M. J. B. Brown; J. Staunton; P. F. Leadlay, *Science* **1995**, *268*, 1487.
48. C. M. Kao; G. Luo; L. Katz; D. E. Cane; C. Khosla, *J. Am. Chem. Soc.* **1995**, *117*, 9105.
49. K. E. H. Weismann; J. Cortés; M. J. B. Brown; A. L. Cutter; J. Staunton; P. F. Leadlay, *Chem. Biol.* **1995**, *2*, 583.
50. I. Böhm; I. E. Holzbaur; U. Hanefield; J. Cortés; J. Staunton; P. F. Leadlay, *Chem. Biol.* **1998**, *5*, 407.
51. C. M. Kao; G. Luo; L. Katz; D. E. Cane; C. Khosla, *J. Am. Chem. Soc.* **1996**, *118*, 9184.
52. C. M. Kao; G. Luo; L. Katz; D. E. Cane; C. Khosla, *J. Am. Chem. Soc.* **1995**, *117*, 9105.
53. C. M. Kao; M. McPherson; R. N. McDaniel; H. Fu; D. E. Cane; C. Khosla, *J. Am. Chem. Soc.* **1997**, *119*, 11337.
54. R. McDaniel; C. M. Kao; H. Fu; P. Hevezi; C. Gustafsson; M. Betlach; G. Ashley; D. E. Cane; C. Khosla, *J. Am. Chem. Soc.* **1997**, *119*, 4309.
55. R. McDaniel; C. M. Kao; S. J. Hwang; D. E. Cane; C. Khosla, *Chem. Biol.* **1997**, *4*, 667.
56. M. Oliynyk; M. J. B. Brown; J. Cortés; J. Staunton; P. F. Leadlay, *Chem. Biol.* **1996**, *3*, 833.
57. H. Zhang; Y. Wang; B. A. Pfeifer, *Mol. Pharm.* **2008**, *5*, 212.
58. S. C. Wenzel; R. Müller, *Curr. Opin. Biotechnol.* **2005**, *16*, 6828.
59. M. J. B. Brown; J. Cortés; A. L. Cutter; P. F. Leadlay; J. Staunton, *J. Chem. Soc. Chem. Commun.* **1995**, 1517.
60. G. F. Liou; J. Lau; D. E. Cane; C. Khosla, *Biochemistry* **2003**, *42*, 200.
61. H. Hong; A. N. Appleyard; A. P. Siskos; J. Carcia-Bernado; J. Staunton; P. F. Leadlay, *FEBS J.* **2005**, *272*, 2373.
62. L. H. Østergaard; L. Kellenberger; J. Cortés; M. P. Roddis; M. Deacon; J. Staunton; P. F. Leadlay, *Biochemistry* **2002**, *41*, 2719.
63. W. D. Celmer, *J. Am. Chem. Soc.* **1965**, *87*, 1799.
64. W. D. Celmer, *J. Am. Chem. Soc.* **1965**, *87*, 1801.
65. M. McPherson; C. Khosla; D. E. Cane, *J. Am. Chem. Soc.* **1998**, *120*, 3267.
66. P. Caffrey, *ChemBioChem* **2003**, *4*, 654.
67. R. Reid; M. Piagentini; E. Rodriguez; G. Ashley; N. Viswanathan; J. Carney; D. V. Santi; C. R. Hutchinson; R. McDaniel, *Biochemistry* **2003**, *42*, 72.
68. S. Donadio; M. J. Staver; J. B. McAlpine; S. J. Swanson; L. Katz, *Science* **1991**, *252*, 675.
69. A. F. A. Marsden; P. Caffrey; J. P. Aparicio; M. S. Loughran; J. Staunton; P. F. Leadlay, *Science* **1994**, *263*, 378.
70. K. J. Weissman; M. Timoney; M. Bycroft; P. Grice; U. Hanefield; J. Staunton; P. F. Leadlay, *Biochemistry* **1997**, *36*, 13849.
71. I. E. Holzbauer; R. C. Harris; M. Bycroft; J. Cortés; C. Bisang; J. Staunton; B. A. M. Rudd; P. F. Leadlay, *Chem. Biol.* **1999**, *6*, 189.
72. C. Y. Kim; V. Y. Alekseyev; A. Y. Chen; Y. Y. Tang; D. E. Cane; C. Khosla, *Biochemistry* **2004**, *43*, 13892.
73. R. Castonguay; W. G. He; A. Y. Chen; C. Khosla; D. E. Cane, *J. Am. Chem. Soc.* **2007**, *129*, 13758.
74. A. Y. Chen; D. E. Cane; C. Khosla, *Chem. Biol.* **2007**, *14*, 784.
75. S. Weist; R. D. Süssmuth, *Appl. Microbiol. Biotechnol.* **2005**, *68*, 141.
76. K. J. Weissman, *Trends Biotechnol.* **2007**, *25*, 139.
77. S. J. Wakil; J. K. Stoops, Structure and Mechanism of Fatty Acid Synthetase. In *The Enzymes*; P. D. Boyer, Ed.; Academic Press: New York, 1983; Vol. XVI, p 3.
78. J. K. Stoops; S. J. Wakil, *J. Biol. Chem.* **1981**, *256*, 5128.
79. J. K. Stoops; S. J. Wakil; E. C. Uberbacher; G. J. Bunick, *J. Biol. Chem.* **1987**, *262*, 10246.
80. A. Witkowski; A. K. Joshi; S. Smith, *Biochemistry* **1996**, *35*, 10569.
81. A. K. Joshi; A. Witkowski; S. Smith, *Biochemistry* **1997**, *36*, 2316.
82. A. K. Joshi; A. Witkowski; S. Smith, *Biochemistry* **1998**, *36*, 2515.
83. A. K. Joshi; V. S. Rangan; S. Smith, *J. Biol. Chem.* **1998**, *273*, 4937.
84. A. Witlowski; A. K. Joshi; V. S. Rangan; A. M. Falick; H. E. Witkowska; S. Smith, *J. Biol. Chem.* **1999**, *274*, 11557.
85. V. S. Rangan; A. K. Joshi; S. Smith, *Biochemistry* **2001**, *40*, 10792.
86. A. K. Joshi; V. S. Rangan; A. Witkowski; S. Smith, *Chem. Biol.* **2003**, *10*, 169.
87. T. Maier; S. Jenni; N. Ban, *Science* **2006**, *311*, 1258.
88. T. Maier; N. Leibunggut; N. Ban, *Science* **2008**, *321*, 1315.
89. S. W. White; J. Zheng; Y. M. Zhang; C. O. Rock, *Annu. Rev. Biochem.* **2005**, *74*, 791.
90. S. Jenni; M. Leibundgut; D. Boehringer; C. Frick; B. Mikolasek; N. Ban, *Science* **2007**, *316*, 254.
91. I. B. Lomakin; Y. Xiong; T. A. Steiz, *Cell* **2007**, *129*, 319.
92. M. Leibundgut; S. Jenni; C. Frick; N. Ban, *Science* **2007**, *316*, 288.
93. E. Plokson; C. J. Arthur; S. E. Evans; C. Williams; J. Crosby; T. J. Simpson; M. P. Crump, *J. Biol. Chem.* **2008**, *283*, 518.
94. G. Bunkoczi; S. Pasta; A. Joshi; X. Q. Wu; K. L. Kavanough; S. Smith; U. Opperman, *Chem. Biol.* **2007**, *14*, 1243.
95. B. Chakravarty; Z. Gu; S. S. Chirala; S. J. Wakil; F. A. Quiocho, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15567.
96. C. W. Pemble IV; L. C. Johnson; S. J. Kridel; W. T. Lowther, *Nat. Struct. Mol. Biol.* **2007**, *14*, 704.
97. M. Leibundgut; T. Maier; S. Jenni; N. Ban, *Curr. Opin. Struct. Biol.* **2008**, *18*, 714.
98. S. Smith, *Science* **2006**, *311*, 1251.
99. C. A. Townsend; J. M. Crawford; T. Bililign, *Chem. Biol.* **2006**, *13*, 349.

100. P. F. Leadlay; J. Staunton; J. F. Aparicio; D. J. Bevitt; P. Caffrey; J. Cortes; A. Marsden; G. A. Roberts, *Biochem. Soc. Trans.* **1993**, *21*, 218.
101. G. A. Roberts; J. Staunton; P. F. Leadlay, *Eur. J. Biochem.* **1993**, *214*, 305.
102. J. F. Aparicio; P. Caffrey; A. F. A. Marsden; J. Staunton; P. F. Leadlay, *J. Biol. Chem.* **1994**, *269*, 8524.
103. J. Staunton; P. Caffrey; J. F. Aparicio; G. A. Roberts; S. S. Bethel; P. F. Leadlay, *Nat. Struct. Biol.* **1996**, *3*, 188.
104. C. M. Kao; R. Pieper; D. E. Cane; C. Khosla, *Biochemistry* **1996**, *35*, 12363.
105. R. S. Gokhale; J. Lau; D. E. Cane; C. Khosla, *Biochemistry* **1998**, *37*, 2524.
106. R. W. Broadhurst; D. Nietlispach; M. P. Wheatcroft; P. F. Leadlay; K. J. Weissman, *Chem. Biol.* **2003**, *10*, 723.
107. C. M. Squire; R. J. M. Goss; H. Hong; P. F. Leadlay; J. Staunton, *ChemBioChem* **2003**, *4*, 1225.
108. R. Mc.Daniel; C. M. Kao; S. J. Hwang; C. Khosla, *Chem. Biol.* **1997**, *4*, 667.
109. K. J. Weissman, *ChemBioChem* **2006**, *7*, 485.
110. K. J. Weissman, *ChemBioChem* **2006**, *7*, 1334.
111. V. Y. Alekseyev; C. W. Liu; D. E. Cane; J. D. Puglisi; C. Khosla, *Protein Sci.* **2007**, *16*, 2093.
112. Y. Y. Tang; A. Y. Chen; C. Y. Kim; D. E. Cane; C. Khosla, *Chem. Biol.* **2007**, *14*, 931.
113. Y. Tang; C. Y. Kim; I. I. Mathews; D. E. Cane; C. Khosla, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11124.
114. A. T. Keatinge-Clay; R. M. Stroud, *Structure* **2006**, *14*, 737.
115. A. T. Keatinge-Clay, *Chem. Biol.* **2007**, *14*, 898.
116. C. D. Richter; D. A. Stanmore; R. N. Miguel; M. C. Moncrieffe; L. Tran; S. Brewerton; F. Meersman; R. W. Broadhurst; K. J. Weissman, *FEBS J.* **2007**, *274*, 2196.
117. E. J. Brignole; S. Smith; F. J. Asturias, *Nat. Struct. Mol. Biol.* **2009**, *16*, 190.
118. P. F. Leadlay; J. Staunton; M. Oliynyk; C. Bisang; J. Cortés; E. Frost; Z. A. Hughes-Thomas; M. A. Jones; S. G. Kendrew; J. B. Lester; P. F. Long; H. A. I. McArthur; E. L. McCormick; Z. Oliynyk; C. B. W. Stark; C. J. Wilkinson, *J. Indus. Microb. Biotechnol.* **2001**, *27*, 360.
119. M. Oliynyk; C. B. W. Stark; A. Bhatt; M. A. Jones; Z. A. Hughes-Thomas; C. Wilkinson; Z. Oliynyk; Y. Demydchuk; J. Staunton; P. F. Leadlay, *Mol. Microbiol.* **2003**, *49*, 1179.
120. T. Schwecke; J. F. Aparicio; I. Molnár; A. König; L. E. Khaw; S. F. Haydock; M. Oliynyk; P. Caffrey; J. Cortés; J. B. Lester; G. A. Böhm; J. Staunton; P. F. Leadlay, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 7839.
121. J. F. Aparicio; I. Molnár; T. Schwecke; A. König; S. F. Haydock; L. E. Khaw; J. Staunton; P. F. Leadlay, *Gene* **1996**, *169*, 9.
122. T. P. Stinear; A. Mve-Obiang; P. L. C. Small; W. Frigui; M. L. Pryor; R. Brosch; G. A. Jenkin; P. D. R. Johnson; J. K. Davies; R. E. Lee; S. Adusumilli; T. Garnier; S. F. Haydock; P. F. Leadlay; S. T. Cole, *Proc. Natl. Acad. Sci. U.S A.* **2004**, *191*, 1345.
123. C. P. Ridley; H. Y. Lee; C. Khosla, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4595.
124. H. Hong; C. Demangel; S. J. Pidot; P. F. Leadlay; T. Stinear, *Nat. Prod. Rep.* **2008**, *25*, 447.
125. A. M. Hill, Polyketide Biosynthesis, Polyethers. In *Wiley Encyclopedia of Chemical Biology*; T. D. Begley, Ed.; John Wiley and Sons, 2008, doi: 10.1002/9780470048672.wecb460.
126. F. G. Riddell, *Chem. Br.* **1992**, *28*, 533.
127. B. M. Harvey; T. Mironenko; Y. H. Sun; H. Hong; Z. X. Deng; P. F. Leadlay; K. J. Weissman; S. F. Haydock, *Chem. Biol.* **2007**, *14*, 703.
128. Y. Sun; X. Zhou; H. Dong; G. Tu; M. Wang; B. Wang; Z. Deng, *Chem. Biol.* **2003**, *10*, 431.
129. T. Liu; D. You; C. Valenzano; Y. Sun; J. Li; Q. Yu; X. Zhou; D. E. Cane; Z. Deng, *Chem. Biol.* **2006**, *13*, 945.
130. Y. Demydchuk; Y. H. Sun; H. Hong; J. Staunton; J. B. Spencer; P. F. Leadlay, *ChemBioChem* **2008**, *9*, 1136.
131. B. Shen; H.-J. Kwon, *Chem. Rec.* **2002**, *2*, 389.
132. H.-J. Kwon; W. C. Smith; J. Scharon; S. H. Hwang; M. J. Kurth; B. Shen, *Science* **2002**, *297*, 1327.
133. J. W. Westley, Polyether Antibiotics – Biosynthesis. In *Antibiotics IV*; J. W. Corcoran, Ed., Springer-Verlag: New York; 1981, pp 41–73.
134. D. E. Cane; T. C. Liang; H. C. Hasler, *J. Am. Chem. Soc.* **1982**, *104*, 7274.
135. D. E. Cane; W. D. Celmer; J. W. Westley, *J. Am. Chem. Soc.* **1983**, *105*, 3594.
136. B. M. Harvey; H. Hong; M. A. Jones; Z. A. Hughes-Thomas; R. M. Goss; M. L. Heathcote; V. M. Bolanos-Garcia; W. Kroutil; J. Staunton; P. F. Leadlay; J. B. Spencer, *ChemBioChem* **2006**, *7*, 1435.
137. A. Bhatt; C. B. W. Stark; B. M. Harvey; A. R. Gallimore; Y. A. Demydchuk; J. B. Spencer; J. Staunton; P. F. Leadlay, *Angew. Chem. Int. Ed.* **2005**, *44*, 7075.
138. A. R. Gallimore; C. B. W. Stark; A. Bhatt; B. M. Harvey; Y. Demydchuck; V. Bolanos-Garcia; D. J. Fowler; J. Staunton; P. F. Leadlay; J. B. Spencer, *Chem. Biol.* **2006**, *13*, 453.
139. J. E. Baldwin, *J. Chem. Soc. Chem. Commun.* **1976**, *11*, 734.
140. W. Hüttel; J. E. Davies; J. B. Spencer; P. F. Leadlay, Unpublished results.
141. C. Bisang; P. F. Long; J. Cortés; J. Westcott; J. Crosby; A. L. Matharu; R. J. Cox; T. J. Simpson; J. Staunton; P. F. Leadlay, *Nature* **1999**, *401*, 502.
142. P. F. Long; C. J. Wilkinson; C. P. Bisang; J. Cortés; N. Dunster; M. Oliynyk; E. McCormick; H. McArthur; C. Mendez; J. A. Salas; J. Staunton; P. F. Leadlay, *Mol. Microbiol.* **2002**, *43*, 1215.
143. A. G. DiLella; R. J. Craig, *Biochemistry* **1991**, *30*, 8512.
144. E. J. Brown; M. W. Albers; T. B. Shin; K. Ichikawa; C. T. Keith; W. S. Cane; S. L. Schreiber, *Nature* **1994**, *369*, 756.
145. P. A. S. Lowden; B. Wilkinson; G. A. Böhm; S. Handa; H. G. Floss; P. F. Leadlay; J. Staunton, *Angew. Chem. Int. Ed.* **2001**, *40*, 777.
146. N. L. Paiva; A. L. Demain; M. F. Roberts, *J. Nat. Prod.* **1991**, *54*, 167.
147. A. König; T. Schwecke; I. Molnár; G. A. Böhm; P. A. S. Lowden; J. Staunton; P. F. Leadlay, *Eur. J. Biochem.* **1997**, *247*, 526.
148. I. Molnár; J. F. Aparicio; S. F. Haydock; L. E. Khaw; T. Schwecke; A. König; J. Staunton; P. F. Leadlay, *Gene* **1996**, *169*, 1.
149. L. E. Khaw; G. A. Böhm; S. Metcalfe; J. Staunton; P. F. Leadlay, *J. Bacteriol.* **1998**, *180*, 809.
150. G. J. Gatto, Jr.; M. T. Boyne; N. L. Kelleher; C. T. Walsh, *J. Am. Chem. Soc.* **2006**, *128*, 3838.
151. M. A. Gregory; S. Gaisser; R. E. Lill; H. Hong; R. M. Sheridan; B. Wilkinson; H. Petkovic; A. J. Weston; I. Carletti; H.-L. Lee; J. Staunton; P. F. Leadlay, *Angew. Chem. Int. Ed.* **2004**, *43*, 2551.

152. M. A. Gregory; H. Hong; R. E. Lill; S. Gaisser; H. Petkovic; L. Low; L. S. Sheehan; I. Carletti; S. J. Ready; M. J. Ward; A. L. Kaya; A. J. Weston; I. R. Challis; P. F. Leadlay; C. J. Martin; B. Wilkinson; R. M. Sheridan, *Org. Biomol. Chem.* **2006**, *4*, 3565.
153. M. A. Gregory; H. Petkovic; R. E. Lill; S. J. Moss; B. Wilkinson; S. Gaisser; P. F. Leadlay; R. M. Sheridan, *Angew. Chem. Int. Ed.* **2005**, *44*, 4757.
154. J. M. Cassady; K. K. Chan; H. G. Floss; E. Leistner, *Chem. Pharm. Bull.* **2004**, *52*, 1.
155. H. G. Floss; T.-W. Yu, *Chem. Rev.* **2005**, *105*, 621.
156. P. R. August; L. Tang; Y. J. Yoon; S. Ning; R. Müller; T.-W. Yu; M. Taylor; D. Hoffman; C.-G. Kim; X. Zhang; C. R. Hutchinson; H. G. Floss, *Chem. Biol.* **1998**, *5*, 69.
157. T. Schupp; C. Toupet; N. Engel; S. Goff, *FEMS Microbiol. Lett.* **1998**, *159*, 201.
158. K. Watanabe; M. A. Rude; C. T. Walsh; C. Khosla, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9774.
159. J. Xu; E. Wan; C.-J. Tam; H. G. Floss, *Microbiology* **2005**, *151*, 2515.
160. T.-W. Yu; Y. Shen; Y. Doi-Katayama; L. Tang; C. Park; B. S. Moore; C. R. Hutchison; H. G. Floss, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *96*, 9051.
161. A. Stratmann; C. Toupet; W. Schilling; R. Traber; L. Oberer; T. Schupp, *Microbiology* **1999**, *145*, 3365.
162. A. El Sayed; J. Hothersall; S. M. Cooper; E. Stephens; T. J. Simpson; C. M. Thomas, *Chem. Biol.* **2003**, *10*, 419.
163. J. Piel, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14002.
164. J. Piel; G. P. Wen; M. Platzer; D. Q. Hui, *ChemBioChem* **2004**, *5*, 93.
165. R. A. Butcher; F. C. Schroeder; M. A. Fisbach; P. D. Straight; R. Kolter; C. T. Walsh; J. Clardy, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1506.
166. J. Moldenhauer; X. H. Chen; R. Borriss; J. Piel, *Angew. Chem. Int. Ed.* **2007**, *46*, 8195.
167. A. S. Rahman; J. Hothersall; J. Crosby; T. J. Simpson; C. M. Thomas, *J. Biol. Chem.* **2005**, *280*, 6399.
168. T. J. Simpson; R. J. Cox, Polyketide Biosynthesis, Fungi. In *Wiley Encyclopedia of Chemical Biology*; T. D. Begley, Ed.; John Wiley and Sons, 2008, doi: 10.1002/9780470048672.wecb458.
169. F. M. Martin; T. J. Simpson, *J. Chem. Soc. Perkin Trans. I* **1989**, 207.
170. J. Wu; S. M. Cooper; R. J. Cox; J. Crosby; M. P. Crump; J. Hothersall; T. J. Simpson; C. M. Thomas; C. L. Willis, *Chem. Commun.* **2007**, 2040.
171. J. Hothersall; J. Wu; A. S. Ratiman; J. A. Shields; J. Haddock; N. Johnson; S. M. Cooper; E. R. Stephens; R. J. Cox; J. Cosby; C. L. Willis; T. J. Simpson; C. M. Thomas, *J. Biol. Chem.* **2007**, *282*, 15451.
172. S. M. Cooper; R. J. Cox; J. Crosby; M. P. Crump; J. Hothersall; W. Laosripaiboon; T. J. Simpson; C. M. Thomas, *Chem. Commun.* **2005**, 1179.
173. S. M. Cooper; W. Laosripaiboon; A. S. Rahman; J. Hothersall; A. K. El-Sayed; C. Winfield; J. Crosby; R. J. Cox; T. J. Simpson; C. M. Thomas, *Chem. Biol.* **2005**, *12*, 825.
174. J. Wu; J. Hothersall; C. Mazzetti; Y. O'Connel; J. A. Shields; A. S. Rahman; R. J. Cox; J. Crosby; T. J. Simpson; C. M. Thomas; C. L. Willis, *ChemBioChem* **2008**, *9*, 1500.
175. Y. Xue; L. Zhao; H.-W. Liu; D. H. Sherman, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12111.
176. J. Wu; W. He; C. Khosla; D. E. Cane, *Angew. Chem. Int. Ed.* **2005**, *44*, 7557.
177. Y. Xue; D. H. Sherman, *Nature* **2000**, *403*, 571.
178. M. L. Heathcote; J. Staunton; P. F. Leadlay, *Chem. Biol.* **2001**, *8*, 207.
179. B. J. Beck; Y. J. Yoon; K. A. Reynolds; D. H. Sherman, *Chem. Biol.* **2002**, *7*, 907.
180. I. Thomas; C. J. Martin; C. J. Wilkinson; J. Staunton; P. F. Leadlay, *Chem. Biol.* **2002**, *9*, 781.
181. C. C. Aldrich; B. J. Beck; R. A. Fecik; D. H. Sherman, *J. Am. Chem. Soc.* **2005**, *127*, 8441.
182. J. Kittendorf; B. J. Beck; T. J. Buchholz; W. Seufert; D. H. Sherman, *Chem. Biol.* **2007**, *14*, 944.
183. S. J. Moss; C. J. Martin; B. Wilkinson, *Nat. Prod. Rep.* **2004**, *21*, 575.
184. C. Olano; B. Wilkinson; C. Sánchez; S. J. Moss; R. Sheridan; V. Math; A. J. Weston; A. F. Braña; C. J. Martin; M. Oliynyk; C. Méndez; P. F. Leadlay; J. A. Salas, *Chem. Biol.* **2004**, *11*, 87.
185. C. Olano; B. Wilkinson; S. J. Moss; A. F. Braña; C. Méndez; P. F. Leadlay; J. A. Salas, *Chem. Commun.* **2003**, 2780.
186. S. W. Haynes; G. L. Challis, *Curr. Opin. Drug. Discovery Dev.* **2007**, *10*, 203.
187. N. Gaitatzis; B. Silakowski; B. Kunze; G. Nordsiek; H. Blocker; G. Höfle; R. Müller, *J. Biol. Chem.* **2002**, *277*, 13082.
188. J. He; C. Hertweck, *Chem. Biol.* **2003**, *10*, 1225.
189. J. He; C. Hertweck, *ChemBioChem* **2005**, *6*, 908.
190. B. Wilkinson; G. Foster; B. A. M. Rudd; N. L. Taylor; A. P. Blackaby; P. J. Sidebottom; D. J. Cooper; M. J. Dawson; A. D. Buss; S. Gaisser; I. U. Böhm; C. J. Rowe; J. Cortés; P. F. Leadlay; J. Staunton, *Chem. Biol.* **2000**, *7*, 111.
191. C. Olano; S. J. Moss; A. F. Braña; R. M. Sheridan; V. Math; A. J. Weston; C. Méndez; P. F. Leadlay; B. Wilkinson; J. A. Salas, *Mol. Microbiol.* **2004**, *52*, 1745.
192. K. J. Weissman; P. F. Leadlay, *Nat. Rev. Microbiol.* **2005**, *3*, 925.
193. A. Kirschning; F. Taft; T. Knobloch, *Org. Biomol. Chem.* **2007**, *5*, 3245.
194. M.-Q. Zhang; B. Wilkinson, *Curr. Opin. Biotechnol.* **2007**, *18*, 478.
195. H. G. Menzelles; C. D. Reeves, *Curr. Opin. Microbiol.* **2007**, *10*, 238.
196. S. C. Wenzel; R. Müller, *Curr. Opin. Biotechnol.* **2005**, *16*, 594.
197. G. Schneider, *Curr. Opin. Struct. Biol.* **2005**, *15*, 629.
198. C. T. Walsh, *Nat. Rev. Microbiol.* **2003**, *1*, 65.
199. R. McDaniel; S. Ebert-Khosla; D. A. Hopwood; C. Khosla, *Science* **1993**, *262*, 1546.
200. R. Ziermann; M. C. Batlach, *Biotechniques* **1999**, *26*, 106.
201. B. J. Rawlings, *Nat. Prod. Rep.* **1999**, *16*, 425.
202. H. Zhang; Y. Wang; B. A. Pfeiffer, *Mol. Pharmacol.* **2008**, *5*, 212.
203. B. A. Pfeifer; S. J. Admiraal; H. Gramajo; D. E. Cane; C. Khosla, *Science* **2001**, *291*, 1790.
204. R. S. Gokhale; S. Y. Tsuji; D. E. Cane; C. Khosla, *Science* **1999**, *284*, 482.
205. Q. Xue; G. Ashley; C. R. Hutchinson; D. V. Santi, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11740.
206. E. S. Sattely; M. A. Fischbach; C. T. Walsh, *Nat. Prod. Rep.* **2008**, *25*, 257.

207. S. Mutak, *J. Antibiot.* **2007**, *60*, 85.
208. S. Pal, *Tetrahedron* **2006**, *62*, 3171.
209. E. W. Hafner; B. W. Holley; K. S. Holdom; S. E. Lee; R. G. Wax; D. Beck; H. A. I. McArthur; W. C. Wernau, *J. Antbiot.* **1991**, *44*, 349.
210. C. J. Dutton; S. P. Gibson; A. C. Goudie; K. S. Holdom; M. S. Pacey; J. C. Ruddock; J. D. Bu'Lock; M. K. Richards, *J. Antibiot.* **1991**, *44*, 357.
211. A. F. A. Marsden; B. Wilkinson; J. Cortés; N. J. Dunster; J. Staunton; P. F. Leadlay, *Science* **1998**, *279*, 199.
212. H. Ikeda; T. Nonomiya; M. Usami; T. Ohta; S. Omura, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *17*, 9509.
213. M. S. Pacey; J. P. Dirlam; R. W. Geldart; P. F. Leadlay; H. A. I. McArthur; E. L. McCormick; R. A. Monday; T. N. O'Connell; J. Staunton; T. J. Winchester, *J. Antibiot.* **1998**, *51*, 1029.
214. I. C. Parsons; J. R. Everett; M. S. Pacey; J. C. Ruddock; A. G. Swanson; C. M. Thompson, *J. Antibiot.* **1999**, *52*, 190.
215. M. S. Brown; J. P. Dirlam; H. A. I. McArtur; E. L. McCormick; B. K. Morse; P. A. Murphy; T. N. O'Connell; M. Pacey; D. M. Rescek; J. Ruddock; R. G. Wax, *J. Antibiot.* **1999**, *52*, 742.
216. J. R. Jacobsen; C. R. Hutchinson; D. E. Cane; C. Khosla, *Science* **1997**, *277*, 367.
217. C. D. Reeves; S. Murli; G. W. Ashley; M. Piagentini; C. R. Hutchinson; R. McDaniel, *Biochemistry* **2001**, *40*, 15464.
218. Y. Kato; L. Bai; Q. Xue; W. P. Revill; T.-W. Yu; H. G. Floss, *J. Am. Chem. Soc.* **2002**, *124*, 5268.
219. T.-W. Yu; L. Bai; D. Clade; D. Hoffman; S. Toelzer; K. Q. Trinh; J. Xu; S. J. Moss; E. Leistner; H. G. Floss, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7968.
220. K. Wu; L. Chung; W. P. Revill; L. Katz; C. D. Reeves, *Gene* **2000**, *251*, 81.
221. S. C. Wenzel; R. M. Williamson; C. Grunanger; J. Xu; K. Gerth; R. A. Martinez; S. J. Moss; B. J. Carroll; S. Grond; C. J. Unkefen; R. Müller; H. G. Floss, *J. Am. Chem. Soc.* **2006**, *128*, 14325.
222. Y. A. Chan; M. P. Boyne; A. M. Podevels; A. K. Klimowicz; J. Handelsman; N. L. Kelleher; M. G. Thomas, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 14349.
223. P. A. Lowden; G. A. Böhm; S. Metcalfe; J. Staunton; P. F. Leadlay, *ChemBioChem* **2004**, *5*, 535.
224. R. J. M. Goss; S. E. Lanceron; N. J. Wise; S. J. Moss, *Org. Biomol. Chem.* **2006**, *4*, 4071.
225. E. I. Graziani; F. V. Ritacio; M. Y. Summers; T. M. Zabriskie; K. Yu; V. S. Bernan; M. Greenstain; G. T. Carter, *Org. Lett.* **2003**, *5*, 2385.
226. F. V. Titacco; E. I. Graziani; M. Y. Summers; T. M. Zabriskie; K. Yu; V. S. Bernan; G. T. Carter; M. Greenstain, *Appl. Environ. Microbiol.* **2005**, *71*, 1971.
227. C. J. Rowe; I. U. Böhm; I. P. Thomas; B. Wilkinson; B. A. M. Rudd; G. Foster; A. P. Blackaby; P. J. Sidebottom; Y. Roddis; A. D. Buss; J. Staunton; P. F. Leadlay, *Chem. Biol.* **2001**, *8*, 475.
228. S.-C. Tsai; L. J. W. Miercke; J. Krudinksi; R. Gokhale; J. C.-H. Chen; P. G. Foster; D. E. Cane; C. Khosla; R. M. Stroud, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14808.
229. F. Pankewitz; M. Hilker, *Biol. Rev.* **2008**, *83*, 209.
230. L. Fieseler; U. Hentschel; L. Grozdanov; A. Schirmer; G. P. Wen; M. Platzer; S. Hrvatin; D. Butzke; K. Zimmerman; J. Piel, *Appl. Environ. Microbiol.* **2007**, *73*, 2144.
231. J. Piel, *Curr. Med. Chem.* **2006**, *13*, 39.
232. L. P. Partida-Martinez; C. Hertweck, *ChemBioChem* **2007**, *8*, 41.
233. J. Piel; D. Hui; G. Wen; D. Butzke; M. Platzer; N. Fusetani; S. Matsunaga, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16222.
234. I. Molnár; T. Schupp; M. Ono; R. E. Zirckle; M. Milnamow; B. Nowak-Thompson; N. Engel; C. Toupet; A. Stratmann; D. D. Cyr; J. Gorlach; J. M. Mayo; A. Hu; S. Goff; J. Schmid; J. M. Ligon, *Chem. Biol.* **2000**, *7*, 97.
235. L. Tang; S. Shah; R. Zimmerman; R. Goldman; L. Katz; B. Julien, *Science* **2000**, *287*, 640.
236. S. Donadio; P. Mociardini; M. Sosio, *Nat. Prod. Rep.* **2007**, *24*, 1073.
237. M. A. Fischbach; C. T. Walsh, *Chem. Rev.* **2006**, *106*, 3468.
238. Y.-Q. Cheng; G.-L. Tang; B. Shen, *J. Bacteriol.* **2002**, *184*, 7103.
239. Y.-Q. Cheng; G.-L. Tang; B. Shen, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3149.
240. Y. A. Chan; A. M. Podevels; B. M. Kevany; M. G. Thomas, *Nat. Prod. Rep.* **2009**, *26*, 90.

Biographical Sketches

Alison M. Hill is a teaching fellow in the School of Biosciences of the University of Exeter. Her interest in polyketides began at the Australian National University where she conducted

her undergraduate research project on lichen polyketide metabolites with Professor Jack Elix. Alison then moved to the University of Cambridge to study the biosynthesis of aspyrone, a small fungal polyketide with Professor Jim Staunton, FRS. Postdoctoral work at Brown University, RI, USA with Professor David Cane on casbene synthase, a diterpene cyclase was followed by a return to the United Kingdom at the end of 1995. Alison joined the Department of Chemistry, King's College London as a lecturer and started work on soraphen A. In 2001 she moved to the University of Exeter, first in the Department of Chemistry where she was a research fellow and then to her current position.

Jim Staunton is emeritus professor of chemical biology in the Chemistry Department of the University of Cambridge, a Life Fellow of St Johns College, Cambridge, and a Fellow of the Royal Society. His research interests over the past 50 years have spanned organic synthesis, organic mechanisms, and tracer technology applied to the study of biosynthetic pathways. He has also made important contributions to technique development in organic analysis, especially NMR studies of tracer isotopes, and electrospray MS studies of organic molecules and proteins. In his academic research over the past 20 years, he has collaborated with Peter Leadlay on structural and mechanisitic studies of modular polyketide synthases. He is a joint founder with Peter Leadlay of Biotica Technology Ltd.

1.11 NRPS/PKS Hybrid Enzymes and Their Natural Products

Christopher M. Rath, Jamie B. Scaglione, Jeffrey D. Kittendorf, and David H. Sherman, Life Sciences Institute University of Michigan, Ann Arbor, MI, USA

© 2010 Elsevier Ltd. All rights reserved.

1.11.1	**Introduction**	454
1.11.1.1	Prototypical Biosynthesis of PK and NRP Natural Products	454
1.11.1.2	Biosynthesis of Hybrid PK/NRP Natural Products	458
1.11.2	**Medicinally Important PK/NRP Natural Products and Pathways**	458
1.11.2.1	Anticancer	458
1.11.2.1.1	Epothilones	458
1.11.2.1.2	Bleomycin	461
1.11.2.1.3	Curacin	462
1.11.2.1.4	FK228	463
1.11.2.1.5	Cryptophycins	464
1.11.2.2	Immunomodulatory	465
1.11.2.2.1	Rapamycin	465
1.11.2.2.2	FK506	465
1.11.2.3	Antimicrobial	466
1.11.2.3.1	Pristinamycin	466
1.11.2.3.2	Antibiotic TA	467
1.11.3	***Trans-AT Domain Pathways – a Rich Source of Unusual Biochemistry***	467
1.11.3.1	*Trans*-AT Hybrid PKS/NRPS Systems – an Introduction	467
1.11.3.1.1	Known *trans*-AT hybrid PK/NRP pathways	468
1.11.3.1.2	Biological activity and structure of *trans*-AT hybrid PK/NRPs	468
1.11.3.1.3	*In vivo* analysis of *trans*-AT hybrid PK/NRP systems	468
1.11.3.1.4	*In vitro* characterization of *trans*-AT hybrid PKS/NRPS pathways	468
1.11.3.1.5	Evolution, biology, and symbiosis of *trans*-AT hybrid PKS/NRPS systems	472
1.11.3.2	Anticancer	472
1.11.3.2.1	Albicidin	472
1.11.3.2.2	Chivosazol	473
1.11.3.2.3	Disorazol	473
1.11.3.2.4	Leinamycin	474
1.11.3.2.5	Onnamide/pederin	475
1.11.3.2.6	Rhizoxin	476
1.11.3.3	Antimicrobial	477
1.11.3.3.1	Antibiotic TA	477
1.11.3.3.2	Bacillaene	478
1.11.3.3.3	Lankacidin	479
1.11.3.3.4	Mycosubtilin	480
1.11.3.3.5	Virginiamycin M	480
1.11.3.4	Unknown Biological Activity	481
1.11.3.4.1	Thailandamide A	481
1.11.4	**Engineering, Technology Development, and Other Future Directions for Hybrid PK/NRP Systems**	481
1.11.4.1	Overview	481
1.11.4.2	Heterologous Expression and Pathway Engineering in Hybrid PK/NRP Systems	482
1.11.4.3	Culture Conditions and Media Optimization in Hybrid PK/NRP Systems	482

1.11.4.4 DNA-Sequencing Strategies in Hybrid PK/NRP Systems 482
1.11.4.5 Mass Spectrometry in Hybrid PK/NRP Systems 483
1.11.4.6 Structural Biology in Hybrid PK/NRP Systems 483
1.11.5 Conclusions 483
References 485

1.11.1 Introduction

Throughout the course of evolution, microorganisms have developed the capacity to generate a myriad of structurally diverse, complex organic compounds through dedicated secondary metabolic pathways. Generally, these systems are not expressed during the normal growth and proliferation of the organism; hence, the compounds they produce are largely considered to be nonessential to the organism. However, under environmentally stressful conditions, these pathways are induced to synthesize a multitude of structurally complex metabolites that are often toxic toward competing organisms, both unicellular and multicellular, occupying the same environmental niche. As such, the biosynthesis of natural products is thought to provide a competitive advantage that is critical to the survival of the producing organism.[1]

It is widely recognized that the inherent activities of natural products can be exploited for the benefit of human health.[2,3] Currently, there are many examples of natural product-derived pharmaceuticals that are employed to battle human disease.[4] Moreover, despite the technological advances achieved within pharmaceutical discovery, natural products continue to be the greatest source of new drug leads.[5] Of these, compounds classified as polyketides (PKs), nonribosomal peptides (NRPs), and hybrid PKs/NRPs have demonstrated a remarkable therapeutic potential (Figure 1). These secondary metabolites, synthesized by many microorganisms, fungi, and plants, display a vast structural diversity and array of pharmacological activities. At present, a variety of natural products from these three related compound classes find clinical utility as antibiotics, antiparasitic agents, antifungals, anticancer drugs, and immunosuppressants.[5]

The impressive chemical diversity and wide-ranging biological activities displayed by PK, NRP, and PK/NRP natural products have stimulated research programs, both in industry and in academia, aimed at discovering novel classes of these metabolites that may possess commercial therapeutic value. Historically, much of this effort has been focused on compound discovery, characterization, and biological screening. In more recent years, the rapidly advancing field of genomics has influenced natural product research, driving efforts aimed at deciphering the genetic blueprints that encode the pathways responsible for the production of PK, NRP, and PK/NRPs compounds.[6–10] This genomics-driven approach has led to the identity of a set of exquisite catalytic tools whose activities can be leveraged for the production of desirable compounds through mutasynthesis, chemoenzymatic synthesis, or semisynthesis.[11] Moreover, it has inspired the emerging field of 'combinatorial biosynthesis,' whereby biosynthetic enzymes are reengineered in combinatorial fashion for the construction of *de novo* biosynthetic pathways that are capable of generating novel compounds less effectively accessed by traditional chemical methods.[12–16]

Throughout the past decade, tremendous strides have been achieved not only in the identification of novel PK, NRP, and hybrid PK/NRP natural products but also in the elucidation of their biosynthetic pathways. The culmination of these efforts has resulted in an enhanced understanding of how these complex metabolites are assembled from relatively simple chemical building blocks. Moreover, these efforts have led to the discovery of several lead compounds that have entered pharmaceutical development. Much of this work, with respect to both PK and NRP natural products, has been summarized elsewhere and the reader is directed to one of the many well-written review articles for in-depth information on these topics.[17–21] Herein, this chapter is dedicated to hybrid PK/NRP natural products and the extraordinary enzymes that are responsible for their production.

1.11.1.1 Prototypical Biosynthesis of PK and NRP Natural Products

Although PKs are derived from sequential condensations of acetate, propionate, and other selected short-chain carboxylic acid precursors and NRPs are composed of amino acids, both proteinogenic and nonproteinogenic,

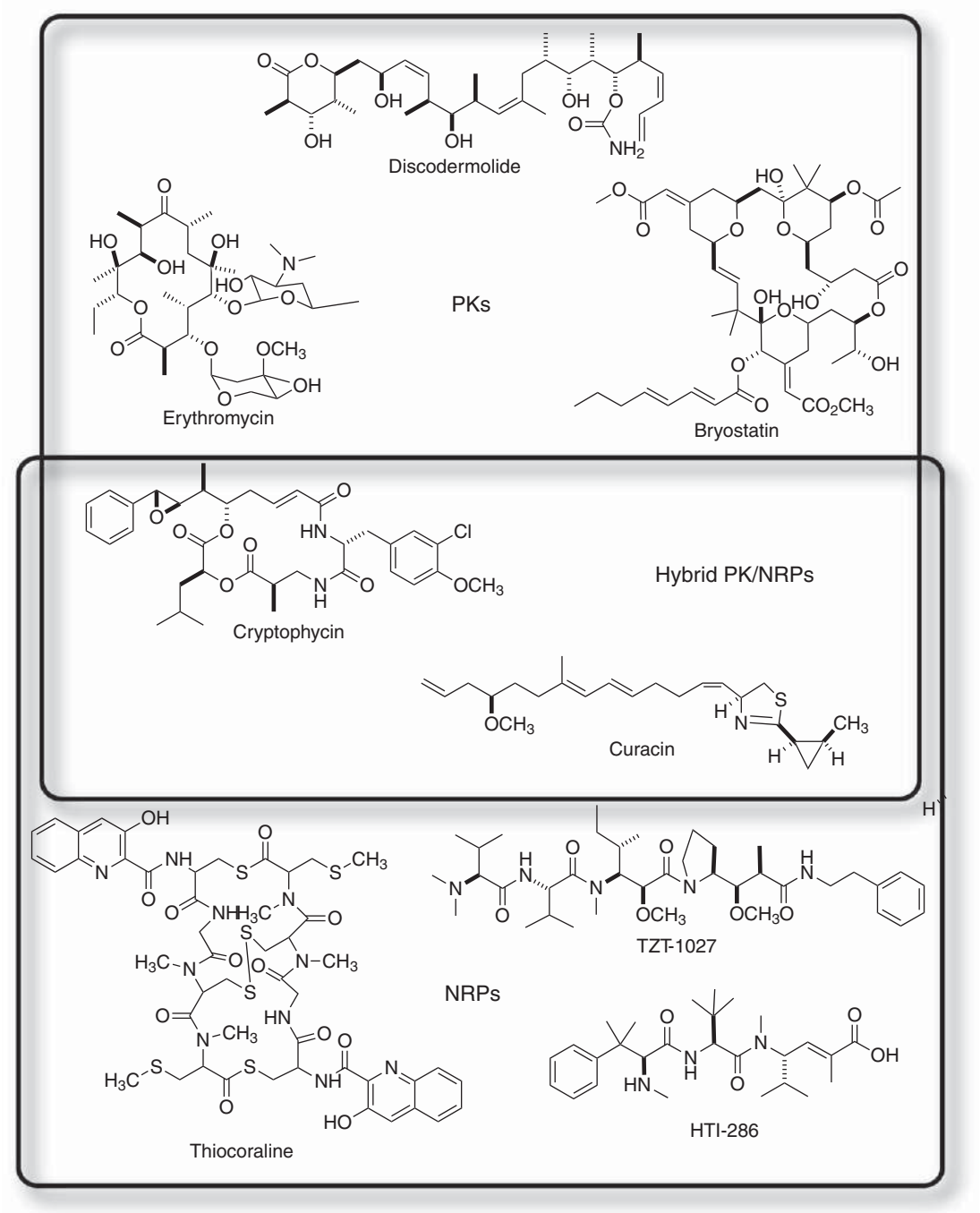


Figure 1 Examples of polyketide (PK), nonribosomal peptide (NRP), and hybrid PK/NRP natural products.

hybrid PK/NRP natural products are composed of a mixture of both short-chain carboxylic acids and amino acids. As such, it is perhaps not surprising to find that the biosynthetic machinery responsible for assembling PK/NRP natural products contains components of both PK biosynthesis and NRP biosynthesis. Hence, we begin our discussion on hybrid PK/NRP compounds with a primer on the biosynthesis of PK and NRP natural products.

The biosynthesis of both PKs and NRPs occurs using a stepwise, assembly line-type mechanism that is catalyzed by type I modular PK synthases (PKSs) and NRP synthases (NRPSs), respectively. Each of these biosynthetic systems is modular in organization, whereby each module contains a set of catalytic activities that are responsible for catalyzing a single elongation cycle and subsequent processing of the intermediate (**Figure 2**). Typically, these systems are considered to be colinear, whereby the organization of the biosynthetic genes, and thus the proteins they encode, is reflective of the order of assembly of the final product. Hence, the composition, number, and sequential arrangement of elongation modules found within a specific PKS or NRPS biosynthetic system often dictate the size and chemical structure of the mature natural product compound. There are, of course, exceptions to this generalization. Examples exist where the pathway genetic architecture does not directly correlate with the assembly order of the natural product.[22] Furthermore, a recent work has identified several pathways, many of which produce PK/NRP hybrids that deviate at the protein level from the colinearity rule. Biosynthetic pathways are known that skip entire elongation modules, iteratively employ modules for multiple elongations, possess an altered arrangement of catalytic domains, and utilize domains *in trans* in the generation of PK/NRP natural products. Several of these pathways are described in detail within this chapter.

At a minimum, each PKS module consists of three core protein domains, an acyltransferase (AT), ketosynthase (KS), and an acyl-carrier protein (ACP) that are absolutely required for PK elongation. The AT domain selects a specific acyl-coenzyme A (acyl-CoA) extender unit (usually malonyl-CoA or methylmalonyl-CoA) from the cellular CoA pool and covalently attaches it through a thioester bond to the ACP domain. The ACP domain serves as a scaffold for PK synthesis, responsible for anchoring the growing PK intermediate and shuttling it to the various active sites within the elongation module. Critical to this function, an active-site serine residue of the ACP domain is posttranslationally modified with the CoA-derived 4′-phosphopantetheine prosthetic group, which serves as the tether for the extender unit. Once the ACP domain is acylated by the AT, the KS domain catalyzes a Claisen-type decarboxylative condensation of the ACP-bound extender unit with the growing PK chain from the preceding module. KS catalysis results in the generation of an ACP-bound β-ketoacyl product, which eventually becomes the substrate for the subsequent elongation module. In addition to these three core domains, each elongation module may contain up to three additional domains that are responsible for reductive processing of the β-keto functionality before the next elongation step. These reductive steps contribute greatly to the overall structural diversity that has been observed among PK natural products. The presence of a ketoreductase (KR) domain gives rise to β-hydroxyl functionality and the presence of both a KR and a dehydratase (DH) domain generates an alkene, while the combination of KR, DH, and enoylreductase (ER) results in complete reduction to the alkane. Finally, a thioesterase (TE) domain, located at the C-terminus of the final elongation module, typically catalyzes the termination of PK biosynthesis, often resulting in the formation of a macrolactone structure.

Similar to PKS systems, NRPS modules also contain three core domains, an adenylation (A) domain, a condensation (C) domain, and a thiolation (T) domain, that act in concert to polymerize amino acid building blocks through amide linkages. Similar to the AT domains of PKS systems, the A domain selects an appropriate amino acid unit and activates it via an ATP-dependent adenylation reaction. Following activation, the A domain transfers the amino acid to the peptidyl carrier protein (PCP) (T) domain. This domain is equivalent to ACP domain in PKS biosynthetic systems, and as such, serves as the scaffold for NRP biosynthesis via an active-site serine residue that is modified with a phosphopantetheine moiety. The C domain catalyzes the peptide bond formation between the PCP-bound amino acid and the growing peptide chain from the preceding NRPS module. As with PKS systems, there are several additional catalytic domains that can reside within NRPS elongation modules that increase structural diversity. A heterocyclization domain (Cy) can, in certain cases, replace a C domain, and is responsible for both peptide bond formation and subsequent cyclization of cysteine, serine, or threonine residues into thiazoline or oxazoline rings. Oxidative (Ox) or reductive (R) domains often serve to further modify the heterocyclic rings to produce the corresponding thiazole/oxazole or thiazolidine/oxazolidine, respectively. Epimerization (E) domains are often found within elongation modules and epimerize the carrier protein thioester of the loaded amino acid. In addition, methyltransferases (MTs) are commonly encountered within NRPS modules to catalyze the N- or C-methylation of amino acid residues, a modification that makes the peptide less susceptible to proteolytic cleavage. Finally, as in PK biosynthesis, a TE domain contained within the final elongation module

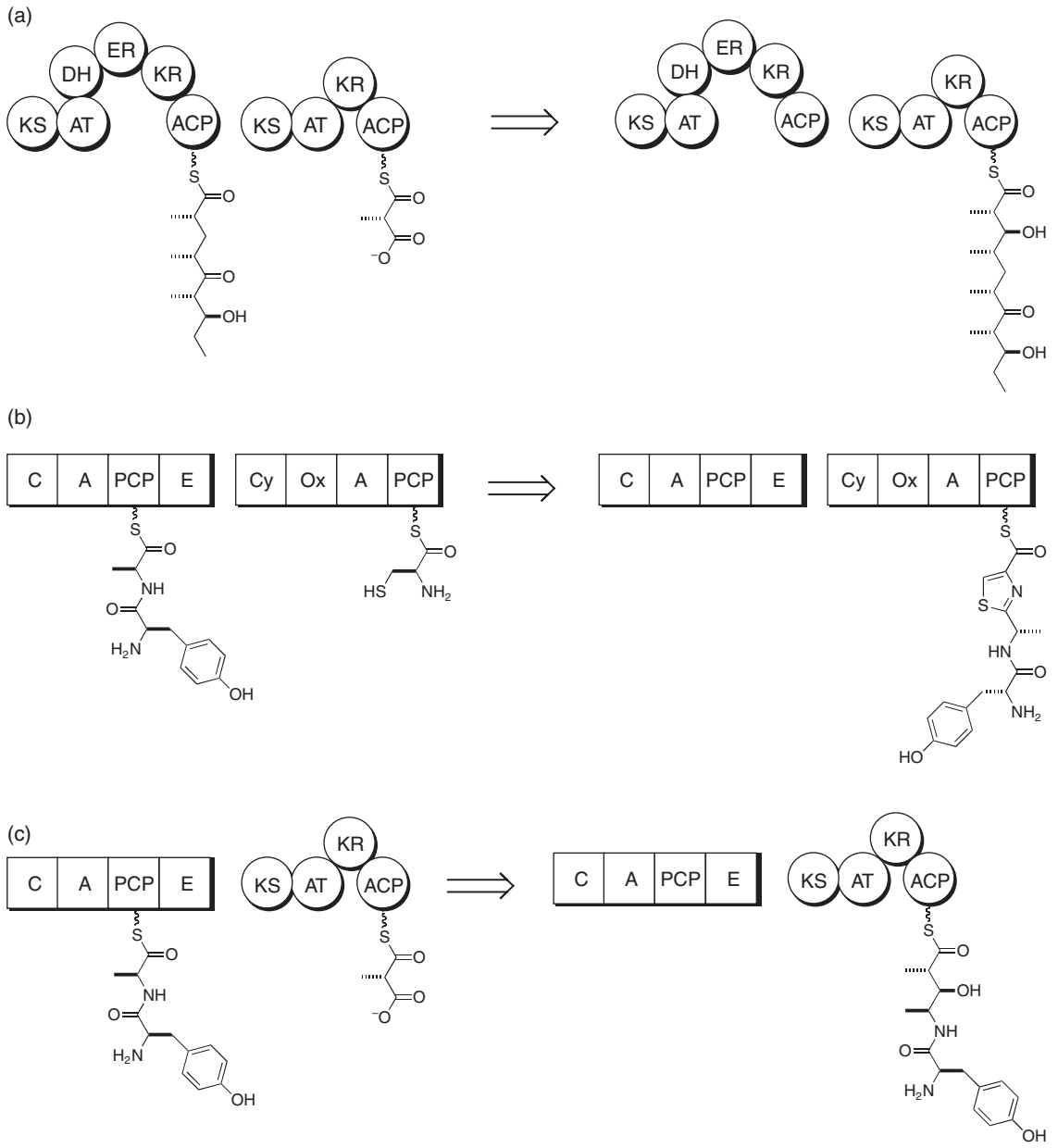


Figure 2 Hypothetical examples of the modular organization in polyketide synthases, nonribosomal peptide synthases, and hybrid polyketide/nonribosomal peptide synthases. (a) Two consecutive PKS elongation modules from a hypothetical polyketide biosynthetic pathway that catalyze the elongation of the growing polyketide intermediate by two carbons (from methylmalonyl-CoA), and subsequent β-keto reduction. (b) Two consecutive NRPS elongation modules from a hypothetical nonribosomal peptide biosynthetic pathway that catalyzes peptide bond formation between the growing peptide intermediate and an activated cysteine residue with subsequent cyclization/oxidation of the incorporated cysteine residue to a thiazole. (c) Consecutive NRPS and PKS elongation modules from a hypothetical hybrid nonribosomal peptide/polyketide biosynthetic pathway that catalyze the two-carbon extension (from methylmalonyl-CoA) of the peptide intermediate and subsequent β-keto reduction.

terminates NRP biosynthesis. This domain is responsible for liberation of the peptide intermediate from the biosynthetic machinery to generate either a linear or macrocyclic product. A variety of additional tailoring or loading modules may also be present before, within, or after the PKS or NRPS modules to introduce additional structural diversity.

1.11.1.2 Biosynthesis of Hybrid PK/NRP Natural Products

Although the metabolites that are composed of both amino acids and short-chain carboxylic acids, the so-called hybrid NRP/PKs, have been known for many years, it is only within the last decade that the mechanisms of their biosynthesis have been elucidated. Shen *et al.*[23] first proposed the idea that the bleomycins, a family of natural products that possess potent anticancer activity, might be constructed by a hybrid megasynthase that contains catalytic components of both NRPS and PKS biosynthetic machinery. The subsequent sequencing and cloning of the bleomycin biosynthetic gene cluster indeed confirmed the hypothesis that a merger of NRPS and PKS systems results in a hybrid megasynthase that is responsible for bleomycin production.[23,24] Since this initial discovery, many hybrid PK/NRP biosynthetic pathways have been identified.[25] As described below, these pathways often contain impressive biochemical catalysts that generate structurally diverse and clinically important natural products, many of which have been developed into human medicine or are currently being investigated in clinical trials. Broadly, these compounds fall into three categories: anticancer, immunomodulatory, and antimicrobial.

1.11.2 Medicinally Important PK/NRP Natural Products and Pathways

1.11.2.1 Anticancer

1.11.2.1.1 Epothilones

1.11.2.1.1(i) Isolation and biological activity Epothilones A and B (Figure 3) were originally discovered in 1987 as bioactive components in extracts of the cellulose-degrading myxobacterium *Sorangium cellulosum* So ce90 during a screen for new antifungal agents by Reichenbach, Hofle, and coworkers at the Gesellschaft fur Biotechnologische (GBF) in Braunschweig, Germany.[26,27] The compounds were termed 'epothilones' to reflect their basic structural features, which include an epoxide moiety, a thiazole-containing side chain, and a single keto(ne) function. In 1995, the compounds were rediscovered by Bollag *et al.*[28] at Merck Research Laboratories during a target-oriented screen for paclitaxel (Taxol)-like tubulin-stabilizing compounds.[29] Similar to Taxol, epothilones A and B were able to promote GTP-independent tubulin polymerization, resulting in microtubule stabilization.[28] *In vitro* studies demonstrated that the epothilones inhibited the growth of cultured cancer cells at low drug concentrations resulting from mitotic arrest and subsequent apoptosis.[28] Drug-binding competition studies revealed that the epothilones competitively inhibited the binding of [^{3}H]Taxol to microtubules, suggesting that both drugs interact with the same, or overlapping, sites in β-tubulin. Most impressively, the epothilones are neither substrates nor inducers of P-glycoprotein (PGP) and thus maintain potency in cells that are resistant to paclitaxel or other anticancer agents.[30]

Although the major products originally isolated from fermentation broths of *S. cellulosum* So ce90 were epothilones A and B, several other related members of this class of natural products have since been discovered as minor products, such as epothilones C and D (**Figure 3**).[31–33] Following determination of the absolute configurations of epothilones A and B through a combination of X-ray crystallography and chemical degradation studies,[26] extensive efforts were initiated to synthesize the natural epothilones and to create a variety of analogues for structure–activity relationship (SAR) studies.[34–38] As a result of these efforts, in late 2007 the US Food and Drug Administration approved the lactam analogue of epothilone B (Ixempra, ixabepilone, BMS-247550, Bristol–Myers Squibb) for use in patients with metastatic or locally advanced breast cancer.[39–45] Furthermore, there are at least six additional epothilone-inspired compounds that are currently undergoing clinical evaluation as potential anticancer agents.[46–49] These include the development of the natural products epothilone B by Novartis (EPO906, patupilone),[50–53] epothilone D by Kosan/Roche (deoxyepothilone B, KOS-862),[46,54–56] a C21-amino-epothilone B analogue by Bristol–Myers Squibb (BMS-310705),[57,58] a C20-desmethyl-C20-methylsulfanyl-epothilone B analogue by Novartis (ABJ879), a 9,10-didehydro-epothilone B analogue at Kosan/Roche (KOS-1584), and a synthetic analogue, ZK-EPO, developed by Schering AG in Berlin.[59]

1.11.2.1.1(ii) Gene cloning, sequence analysis The epothilone biosynthetic gene cluster (**Figure 4**) from myxobacterium *S. cellulosum* was independently cloned by Molnar *et al.* and Tang *et al.*[60–62] Molnar *et al.*[61] screened a bacterial artificial chromosome (BAC) library of *S. cellulosum* So ce90 with heterologous DNA probes

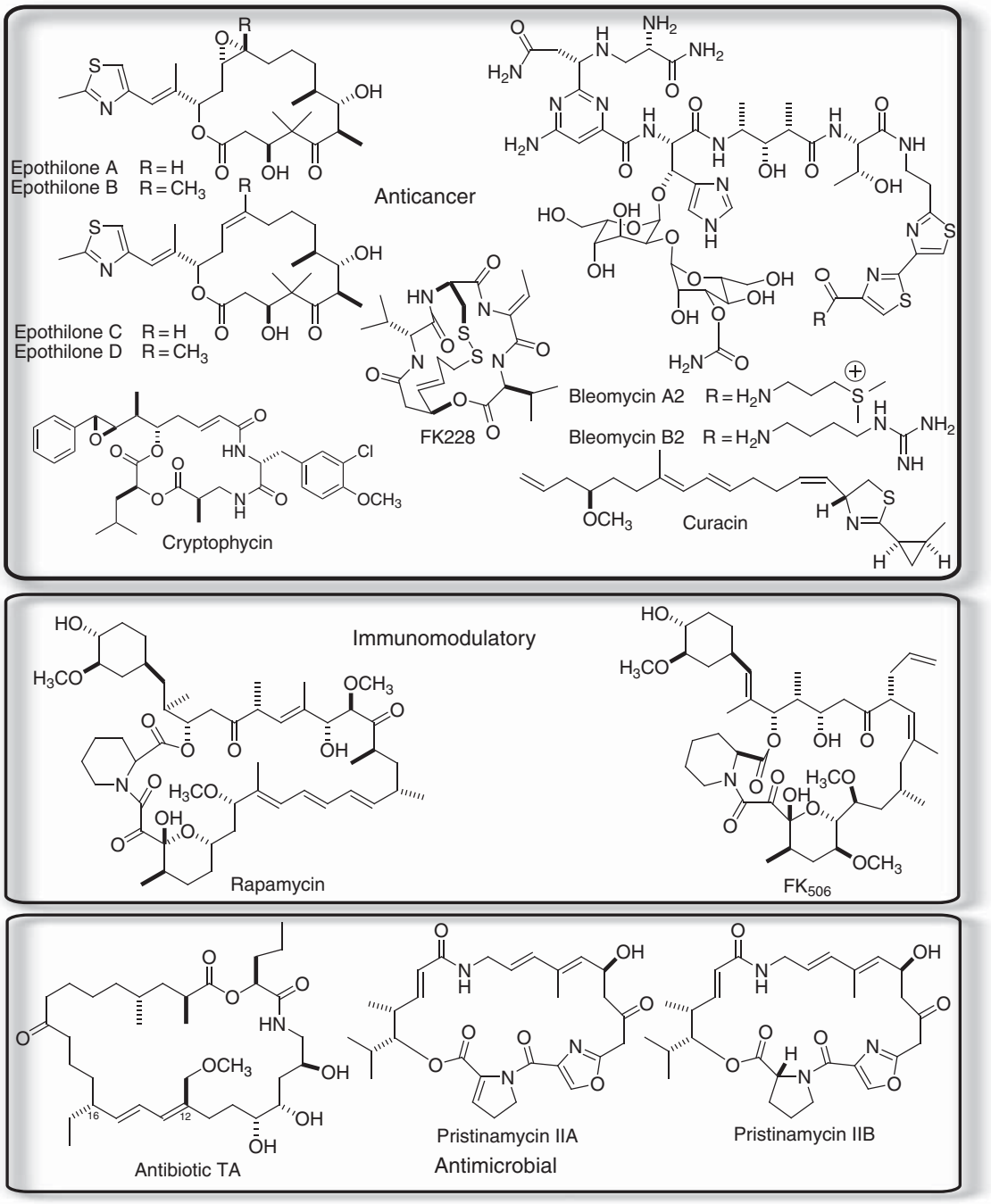


Figure 3 Structures of medicinally important hybrid PKS/NRPS products. Compounds are grouped by primary bioactivity.

derived from the keto synthase (KS) regions of other PKS gene clusters. Tang *et al.*[62] used degenerate KS polymerase chain reaction (PCR)-generated hybridization probes to isolate overlapping cosmid clones from a genomic library of *S. cellulosum* SMP44.[60]

The epothilone gene cluster contains five type I PKS genes (*epo*A, *epo*C, *epo*D, *epo*E, and *epo*F) that encode nine PKS modules and one NRPS gene (*epo*B) that encodes a single NRPS elongation module. In addition, the epothilone gene cluster contains one cytochrome P-450 gene, *epo*K. Epothilone biosynthesis is initiated

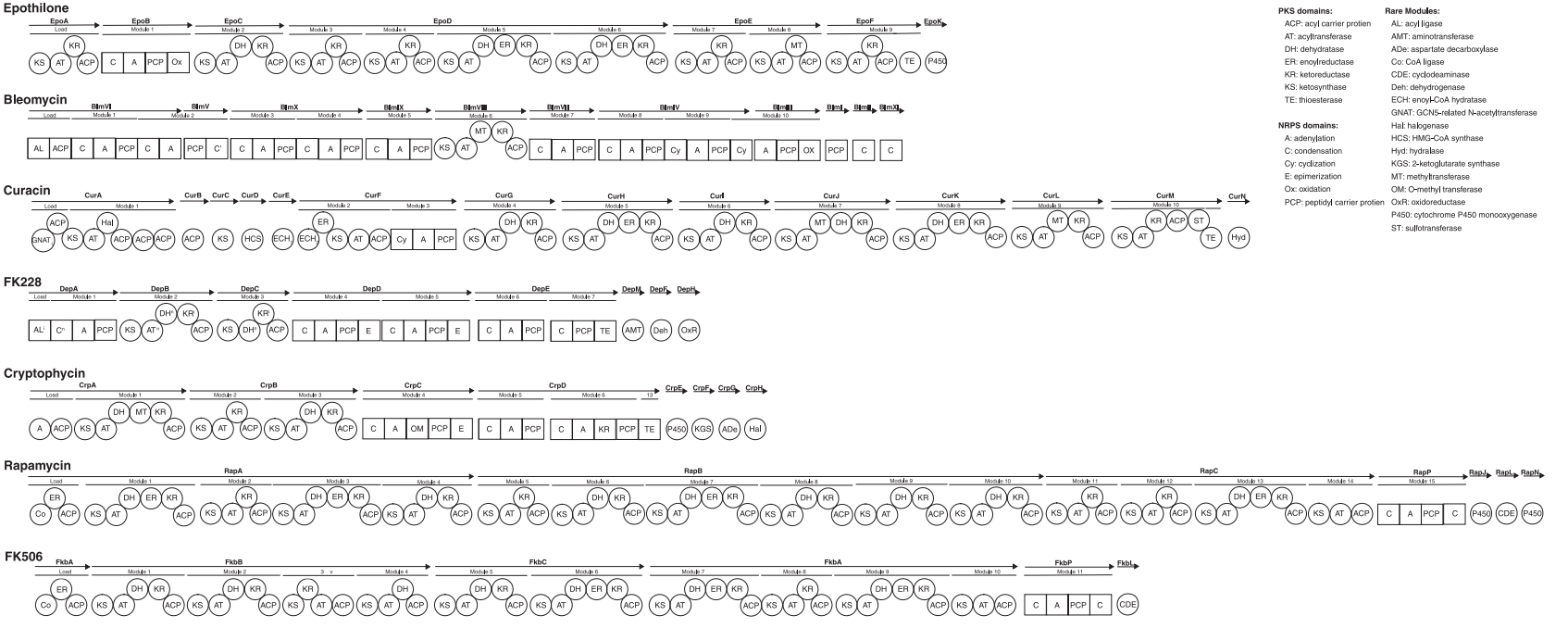


Figure 4 Biosynthetic pathways of medicinally important hybrid PKS/NRPS products. Pathways are arranged based on predicted biosynthetic schemes, with modules and domains labeled. See **Figure 7** for the antibiotic TA biosynthetic pathway.

at epothilone A, a polypeptide consisting of four PKS domains: KS^Y, AT, ER, and ACP. The AT selects malonyl-CoA from the cellular CoA pool and loads it onto the ACP domain to give a malonyl–ACP intermediate. The KS^Y domain, in which the active site cysteine is replaced by a tyrosine (Y),[63–66] is presumed to be analogous to the biochemically characterized KS^Q domains, responsible for decarboxylating primed malonyl-ACP into an acetyl-ACP starting unit.[67]

Following the EpoA PKS module is the EpoB NRPS module that contains four domains: Cy, A, Ox, and PCP. The EpoB NRPS module is responsible for cysteine incorporation and condensation with the acetyl group on the upstream EpoA PKS-loading module, with a subsequent intramolecular cyclization, dehydration, and oxidation to form the thiazole moiety. Together EpoA and EpoB catalyze the formation of the EpoB 2-methylthiazolylcarboxy-S-PCP intermediate that is elongated by EpoC and the seven remaining PKS modules to provide the epothilone 16-membered ring macrolactone backbone. Sequence analysis of AT domains in the epothilone PKS modules suggests that modules 3, 5, and 9 utilize malonyl-CoA, whereas modules 2, 6, 7, and 8 utilize methylmalonyl-CoA as an extender unit. The AT domain of module 4 has an unusual alanine in position of the critical residue for substrate specificity, which is suggested to confer the relaxed specificity to module 4 for recognizing both malonyl-CoA and methylmalonyl-CoA extender units, as evidenced by the production of both epothilone A and epothilone B. A TE domain, located at the terminus of EpoF, catalyzes off-loading and macrolactonization of the epothilonyl-S-ACP linear intermediate to generate epothilone C or epothilone D. Finally, the cytochrome P-450 monooxygenase, EpoK,[68,69] catalyzes epoxidation of epothilone C or epothilone D to give epothilone A or epothilone B. Note that the target for epoxidation is a striking *cis*-alkene, as is present in epothilones C and D.

EpoA–EpoF constitutes a natural hybrid PKS/NRPS/PKS megasynthetase. The modular architecture lends itself to modifications through combinatorial biosynthesis to create improved epothilones and provide novel building blocks for *de novo* PKS engineering.

1.11.2.1.2 Bleomycin

1.11.2.1.2(i) Isolation and biological activity

The bleomycins, such as bleomycinic acid, bleomycin A2, and bleomycin B2 (**Figure 3**), are a family of glycopeptide-derived antibiotics originally isolated from *Streptomyces verticillus* by Umezawa *et al.*[70] Through a series of chemical degradation studies coupled with X-ray crystallographic analysis of the products, the structure of bleomycin A2, complete with the assignment of its relative and absolute stereochemistry, was originally established.[71,72] The bleomycin structure was revised in 1978[73] to correct the structural assignment of the N-terminus and was successively confirmed by total synthesis in 1982.[74,75] Feeding studies suggested that the compound is derived from nine amino acid residues (two asparagines, histidine, alanine, threonine, serine, β-alanine, and two cysteines), one acetate moiety, and two *S*-adenosylmethionine units. Several structural variations of the naturally occurring bleomycins, differing primarily at the C-terminus of the glycopeptide, have been identified from fermentation broths.[76]

Bleomycins exhibit potent antitumor activity and are currently used clinically (Blenoxane, Bristol–Myers Squibb) in combination with a number of other agents for the treatment of several types of tumors, notably squamous cell carcinomas and malignant lymphomas. The drugs exert their biological effects through a metal-dependent, oxidative, sequence-selective DNA and RNA binding and degradation process. Although both single- and double-stranded cleavage are observed, the latter has been regarded by many as the biologically more important event.[77–79]

Each structural unit of bleomycin contributes functionally to its biological activity. Extensive structural, chemical, biophysical, and biological studies on bleomycin A2 and its derivatives have dissected the bleomycins into four domains: (1) the N-terminal pyrimidioblamic acid subunit, including its β-aminoalanine amide side chain and the linked β-hydroxy-L-histidine provide the coordination sites required for Fe(II) complexation and molecular oxygen activation; (2) the DNA-binding affinity originates from the C-terminus with the bithiazole and C-terminal amine providing the majority of the affinity; (3) the 4-amino-3-hydroxy-2-methylpentanoic acid subunit not only provides the connectivity between the metal-binding and DNA-binding sites but also plays an important role in potentiating the cytotoxicity and DNA cleavage efficiency of bleomycins;[80] and (4) the glycoside domain has been less extensively examined, but is believed to participate in cell recognition by bleomycins and possibly in cellular uptake and metal-ion coordination.[81–86] Bleomycin is unique among anticancer drugs in that it does not cause myelosuppression, enabling its wide application in combination

chemotherapy. However, the major clinical limitations of bleomycins have been the early development of drug resistance and cumulative pulmonary toxicity.[87,88] Thus, there have been ongoing attempts to isolate or develop new bleomycin analogues with better therapeutic efficacy and lower toxicity, as well as further define the functional roles of the individual domains.[89,90]

1.11.2.1.2(ii) Gene cloning, sequence analysis In the mid-1990s, two bleomycin resistance genes, *blm*A and *blm*B, were cloned from *S. verticillus* ATCC 15003.[91] Subsequently, Shen and coworkers constructed a genomic library from this microorganism and were ultimately able to identify and sequence the bleomycin gene cluster as an 85-kb region of DNA upstream of the *blm*AB locus.[92] The cluster was found to contain one PKS gene (*blm*VIII), ten NRPS genes (*blm*X, *blm*IX, *blm*VII, *blm*VI, *blm*V, *blm*IV, *blm*III, *blm*I, *blm*II, and *blm*XI), five sugar biosynthetic genes (*blm*C, *blm*D, *blm*E, *blm*F, and *blm*G), and several other genes believed to be involved in regulation, resistance, and product tailoring. However, unlike most NRPS and PKS biosynthetic clusters, the genes are not arranged colinearly on the chromosome. Shen and coworkers have proposed a model to explain the organization of the Blm megasynthetase complex (**Figure 4**). Biochemical studies on the Blm biosynthetic system have been limited to the characterization of an Ox domain responsible for conversion of a thiazoline into a thiazole and identification of the phosphopantetheinyl transferase (PPT) required for functional NRPS and PKS activity.[24]

1.11.2.1.3 Curacin

1.11.2.1.3(i) Isolation and biological activity Curacin A (**Figure 3**) is a mixed PK/NRP natural product with potent antiproliferative and antimitotic activity against colon, renal, and breast-cancer-derived cell lines.[93] The compound was originally isolated from strains of the tropical marine cyanobacterium *Lyngbya majuscula* discovered in Curaçao by Gerwick *et al.*[94] and was found to possess unusual structural features, including a cyclopropane group, thiazoline moiety, *cis*-alkenyl group, and terminal double bond. Curacin A has been shown to block cell cycle progression by interacting with the colchicine-binding site on tubulin and inhibiting microtubule polymerization.[95] The clinical development of curacin has been hindered by its high lipophilicity; however, structural analogues having improved water solubility and potency have been recently synthesized to enable continued preclinical studies.[96,97]

1.11.2.1.3(ii) Gene cloning, sequence analysis Together, the Sherman and Gerwick laboratories have conducted a series of isotope-labeled precursor feeding and NMR studies that established the metabolic origin of all curacin A atoms and their order of assembly.[98] The studies indicated that the compound is composed of one cysteine residue, ten acetate units, and two *S*-adenosylmethionine-derived methyl groups, thus suggesting that curacin A was of mixed PKS–NRPS origin. Through the creation and screening of a cosmid library from *L. majuscula* using a general PKS probe, a 64-kb metabolic gene cluster containing 14 open-reading frames (ORFs) was identified. As predicted by the precursor incorporation experiments, the biosynthetic gene cluster (**Figure 4**) was found to contain nine PKS modules and one NRPS module. Interestingly, this biosynthetic system is unique in that all PKS modules, with the exception of the CurF hybrid PKS/NRPS module, are monomodular.[98]

Curacin A biosynthesis is initiated by the unique CurA PKS. Bioinformatic analysis of the AT domain located at the N-terminus of CurA indicated homology with the GCN5-related *N*-acetyltransferase (GNAT) domain, PedI, from the putative pederin gene cluster.[98,99] Interestingly, a recent study has revealed an astonishing biochemical chain-initiation strategy for the loading module of curacin A that involves an unusual loading tridomain found at the N-terminus of CurA. This tridomain comprises of an adapter domain, a GNAT domain, and an ACP domain. *In vitro* biochemical studies of the isolated tridomain have shown that the GNAT has unprecedented bifunctional activity, because it is capable of decarboxylating malonyl-CoA to acetyl-CoA first and then directing the transfer from acetyl-CoA onto the ACP domain to produce the acetyl-ACP intermediate.[100]

Three tandem ACP domains reside at the C-terminus of the CurA polypeptide that together with four ORFs encoding CurB–CurE, as well as the first two domains of CurF, was predicted to direct formation of the unique cyclopropyl ring in curacin A. Indeed, recent biochemical and structural studies from the Hakansson, Gerwick, Smith, and Sherman laboratories confirm that the CurE/CurF ECH_1–ECH_2 enzyme pair catalyze

the successive dehydration and decarboxylation of (*S*)-3-hydroxy-3-methylglutarate (HMG)-ACP to generate a 3-methylcrotonyl-ACP intermediate for subsequent formation of the cyclopropane ring.[101,102]

The remainder of the molecule is assembled by seven PKS monomodules, CurG–CurM, that catalyze seven successive rounds of condensation with malonyl-CoA extender units. Furthermore, embedded MT domains in CurJ and CurL are predicted to catalyze transfer of the C-17 and O-13 methyl groups, respectively. Of final interest is the atypical biosynthetic termination mechanism that is predicted to function in both product release and decarboxylative dehydration to form the unusual terminal alkene. Like the majority of other known PK/NRP biosynthetic pathways, the final elongation module of the curacin pathway, CurM, contains a terminal TE domain that presumably plays a role in formation of the terminal olefin; however, biochemical evidence for such a role remains to be established. Interestingly, analysis of CurM also predicts the presence of a sulfotransferase (ST) domain immediately preceding the TE. ST domains are typically responsible for transferring a sulfuryl group from a donor molecule (such as 3′-phosphoadenosine-5′-phosphosulfate, PAPS) to a variety of acceptor carbohydrates, proteins, and other low-molecular-weight metabolites.[103] Although STs have been characterized from both eubacterial and eukaryotic organisms, the presence of an ST domain within a PKS system is unprecedented.

While the particular functions of the ST and TE domains in terminal olefin formation are still unknown, one possible mechanism currently under consideration involves the transfer of a sulfuryl group by the ST domain to the 3-hydroxyl group of the ACP-bound thioester chain, followed by hydrolytic termination of curacin A biosynthesis by the TE to produce the linear free acid. At this point, one of the several chemical steps can be envisaged. For example, the TE may also catalyze decarboxylation of the free acid, after which formation of the double bond would be energetically driven by displacement of the sulfate-leaving group. Alternatively, a yet unidentified enzyme might be responsible for decarboxylating the free acid generated by the TE domain. Finally, it is also conceivable that upon TE-catalyzed hydrolysis, the decarboxylation reaction occurs spontaneously as a result of the presence of the sulfate-leaving group at C3. Deciphering the mechanism of terminal alkene formation in the curacin biosynthetic pathway is currently ongoing in our laboratory.

1.11.2.1.4 FK228

1.11.2.1.4(i) Isolation and biological activity FK228 (also known as FR901228, depsipeptide, NSC630176, or Romidepsin) was originally isolated as an antitumor antibiotic from the fermentation broth of *Chromobacterium violaceum* No. 968.[104] The compound demonstrated potent *in vitro* and *in vivo* antitumor activity against both murine and human tumor cell lines,[105] and has since shown activity against many members of the US National Cancer Institute standard panel of 60 cell lines.[106,107] While FK228 was originally identified as an anti-*ras* agent,[104,105] it was later found to interfere with mitogen-induced signaling pathways, and today is classified as a potent histone deacetylase (HDAC) inhibitor.[108]

The structure of FK228 (**Figure 3**) was determined by spectroscopic and X-ray crystallographic analyses,[109] and later confirmed by total synthesis.[110] The compound is structurally unique among HDAC inhibitors, consisting of a bicyclic depsipeptide core that features one 16-membered macrolactone ring bearing an ester linkage and one 17-membered macrolactone ring bearing both an ester linkage and a disulfide bond. This disulfide bond of the 17-membered macrolactone enables the novel cytotoxic mechanism of FK228. Essentially, the disulfide form of FK228 acts as a stable prodrug, that in the presence of cellular reducing agents is rapidly converted into a compound containing two free sulfhydryl groups.[111] The liberated sulfhydryl on the longer aliphatic tail of reduced FK228 lines the binding-site pocket of class I HDACS, interacting with active-site zinc, and preventing substrate access. Due to its remarkable nanomolar IC_{50} and selective cytotoxicity for inducing apoptosis in cells from patients with chronic lymphocytic leukemia (the most common leukemia in the Western hemisphere),[112] Romedepsin (Gloucester Pharmaceuticals) is currently in phase 2b clinical trials for use in both cutaneous and peripheral T-cell lymphoma.[106]

1.11.2.1.4(ii) Gene cloning, sequence analysis Through the creation, sequencing, and scanning of a genome sampling library from *C. violaceum* No. 968 for genes encoding an apparent NRPS or PKS, a 36.4-kb gene cluster for FK228 biosynthesis has recently been identified, cloned, and partially characterized.[113] The FK228 biosynthetic gene cluster (**Figure 4**) is predicted to encompass 14 genes, designated *dep*A–*dep*N, with nine of these gene products thought to participate in PK/NRP assembly and tailoring of FK228. The remaining

five genes are presumed to have regulatory functions. Like many other PK/NRP natural products, the elongation modules are arranged in a colinear fashion. FK228 biosynthesis commences with the activation of a cysteine residue by the A domain of the NRPS DepA and subsequent loading onto the PCP. Next, the aminotransferase (AMT), DepM, acts *in trans* to remove an amino group from the PCP-bound intermediate, generating 4-mercaptobutanyl-*S*-PCP. The PKS monomodules DepB and DepC elongate this intermediate with two C_2 units derived from malonyl-CoA. Following elongation by DepB, the FadE2-like acyl-CoA dehydrogenase, DepF, is hypothesized to install a double bond on the β-hydroxyl-5-mercaptopentanoyl-*S*-ACP intermediate, functionally serving as an ER domain. The final two bimodular NRPS polypeptides, DepD and DepE, are expected to extend the growing intermediate chain with activated D-Val, D-Cys, 2,3-dehydrothreonine, and Val. Interestingly, module 7 of DepE lacks an A domain. It has been proposed, although not experimentally verified, that the A domain of module 4 acts *in trans* to aminoacylate the PCP domain in module 7. Finally, DepH, a proposed flavin adenine dinucleotide-dependent pyridine nucleotide-disulfide oxidoreductase (OxR), is believed to bring together the two free sulfhydryl groups and catalyze the intramolecular disulfide bond formation.[113]

1.11.2.1.5 Cryptophycins

1.11.2.1.5(i) Isolation and biological activity Cryptophycins, a large class of peptolides, were originally isolated from the cyanobacteria *Nostoc* sp. ATCC 53789 by researchers at Merck as a potent fungicide. A gross structure was proposed, but Merck abandoned the project because the compounds were highly toxic to be developed as antifungals.[114] Several years later, interest in the cryptophycins was renewed when a screen of the lipophilic extract of *Nostoc* sp. GSV 224 exhibited potent cytotoxic activity.[115] This activity was attributed to the cryptophycin natural products, which have since been found to have antimitotic activity and cytotoxicity toward tumor cells in culture, as well as anticancer activity against murine solid tumor models and human tumor xenografts.[116–118] While there are more than 25 naturally occurring analogues, the major compound from both *Nostoc* sp. ATCC 53789 and *Nostoc* sp. GSV 224, cryptophycin 1 (**Figure 3**), consists of four subunits: δ-hydroxyoctenoic acid (Unit A), 3-chloro-*O*-methyl-D-tyrosine (Unit B), methyl-β-alanine (Unit C), and L-leucic acid (Unit D), linked in a cyclic clockwise sequence.[115,119,120] Other naturally occurring cryptophycins are analogues that differ from cryptophycin 1 by one or two of these subunits.

Cryptophycin 1 is one of the most potent tubulin-destabilizing agents ever discovered, resulting in cellular arrest at the G2/M phase via hyperphosphorylation of Bcl-2, thereby triggering the apoptotic cascade.[121] Cryptophycins are also attractive as chemotherapeutics because they are active against multidrug-resistant tumor cell lines and are not substrates for *p*-glycoprotein pumps.[118] A synthetic analogue, cryptophycin 52 (LY355703), was developed by Eli Lilly & Co., and ultimately reached phase 2 clinical trials; however, high production costs coupled with dose-limiting toxicity halted its development.[122] In spite of this, a subsequent phase 2 clinical trial involving patients with platinum-resistant advanced ovarian cancer concluded that the rate of disease stabilization in the absence of adverse events might justify further investigation of cryptophycin 52.[123] A second generation of cryptophycin 1 analogues with improved solubility properties has been synthesized and preclinical studies indicate a marked increase in efficacy against a variety of tumors.[124]

1.11.2.1.5(ii) Gene cloning, sequence analysis The Sherman and Moore laboratories collaboratively isolated and characterized the cryptophycin gene cluster (**Figure 4**) using a strategy that relied on comparative metabolomic analysis.[125] In this approach, the A and KS domain sequences of *Nostoc* sp. ATCC 53789 were compared to that of *Nostoc punctiforme* ATCC 29133, a strain that lacks cryptophycin productivity. This comparative method resulted in the identification of six A domain sequences that were present in *Nostoc* sp. ATCC 53789 but not in *N. punctiforme*. Of these six, a single A domain appeared to be a candidate for the cryptophycin pathway. From this initial lead, the 40-kb cryptophycin biosynthetic gene cluster was identified by cosmid library screening.[125]

The cryptophycin gene cluster consists of two modular PKS genes (*crp*A and *crp*B) and two modular NRPS genes (*crp*C and *crp*D). In total, these ORFs encode seven elongation modules that contain the requisite catalytic domains for assembly of the cryptophycin macrocyclic core structure. In addition, a series of ORFs, designated *crp*E–*crp*H, is located downstream of *crp*A–*crp*D, and is predicted to encode enzymes that modify the nascent macrocycle to yield cryptophycin 1. These include a cytochrome P-450 epoxidase (*crp*E), a putative

2-ketoglutarate-dependent enzyme (*crp*F), an aspartate decarboxylase (ADe) (*crp*G), and a flavin-dependent halogenase (Hal) (*crp*H). In addition to the characterization of the gene cluster, Sherman and coworkers have produced novel cryptophycin analogues by precursor-directed biosynthesis.[125] *In vitro* biochemical work has also been performed to characterize the TE domain of CrpD,[126,127] the ADe, CrpG,[128] and the P-450 epoxidase, CrpE.[129] These studies have unveiled the genetic blueprint of cryptophycin biosynthesis in *Nostoc* sp. ATCC 52789, thereby providing access to a set of catalytic tools for chemoenzymatic construction and modification of new cryptophycin analogues.

1.11.2.2 Immunomodulatory

1.11.2.2.1 Rapamycin

1.11.2.2.1(i) Isolation and biological activity Rapamycin (sirolimus, Rapamune, Wyeth) was isolated in 1975 from a species of *Streptomyces hygroscopicus* native to Easter Island (Rapa Nui).[130,131] Rapamycin's (**Figure 3**) 31-membered macrocyclic structure containing both lactam and lactone linkages was initially determined by X-ray crystallography, with complementary NMR and degradation studies reported shortly thereafter.[132,133] Despite the complex stereochemical and functional architecture of rapamycin, four total syntheses of the molecule have been reported.[134–137]

While rapamycin was initially described as an antifungal with weak antibiotic activity, and its antiproliferative properties have very recently been explored,[138] it is best known for its immunosuppressive action.[139] Like FK506 (described below), rapamycin binds FK506-binding protein (FKBP), and the rapamycin–FKBP complex acts in a downstream-signaling pathway by binding and inhibiting mammalian target of rapamycin (mTOR), a member of the phosphatidylinositol kinase-related kinase (PIKK) family. Because mTOR is responsible for the progression of cells from the G1 phase into the S phase, rapamycin effectively halts T-cell proliferation, inhibiting the immune response.[140] While the side effects of rapamycin are less serious than the calcineurin inhibitors FK506 or cyclosporin, it is often used in combination with cyclosporin for a synergistic immunosuppressive effect and was approved by the US FDA in 1999 (Rapamune, Wyeth).[141]

1.11.2.2.1(ii) Gene cloning, sequence analysis The structure of rapamycin suggested that it was mainly PK in origin, assembled from seven acetate and propionate units, with *O*-methyl groups installed by *S*-adenosylmethionine cofactors. The starting dehydroxycyclohexene carboxylate unit is derived by a reductive pathway from shikimic acid, whereas the pipecolate moiety comes from L-lysine.[142] The rapamycin gene cluster (**Figure 4**) was cloned by Leadlay and coworkers in 1995 using probes derived from erythromycin PKS genes, followed by subsequent chromosome walking.[143] The cluster was found to contain three PKS genes (*rap*A, *rap*B, *rap*C) that encode 15 PKS elongation modules. Also a monomodular NRPS gene, *rap*P, and several additional genes that encode tailoring, regulatory, or transport enzymes were identified. These include the RapJ and RapN cytochrome P-450s that are responsible for installing the C9 keto and C27 hydroxyl functionalities, as well as the cyclodeaminase (CDE), RapL, that catalyzes conversion of L-lysine into L-pipecolate.[143–145]

The combined RapA–RapC and RapP megasynthetase is one of the largest and most complex ever identified, containing 69 PKS and NRPS domains. The loading module requires three domains, including CoA ligase (Co), ER, and ACP.[143–145] The Co domain shows homology to ATP-dependent carboxylic acid-Cos and is responsible for converting the dihydroxycyclohexene carboxylate starter unit into an activated acyl unit for loading onto the ACP. Following ACP loading, the ER domain reduces this dihydroxycyclohexene starter unit to a dihydroxycyclohexane for subsequent PK chain elongation. Unlike other hybrid PKS/NRPS pathways, a TE domain does not catalyze rapamycin chain termination. Instead, after 14 cycles of PK chain elongation, the RapC ACP-bound intermediate is elongated with a pipecolate unit by the RapP NRPS.[144] The same RapP has an extra C domain at its C-terminus that catalyzes the formation of the ester bond to pipecolate, which results in product release.

1.11.2.2.2 FK506

1.11.2.2.2(i) Isolation and biological activity FK506 (tacrolimus) was isolated from the fermentation broth of *Streptomyces tsukubaensis* No. 9993 in 1987[146,147] and the structure of the 23-membered ring macrolactone (**Figure 3**) was determined shortly thereafter through chemical degradation and NMR studies.[148] The

compound is a potent immunosuppressant, associating with the intracellular receptor FKBP to form a complex that subsequently binds calcineurin. Inhibition of this calcium-dependent phosphatase prevents the dephosphorylation of transcription factor nuclear factor of activated T cells (NFAT), which is required for its translocation into the nucleus. The normal NFAT function in regulating gene expression of cytokines, in particular the T-cell activator IL-2, is thereby prevented, leading to suppression of T-cell activation.[149] However, because calcineurins are also present in the kidney, liver, and other organs, the compound is nonspecific, which results in severe toxic side effects.[150] FK506 was initially approved by the US FDA for use after liver transplant, but has since been approved for use following kidney, heart, lung, and several other organ transplants. It is marketed as Prograf, Advagraf, and Protopic (Astellas Pharma).

1.11.2.2.2(ii) Gene cloning, sequence analysis The biosynthetic gene cluster for FK506 (**Figure 4**) is strikingly similar to that of rapamycin. The cluster contains three PKS genes (*fkb*B, *fkb*C, and *fkb*A) encoding ten elongation modules, one NRPS-encoding gene, *fkb*P, and one gene, *fkb*L, that encodes a lysine CDE that is responsible for the construction of pipecolate. The loading module contains a Co, ER, and PCP tridomain analogous to RapA, and chain termination/product release by FkbP is believed to be mediated by a similar mechanism as RapP.[151,152]

1.11.2.3 Antimicrobial

1.11.2.3.1 Pristinamycin

1.11.2.3.1(i) Isolation and biological activity The antibiotic pristinamycin (Pyostacine, Sanofi-Aventis), produced by *Streptomyces pristinaespiralis*, belongs to the family of streptogramin (also called virginiamycin-like or mikamycin-like) antibiotics.[153] Pristinamycin exists as a mixture of two types (termed A and B) of structurally distinct macrocyclic lactone peptides that exhibit synergistic antibacterial activity.[154,155] Pristinamycins I (type B) are NRPS-derived branched cyclic hexadepsipeptides, whereas pristinamycins II (type A) are hybrid PK/NRP macrocyclic lactones. The discussion will focus primarily on type A (**Figure 3**), exemplified by pristinamycin II_A (PII_A) and pristinamycin II_B (PII_B).

The pristinamycins are potent protein synthesis inhibitors, binding to the 50S subunit of the ribosome, but as noted above, have a unique synergistic mechanism.[155] If either a type A or type B compound is present alone, only bacteriostatic activity is observed.[154] However, when both types are present in combination, a synergistic bactericidal activity takes place that is both qualitative and quantitative (each component enhances the antibiotic power of its partner 100-fold). Mechanistically, this cooperativity is believed to arise from a conformational change of the 50S subunit that occurs after the binding of a type A compound. Presumably, this conformation change facilitates more effective binding of a type B compound.[154] Pristinamycin is marketed as an oral tablet in Europe as Pyostacine for the treatment of staphylococcal and streptococcal infections. It is particularly effective against erythromycin-resistant infections as well as methicillin-resistant *Staphylococcus aureus* (MRSA). Nevertheless, for severe infections its therapeutic use is restricted because of its poor solubility that limits intravenous formulation. This prompted the development of the quinupristin/dalfopristin pristinamycin-derivative (Synercid, King Pharmaceuticals), a water-soluble intravenous formulation, approved by the US FDA for serious infections in 1999.[156]

1.11.2.3.1(ii) Gene cloning, sequence analysis While isotope-labeled precursor-feeding studies of PII_A established the biosynthetic origin of the 28 pristinamycin carbon atoms (seven acetate units, amino acids valine, glycine, serine, and proline, and the methyl group of methionine) in the 1980s[157–160] and genetic and biochemical pristinamycin biosynthesis studies was first reported in 1995,[161–165] the complete pristinamycin biosynthetic gene cluster has yet to be cloned and characterized. Overexpression and purification of two enzymes responsible for the conversion of PII_B into PII_A have been reported. PII_A synthase, a heterodimer composed of the SnaA and SnaB proteins, catalyzes the oxidation of PII_B to PII_A, while the SnaC protein, an NADH:riboflavin 5′-phosphate OxR provides reduced flavin mononucleotide for the reaction.[161,165,166] Although a few of the NRPSs involved in type B pristinamycin biosynthesis have been cloned and characterized, the NRPS–PKS system for type A biosynthesis has not been reported.[162–164]

1.11.2.3.2 Antibiotic TA

1.11.2.3.2(i) Isolation and biological activity Antibiotic TA (myxovirescin A) was originally isolated from the culture broth of *Myxococcus xanthus* and its structure (**Figure 3**) was determined by chemical degradation and NMR spectroscopy analysis.[167–170] The compound is a wide-spectrum antibiotic active against Gram-negative bacteria through inhibition of cell wall synthesis. Antibiotic TA is exceptionally adhesive, binding to a variety of tissues and inorganic surfaces, while maintaining antibiotic activity in the bound form.[171] Due to this extraordinary property, the antibiotic is considered a good lead for the treatment of gingivitis, as well as an antibacterial coating for urethral catheters.[172,173]

1.11.2.3.2(ii) Gene cloning, sequence analysis The entire antibiotic TA gene cluster (**Figure 4**) from *M. xanthus* DK1622 has recently been annotated.[174] The cluster contains four ORFs that encode type I PKSs (TaI, TaL, TaO, TaP) and one ORF that encodes a PKS/NRPS hybrid (Ta-1). In addition, there are several genes that encode a variety of enzymes involved in oxidative and methyltransfer processes.[174] The antibiotic TA megasynthetase displays a number of unusual features, including a TE that is not preceded by an ACP and the presence of gene coding for a discrete *trans*-AT enzyme. A more thorough discussion of this interesting pathway will be presented (see below) with a focus on hybrid PKS/NRPS systems that include unusual biochemistry and domains.

1.11.3 *Trans*-AT Domain Pathways – a Rich Source of Unusual Biochemistry

1.11.3.1 *Trans*-AT Hybrid PKS/NRPS Systems – an Introduction

One emerging subclass of hybrid PKS/NRPS pathways are the '*trans*-AT' hybrid biosynthetic systems. Rather than containing embedded AT domains within their PKS modules, the *trans*-AT systems feature a separately encoded, discrete AT domain that is responsible for loading ACP domains with the appropriate CoA substrate (**Figure 5**). Interestingly, remnants of embedded ATs are found within *trans*-AT hybrid pathways, and have been proposed to act as 'AT-docking domains,' or recognition elements, thus providing evidence of an evolutionary link between the two types of pathways.[175]

Aspects of known NRPS/PKS *trans*-AT hybrid biosynthetic pathways including products, pathway organization, *in vitro* and *in vivo* biochemistry, biological roles, and potential for future engineering efforts will be explored within this section. For further information, the reader is directed toward one of the many reviews on

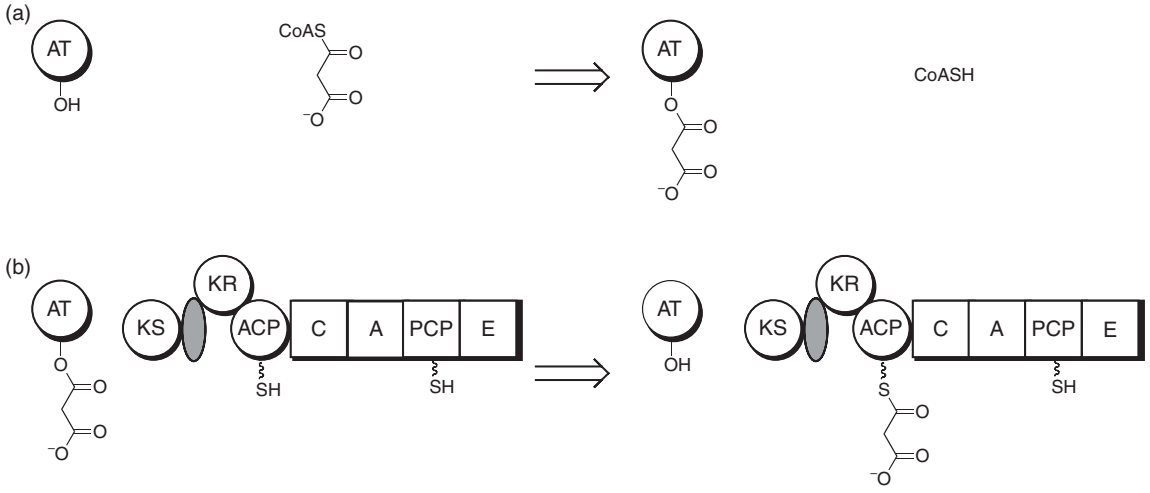


Figure 5 A schematic representation of a *trans*-AT reaction utilizing a hybrid PKR/NRPS biosynthetic module. (a) Loading of malonyl CoA onto the AT-active site serine. (b) Transfer from *trans*-AT to the phosphopantetheine arm of the PKS module ACP.

that have been recently published on closely related topics. These include hybrid NRPS/PKS pathways in general,[176,177] developing molecular tools to engineer these systems,[13] compounds derived from marine invertebrates and bacteria,[6] symbiotic bacteria-produced secondary metabolites,[178,179] and absence of colinearity including skipping and iteration.[180]

1.11.3.1.1 *Known* trans-*AT hybrid PK/NRP pathways*

Over the past several years, *trans*-AT hybrid biosynthetic pathways have been discovered in a wide range of bacterial species (summarized in **Table 1**). Moreover, in several cases (e.g., onnamide, pederin, and rhizoxin), the bacterial species is engaged in important symbiotic relationships with multicellular hosts. These symbiotic relationships often complicate the identification of the species of origin of the hybrid PK/NRP natural product. With few exceptions, such as those involved in the biosynthesis of mycosubtilin and albicidin, the majority of *trans*-AT hybrid pathways are predominantly composed of PKS elongation modules. In fact, it should be noted that *trans*-AT pathways that entirely consist of PKS modules have been characterized: CpPKS1,[181] mupirocin,[182] macrolactam,[183] difficidin,[184] and bryostatin.[185,186]

1.11.3.1.2 *Biological activity and structure of* trans-*AT hybrid PK/NRPs*

Known hybrid PK/NRP natural products that are derived from *trans*-AT biosynthetic pathways can be broadly classified as either antimicrobials or cytotoxic chemotherapeutics based on their biological activity (**Table 1**, Figure 6). Certain antimicrobial compounds, such as albicidin and mycosubtilin, have well-defined activities whereas others, such as bacillaene, chivosazol, and lankacidin, have yet to be rigorously characterized in terms of biochemical targets. Given their ability to cause damage to rapidly growing cells, leinamycin, onnamide, and disorazol, each has potential as an anticancer therapeutic. As discussed above, many hybrid PK/NRP natural products have demonstrated clinical utility; however, with the exception of antibiotic TA, none of the compounds are derived from *trans*-AT hybrid PKS/NRPS pathways.

1.11.3.1.3 In vivo *analysis of* trans-*AT hybrid PK/NRP systems*

To date, 13 *trans*-AT hybrid PK/NRP biosynthetic pathways have been described in the literature (**Figure 7**). Intriguingly, aside from the presence of the *trans*-AT domain, each of these pathways displays multiple deviations from the typical PKS modular organization and composition. In fact, only five of these pathways are colinear with genetic organization, as is typically observed in bacterial PKS or NRPS systems. Module splits, whereby the domains of a single module are divided among multiple polypeptides, are frequently observed within *trans*-AT hybrid pathways. HMG synthase (HMGS) cassettes are also commonly found in these biosynthetic pathways. These cassettes are responsible for the insertion of β-branch points into the middle of the growing PK chain.[187] Repeated 'tandem' domains, and unusual or uncharacterized enzymatic domains, are also often present in biosynthetic pathways that employ *trans*-AT domains.[180] Additional nonstandard features, such as iteratively acting modules or inactive modules, can often be inferred from the chemical structure of the natural product.

Putative *trans*-AT biosynthetic pathways can be linked to a specific natural product through comprehensive bioinformatic analysis, which is used to determine domain compositions and predicted acyl[188] or peptidyl substrate[189] incorporation of AT and A domains, respectively.[190] However, in nonlinear pathways or in pathways that skip or iterate elongation modules, these predictions are challenging and can be misleading. Pathway assignment is typically obtained through genetic disruption and complementation. For example, if the inactivation of a key biosynthetic gene results in a nonproducing phenotype, the pathway or product link is verified. However, this genetic approach is not a viable option for bacteria that are not culturable in the laboratory (often the case for symbionts) or for microorganisms that are not amenable to genetic manipulation. In such cases, the final proof may require complete pathway reconstitution in a heterologous host – a task not yet accomplished for any symbiont pathway.

1.11.3.1.4 In vitro *characterization of* trans-*AT hybrid PKS/NRPS pathways*

Given the many unusual features of the *trans*-AT hybrid PK/NRP biosynthetic pathways, the precise sequence of compound assembly and the exact role of specific domains cannot always be easily ascertained from either sequence analysis or *in vivo* biochemistry. A more direct approach toward understanding these issues is to

Table 1 Known hybrid PKS/NRPS *trans*-AT pathways

Name	*Activity*	*Bacterial class*	*Producing bacteria*	*Host*	*Designation*	*PKS*	*NRPS*
Albicidin	Cytotoxic	γ-Proteobacteria	*Xanthomonas albilineans*	NA	Alb/*alb*	3	7
Chivosazol	Cytotoxic	δ-Proteobacteria	*Sorangium cellulosum*	NA	Chi/*chi*	16	1
Disorazol	Cytotoxic	δ-Proteobacteria	*S. cellulosum*	NA	Dis/*dis*	10	1
Leinamycin	Cytotoxic	Actinobacteria	*Streptomyces atroolivaceus*	NA	Lnm/*lnm*	7	2
Onnamide	Cytotoxic	Unknown		Sponge: *Theonella swinhoei*	Onn/*onn*	?	?
Pederin	Cytotoxic	γ-Proteobacteria	*Pseudomonas aeruginosa*	*Beetle: Paederus* sp.	Ped/*ped*	9	2
Rhizoxin	Cytotoxic	β-Proteobacteria	*Burkholderia rhizoxina*	Fungi: *Rhizopus microsporus*	Rhi/*rhi*	11	1
Antibiotic TA	Antimicrobial	δ-Proteobacteria	*Myxococcus xanthus*	NA	Ta/*ta*	11	1
Bacillaene	Antimicrobial	Bacilli	*Bacillus subtilis*	NA	Pks/*pks*	13	2
Lankacidin	Antimicrobial	Actinobacteria	*Streptomyces rochei*	NA	Lkc/*lkc*	5	1
Mycosubtilin	Antimicrobial	Bacilli	*Bacillus subtilis*	NA	Myc/*myc*	1	7
Virginiamycin M	Antimicrobial	Actinobacteria	*Streptomyces virginiae*	NA	Vir/*vir*	8	2
Thailandamide A		Bacilli	*Bacillus thailandensis*	NA		16	1

The known hybrid *trans*-AT PK/NRP products are listed above by bioactivity, with bacterial class and bionomial. For symbiont products, the host organism is also designated. Protein and gene abbreviations are noted along with PKS and NRPS module composition.

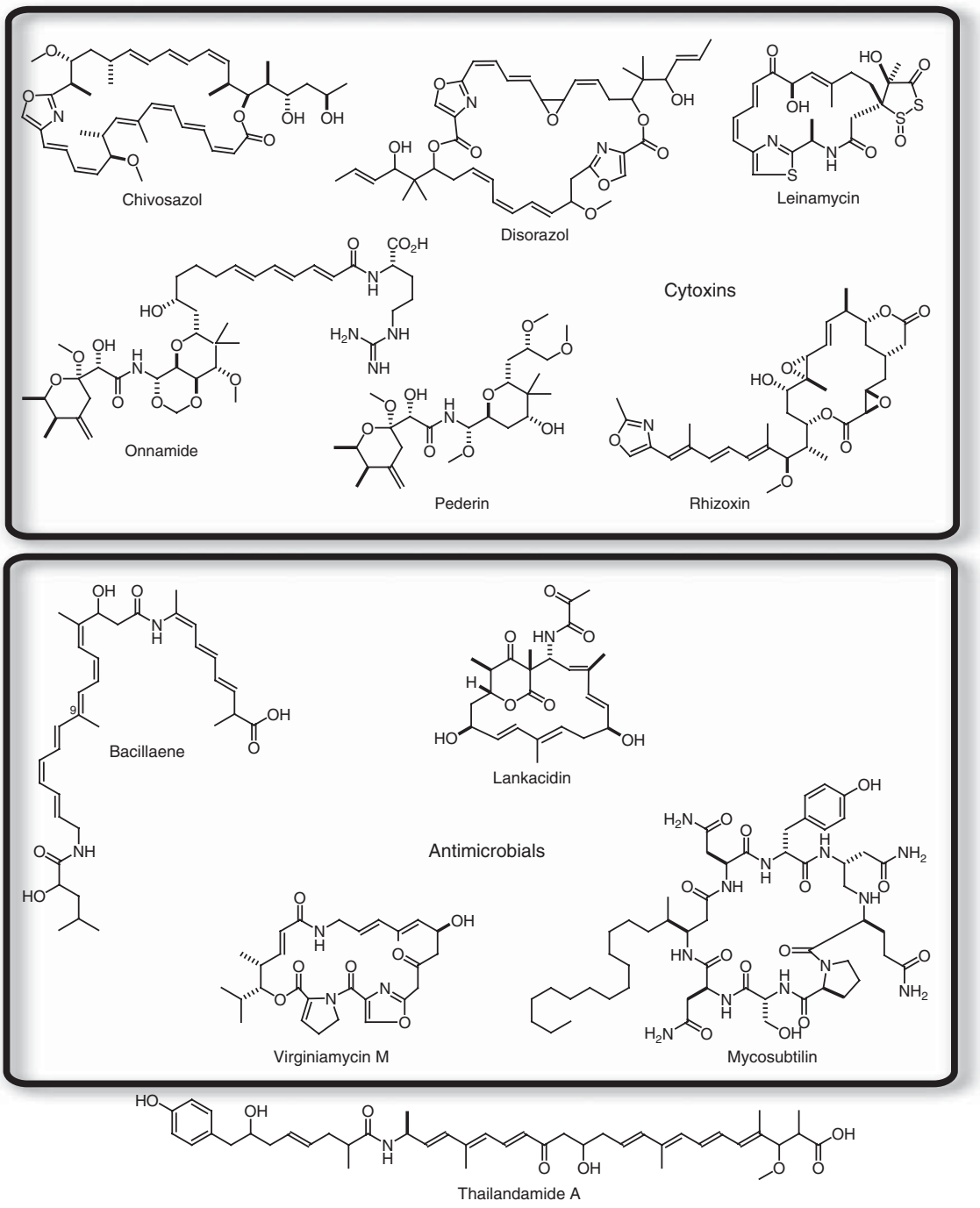


Figure 6 Structures of *trans*-AT hybrid PKS/NRPS products. Structures are displayed above for known hybrid *trans*-AT systems with characterized products. See **Figure 4** for the *trans*-AT hybrid PK/NRP antibiotic TA.

perform detailed *in vitro* biochemical investigations employing recombinant enzymes. Using defined assay conditions, detailed enzymology studies can provide important details of these hybrid megasynthetases. Additionally, heterologous expression and purification of recombinant proteins enable the possibility of gaining key structural data, and might eventually inform new avenues toward pathway reengineering *in vitro*. It is

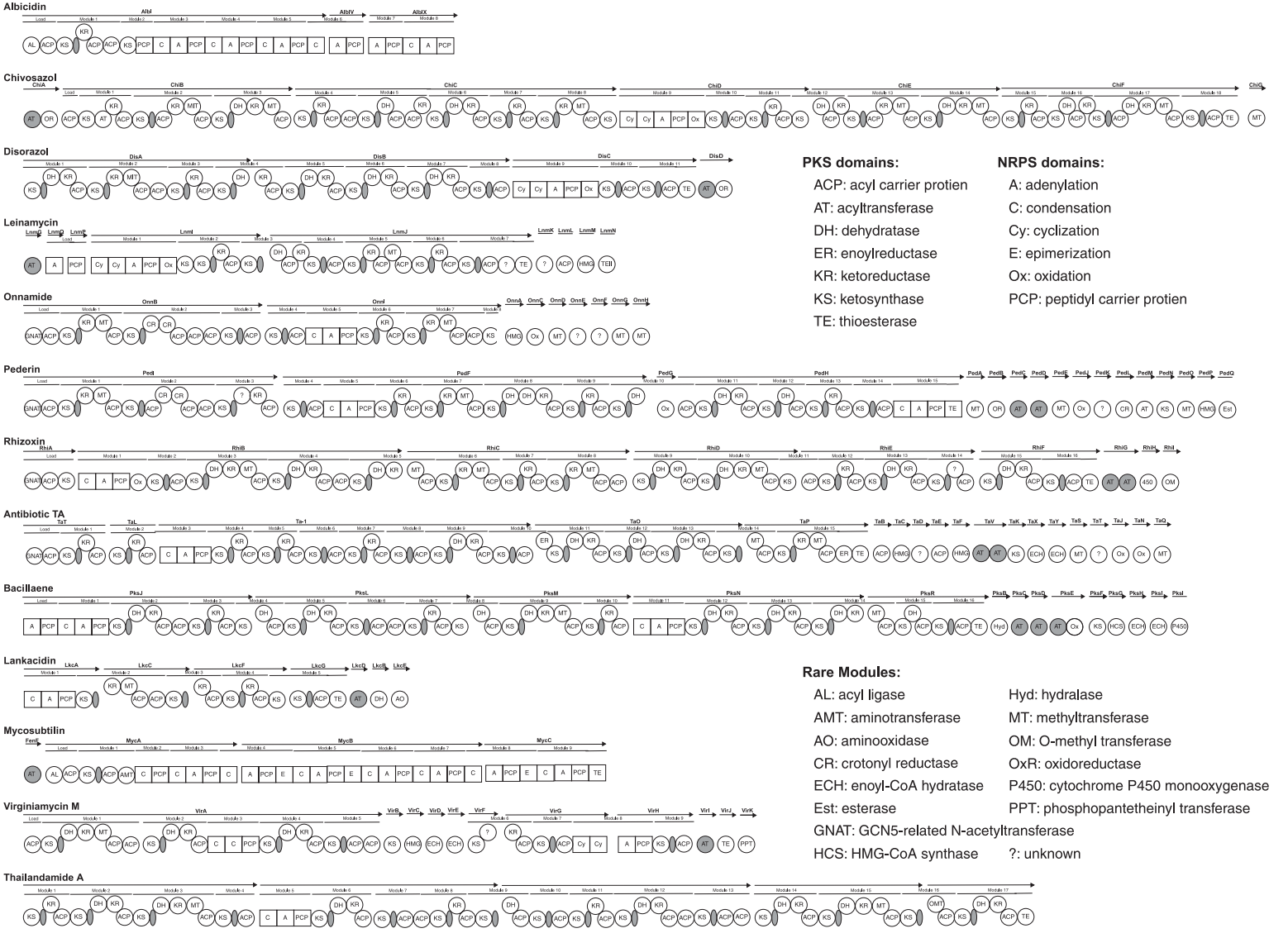


Figure 7 Biosynthetic pathways of *trans*-AT hybrid PKS/NRPS products. Pathways are arranged based on predicted biosynthetic schemes, with modules and domains labeled. *Trans*-AT domains are noted in gray, with likely post-ketosynthase (KS)-docking sites also in gray.

becoming increasingly apparent that rigorous *in vitro* examination of enzymes from a few selected pathways has dramatically improved our understanding of the role of unusual domains and architecture in these pathways.[187,191–197]

1.11.3.1.5 Evolution, biology, and symbiosis of trans-AT hybrid PKS/NRPS systems

Elucidation of the biological roles of hybrid PK/NRP natural products that are produced by *trans*-AT PKS/NRPS hybrid synthetases is a rapidly emerging field of research, particularly in relation to the developing models of microbial symbiosis in natural product biosynthesis. Why do organisms expend so much energy and genome composition to generate these elaborate natural products? While the chemical ecology of some of these compounds has been explored, considerable work remains to be done to understand the role of these compounds in serving for the producing organism.[198] Indeed, even the identity of the organism (host vs. symbiont) that is responsible for natural product biosynthesis is an area of intense interest. Macroscopic eukaryotes including insects, plants, marine sponges, and tunicates have long been recognized as sources of diverse natural products. Yet time and time again, evidence that strongly suggests associated microorganisms are responsible for natural product biosynthesis, especially when similar compounds are isolated from taxonomically diverse producers is developed. Cell separation experiments, as performed in the marine sponge *Theonella swinhoei,* have shown that the isolated fraction of bacteria colocalized with the secondary metabolite production.[199] The complexity of this problem becomes evident when it is recognized that up to 40% of the mass of a sponge may be composed of bacteria, fungi, and other microorganisms.[200]

1.11.3.2 Anticancer

1.11.3.2.1 Albicidin

1.11.3.2.1(i) Biological activity and structure of trans-AT hybrid PK/NRP Albicidin is an antibiotic phytotoxin, produced by the bacterium *Xanthomonas albilineans* that targets bacterial DNA gyrase. Binding of albicidin to DNA gyrase effectively stabilizes the enzyme–DNA complex, leading to the inhibition of replication and eventual cell death. The mechanism of albicidin is considered to be unique since it does not inhibit the ATPase activity of DNA gyrase, but does requires ATP to be bound for high potency.[201] In nature, this metabolite has been implicated as the primary agent responsible for leaf scald disease in certain plants. The producing organism, *X. albilineans*, employs this cytotoxic secondary metabolite to establish a niche to colonize the plant host and thus acquire essential nutrients.[202] Complete structural elucidation of the natural product produced by albicidin pathway has yet to be achieved, but has been proposed based on limited structural analysis to be a nonreduced, cyclic product.[203] Given current product yields (milligrams), structural identification is likely to occur in the near future. Once the structure is determined new insight may be gained into treating plants infected with the pathogen, or alternatively, it may serve as a lead for the development of human therapeutic agents based on its unique biological activity.

1.11.3.2.1(ii) In vivo biochemistry of trans-AT hybrid PK/NRP The albicidin biosynthetic pathway (**Figure 7**) consists of three polypeptides, AlbI, AlbIV, and AlbX, that are composed of a total of three PKS and seven NRPS elongation modules.[202,203] Natural product biosynthesis is initiated by the acyl ligase (AL)-loading domain, which is located at the N-terminus of AlbI. This polypeptide also contains a single PKS module, a tandem di-ACP domain, and three domains of an NRPS elongation module. It appears likely that the albicidin pathway is not colinear, and based on the ordering of the split modules, it is reasonable to hypothesize that biosynthesis proceeds from AlbI to AlbIV, which consists of a split NRPS monomodule that would correspond to the NRPS domains at the terminus of AlbI. Finally, AlbX contains both a split and a full NRPS elongation module that terminates with a TE domain, thereby suggesting it is the final polypeptide in the biosynthetic sequence. Gene disruption of the putative pathway resulted in the abrogation of product synthesis and biological activity, further implying that this pathway is responsible for the production of the natural product.[203,204]

1.11.3.2.1(iii) Evolution, biology, and symbiosis of trans-*AT hybrid PKS/NRPS systems* In certain cases (e.g., albicidin), the biological role of the secondary metabolite for the producing organism is clear. Albicidin has been implicated as the primary agent responsible for leaf scald disease in certain plants. *Xanthomonas albilineans* employs its cytotoxic secondary metabolite to establish a niche to colonize the plant host and thus acquire essential nutrients.[202]

1.11.3.2.2 Chivosazol

1.11.3.2.2(i) Biological activity and structure of trans-*AT hybrid PK/NRP* Chivosazol (**Figure 6**), produced by *S. cellulosum* So ce56, was isolated as a producer of active compounds from a screen of myxobacterium.[205] This compound is a cytotoxic antifungal agent that inhibits yeast cell morphology and mammalian cell membrane stability and integrity. The ability to interfere with membrane stability suggests that chivosazol has the potential as an anticancer lead compound.[206] Chivosazol contains a large macrocyclic polyene structure with a series of conjugated double bonds that incorporate both *cis* and *trans* isomers within the molecule.

1.11.3.2.2(ii) In vivo *biochemistry of* trans-*AT hybrid PK/NRP* The chivosazol biosynthetic pathway (**Figure 7**) was first identified by gene disruption studies that lead to a phenotype that was devoid of metabolite production.[207] Portions of the biosynthetic gene cluster near the mutation were identified, and a BAC containing the entire gene cluster was isolated. Sequence analysis to the gene cluster identified 16 PKS elongation modules and a single NRPS module.[207] The chivosazol pathway begins with the loading module located at the amino terminus of the ChiB polypeptide (**Figure 7**). Once initiated, biosynthesis proceeds through the three PKS modules on ChiB, after which the chain elongation intermediate is passed to ChiC. This polypeptide features five additional PKS elongation modules. Interestingly, ChiC terminates with an aberrant KS domain. The next polypeptide in the biosynthetic sequence, ChiD, harbors a single NRPS elongation module that contains tandem cyclization domains as well as tandem PCP domains. Also present on ChiD are two full PKS elongation modules, as well as a split PKS module. This split module appears to span to the amino terminus of ChiE, which begins with an unusual DH-ACP-KR domain arrangement. Furthermore, ChiE features two additional PKS modules. Finally, the biosynthetic machinery is completed with the four additional PKS elongation modules of ChiF. In addition to the described elongation modules, additional proteins have been linked to this pathway. These include ChiA, a *trans*-AT/OxR, and a discrete MT, ChiG. Interestingly, based on the chemical structure of the final product, it appears that a select number of KS and MT domains within this pathway are inactive.[207]

Recently, media optimization experiments were conducted in attempts to increase chivosazol production.[208] During these studies, it was observed that the transcription of *chi* biosynthetic genes occurs at high levels during logarithmic growth, with all strains exhibiting a reduction in transcription during stationary phase in all media conditions examined. In many other cases, natural product biosynthesis is thought to occur largely in stationary phase.[209] The 13-Mb genome sequence for the chivosazol producer, *S. cellulosum* So ce56, has been reported.[210]

1.11.3.2.3 Disorazol

1.11.3.2.3(i) Biological activity and structure of trans-*AT hybrid PK/NRP* Disorazol (**Figure 6**) exhibits promising anticancer activity, both as a growth inhibitor at low concentrations and as a cell cycle inhibitor and apoptosis inducer at high concentrations. It was serendipitously isolated during the fermentation of a myxobacterial strain for the natural product sorangicin.[211] Based on *in vitro* and *in vivo* studies, disorazol appears to inhibit microtubule polymerization.[212] The chemical structure of disorazol is notable for the macrocycle formed by dimerization of the monomeric subunit. Other interesting structural features of this natural product include an epoxide, an extended conjugated double bond system, an oxazole ring, and gem-dimethylation.

1.11.3.2.3(ii) PKS/NRPS pathways and in vivo *analysis of* trans-*AT hybrid PK/NRP* The disorazol synthetase (**Figure 7**) is a myxobacterial *trans*-AT hybrid PKS/NRPS pathway. Unusual features of this biosynthetic pathway include split modules and tandem NRPS cyclization domains.[213] Biosynthesis of disorazol begins with the DisA polypeptide that consists of three PKS elongation modules, the second of which contains a tandem ACP. A split PKS module is located at the carboxy terminus of DisA, the remainder of which is found at the amino terminus of DisB. Four additional PKS elongation modules complete DisB. The terminal protein in this

pathway, DisC, begins with an NRPS module that bears tandem cyclization domains followed by two PKS elongation modules. The *trans*-AT that operates along the biosynthetic pathway, DisD, is a didomain. As with several other hybrid PKS/NRPS pathways, this didomain contains a putative OxR of unknown function.[213]

1.11.3.2.4 Leinamycin

1.11.3.2.4(i) Biological activity and structure of* trans-*AT hybrid PK/NRP Leinamycin (**Figure 6**) was discovered during screening of soil samples for producers of anticancer activities and is produced by *Streptomyces*.[214] Leinamycin has potential utility as an anticancer drug lead given its defined activity as a thiol-triggered DNA-alkylating agent (although thiol-independent mechanisms are also possible). The bioactivity of leinamycin appears to arise from a dual-action mechanism, whereby a rearrangement of the 1,2-dithiolan-3-one 1-oxide heterocycle releases hydrosulfide and polysulfide species that react to generate reactive oxygen species and subsequent oxidative DNA damage. Furthermore, the remaining portion of leinamycin produces an episulfonium intermediate that has been shown to alkylate guanine nucleobases.[215] Other chemical moieties of interest within leinamycin include a *cis* double bond and a thiazole ring.

1.11.3.2.4(ii) PKS/NRPS pathways and* in vivo *analysis of* trans-*AT hybrid PK/NRP The leinamycin biosynthetic pathway (**Figure 7**) was initially identified using a DNA probe developed against a thiazole-forming NRPS synthetase.[216] The complete gene cluster was then isolated from a genomic DNA library and linked to the natural product through gene disruption and confirmation of a nonproducing phenotype.[175] As with many of the *trans*-AT hybrid pathways, the proposed model for leinamycin biosynthesis appears to be noncolinear and includes both split elongation modules and unusual domain arrangements. Leinamycin biosynthesis is uniquely initiated by the activity of a discrete A domain (LnmQ), which loads a discrete PCP domain (LnmP).[175] Biosynthesis then continues with LnmI, which features an NRPS module-containing tandem cyclization domains and a presumed oxidase. Intriguingly, LnmI is also composed of a PKS elongation module whose amino terminus contains redundant KS domain. This NRPS and PKS elongation modules are followed by the KS of a split PKS module, the remainder of which is found at the amino terminus of LnmJ. It is interesting to note that the first PKS module of LnmJ contains an unusual ACP domain that is embedded between the DH and KR. Four more PKS modules then follow in LnmJ with another unusual ACP insertion, this time between a KR and MT domain. The last PKS module of LnmJ contains a domain of unknown function that precedes the terminal TE. The didomain *trans*-AT that operates along this pathway appears to also include a third domain of unknown function.

The boundaries of the leinamycin gene cluster were established by a series of gene disruption experiments and the essential function of the *trans*-AT domain was similarly confirmed.[175] A series of discrete proteins are also present within the biosynthetic pathway. These include LnmK (unknown function), LnmL (ACP), LnmM (an HMG-CoA synthase (HCS)), and LnmN (type II thioesterase).[191,216]

1.11.3.2.4(iii)* In vitro *biochemical characterization of* trans-*AT hybrid pathways Unlike most *trans*-AT hybrid biosynthetic pathways, the leinamycin biosynthetic system has been extensively investigated *in vitro*. Pathway ACP domains, the *trans*-AT domain and a full module, have been cloned, overexpressed, and purified. ACP-loading assays have been developed using both radio-sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and high-performance liquid chromatography (HPLC)–mass spectrometry (MS). These assays have demonstrated that malonyl-CoA can be loaded onto all expected ACP partners by the *trans*-AT domain.[191] A later study investigated the NRPS portion of the pathway, specifically the single proteins LnmQ (an A domain) and LnmP (a PCP domain), using a combination of radio-SDS-PAGE, HPLC, and electrospray ionization (ESI)–MS.[217] In these experiments, it was observed that LmnQ directly activated D-alanine as the adenylate and subsequently loaded the activated amino acid onto LnmP.[217] This is a surprising observation, as typically L-amino acids are loaded and then epimerized to give the D-amino acid product. The interaction between LnmQ and LnmP appears specific, and channeled, thus underscoring the importance of protein–protein interactions between discrete domains. Steady-state kinetic analysis using a variety of amino acids revealed that D-Ala was the preferred substrate, although a small amount of glycine was observed to react as well; however, the lower substrate affinity of glycine (K_M) and reaction rate were much lower.[217]

An additional feature of the leinamycin biosynthetic pathway is the presence of the tandem ACP domains in LnmJ. Both *in vitro* and *in vivo* experiments have demonstrated that these ACPs are redundant.[192]

1.11.3.2.5 Onnamide/pederin

1.11.3.2.5(i) Biological activity and structure of trans-AT hybrid PK/NRP Both onnamide and theopederin, a close analogue of pederin, (**Figure 6**) are nanomolar inhibitors of protein synthesis, leading to the induction of ribotoxic stress response, p38 kinase activity, and apoptosis. These activities were discovered during a screen for activators of transforming growth factor-β (TGF-β) expression.[218] It has been hypothesized that these compounds may directly bind to the eukaryotic ribosome, thus resulting in downstream activation of apoptotic pathways.[218]

Onnamide and pederin each contain a similar carbon backbone with two tetrahydropyran rings linked through an amide bond. Furthermore, each features an exocyclic double bond. Differences between the two natural product compounds include the presence of an additional hydroxyl group in onnamide that results in the formation of a third six-membered ring. Onnamide also contains a longer, conjugated side chain that terminates in an arginine residue. The structural similarity shared between onnamide and pederin was proposed as evidence for the existence of a related symbiont producer in natural product biosynthesis[219,220] well before Piel obtained the gene clusters from the producing organisms.[221,222]

1.11.3.2.5(ii) In vivo biochemistry of trans-AT hybrid PK/NRP The biosynthesis of pederin and onnamide will be discussed together due to the close chemical and biosynthetic similarities that are shared between the compounds and pathways.[221–223] The role of these pathways (**Figure 7**) in settling the long-standing debate over the source of natural products in marine invertebrates (e.g., sponges, tunicates) is discussed below. Both pederin and onnamide are initiated from a PKS elongation module (OnnB/PedI) that begins with a GNAT-loading domain. A similar initiation mechanism is found in the curacin biosynthetic pathway.[100] The domain composition and arrangement are identical for each pathway over the first 12 domains, encompassing two PKS elongation modules. Interestingly, two unusual domains exist within these first 12 domains, and have been annotated as crotonyl-CoA reductases. Divergence in biosynthesis occurs beyond this point. The onnamide biosynthetic machinery proceeds with a tandem triple ACP, whereas pederin biosynthesis continues with a tandem di-ACP. In addition, the final PKS module of OnnB contains only KS and ACP domains, whereas the final module in PedB features a KS domain, a domain of unknown function, a KR domain, and finally an ACP domain. The subsequent polypeptides, OnnI or PedF, share a high degree of similarity, with the arrangement of the first 12 domains being identical. Briefly, each polypeptide begins with a PKS elongation module, followed by an NRPS module, and a second PKS module. At this point in the biosynthetic pathways, divergence is observed. Here, OnnI terminates with tandem ACP domains, followed by a KS domain. It should be noted that the DNA sequence of onnamide biosynthetic pathway is presumably incomplete, and therefore OnnI does not likely represent the terminus of the biosynthetic pathway. This is because out of a 500 000 member clone library, only one incomplete cosmid containing the onnamide gene cluster was isolated.[222] In comparison, PedF continues with a single ACP domain that is followed by two additional PKS modules, the first of which contains a tandem DH didomain. At this point, two biosynthetic possibilities have been proposed. First, PedG catalyzed hydrolysis of the chain elongation intermediate from PedI and yields the pederin nascent intermediate (pretailored). Alternatively, chain extension continues through PedH, which results in a product having a very similar structure to onnamide. This 'onnamide'-type intermediate would then presumably be cleaved to give the beetle-derived product, pederin. The domain arrangement of PedH involves a presumed oxygenase, PedG, of the previous split module, PedF. PedH continues with four PKS elongation modules before the final arginine-incorporating NRPS module that terminates with TE. Additional discrete proteins are also present in the pederin pathway. These include PedA/E/Q (MTs), PedB (OxR), PedC/D/M (*trans*-AT tridomain), PedJ (oxidase), PedK (unknown function), PedL (crotonyl-CoA reductase), PedN (KS), PedP (HCS), and PedO (esterase (Est)). Several proteins with high sequence similarity are found in the onnamide pathway. These include OnnA (HCS), OnnC (oxidase), OnnD/G/H (MTs), and OnnE/F (unknown function).[221–223] As a final word of caution, it is important to note that these pathway assignments be considered putative, as they have not been confirmed as the metabolite source through heterologous expression or through detailed biochemical analysis of corresponding purified proteins.

1.11.3.2.5(iii) Evolution, biology, and symbiosis of* trans-*AT hybrid PKS/NRPS systems One of the mysteries that have intrigued natural product chemists for years is how structurally similar natural products can be isolated from evolutionarily distinct hosts (e.g., marine sponges, myxobacteria). These discoveries have led to the hypothesis that microorganisms are the likely producers of marine invertebrate-derived natural products. Support for this hypothesis was offered by Piel, who isolated, characterized, and comparatively analyzed the DNA encoding the biosynthetic pathways of pederin and onnamide from the rove beetle and a marine sponge, respectively. The isolation and subsequent screening of DNA from the gut bacteria of the pederin source (rove beetle) eventually lead to the identification of the pederin pathway that originated from an unculturable symbiont.[221] Subsequently, Piel hypothesized that a similar symbiotic relationship accounted for the existence of a homologous pathway for onnamide production in the marine sponge *T. swinhoei*. Screening of a *T. swinhoei* metagenomic library led to the identification of a biosynthetic pathway that shared high similarity with that of pederin.[224] The relationship of pederin and onnamide represents a fascinating example in which similar natural product biosynthetic gene clusters are derived from unique, phylogenetically distinct strains from widely disparate macroorganisms. Extensive gene sequencing has identified that *Pseudomonas aeruginosa* is the beetle endosymbiont responsible for pederin production.[225] To date, the microbial symbiont of the sponge has yet to be determined; the DNA appear to be of bacterial origin.[223] Further questions regarding evolution and divergence of the two pathways and host diversity remain to be fully explored.

Recent work to investigate the evolution of *trans*-AT systems, the main class of synthetase isolated to date from symbiotic organisms, may help frame the pertinent biological questions. Piel recently grouped known *trans*-AT synthetases and subjected them to multiple amino acid sequence alignments.[190] Interestingly, only the KS and MT domains showed conservation in all examined sequences. *Trans*-AT KS-specific clades are formed in an intriguing manner. Domains did not cluster based on whether or not they were from the same gene cluster, as seen for *cis*-ATs, but rather based on what is final extension unit generated. From this analysis, several insights are gained. First, it is possible to predict *trans*-AT PKS product structure to a reasonable degree of accuracy, even in cases where pathways have reductive or β-branching domains acting *in trans*. Such a prediction was illustrated for thailandamide.[190] Second, and perhaps most importantly, Piel and coworkers[190] revealed that *trans*-AT systems have likely evolved through a very different mechanism compared to their *cis*-AT PKS counterparts. In *trans*-AT pathways, horizontal gene transfer and recombination appear to be the driving force, as opposed to recombination. This suggests that *trans*-AT PKS pathways may be classified separately from their *cis*-AT counterparts, much in the sense that PKS and NRPS pathways are to each other.

The rapidly growing number of complete genome sequences from free-living and symbiotic bacteria, as well as from environmental samples, is expected to lead to an increase in the number of characterized *trans*-AT hybrid PKS/NRPS pathways.[190] Previous comparative studies made with significantly smaller amounts of sequence data began to show evidence of distinct clades forming between methylmalonyl-CoA *cis*-ATs, malonyl-CoA *cis*-ATs, and *trans*-AT modular PKS domains.[191] Interestingly, it appears that *trans*-AT PKS or PKS/NRPS systems may be underrepresented in current databases, as relatively larger numbers of *trans*-AT PKS systems have been observed (relative to *cis*-AT PKS systems) in random sequencing of bacterial strains.[226]

1.11.3.2.6 Rhizoxin

1.11.3.2.6(i) Biological activity and structure of* trans-*AT hybrid PK/NRP Rhizoxin (**Figure 6**) is a particularly interesting hybrid PK/NRP natural product that has recently been identified as the primary cause of rice seedling blight, likely due to its ability to interfere with eukaryotic cell growth by modulating tubulin function.[227] Interestingly, rhizoxin was long thought to be of fungal origin, but recent work has demonstrated that it is the product of a bacterial endosymbiont of the pathogenic fungus *Rhizopus microsporus*.[228–231] Because of its mechanism of action, rhizoxin has been investigated as a potential anticancer therapeutic and has been examined in phase 2 clinical trials for melanoma, breast, head, neck, and nonsmall cell lung cancers.[232,233] Throughout the course of these trials, dose-limiting toxicity was observed, a common side effect of compounds that disrupt microtubule formation.[232,233] Although rhizoxin was relatively well tolerated for its class, little efficacy was observed during clinical trials, a result attributed to low bioavailability and rapid clearance.[232,233] Continuous IV administration of the drug for up to 72 h in patients with solid malignancies also failed to elicit a clinical response.[234] Mechanistically, rhizoxin binds β-tubulin, thus disrupting mitotic cell division. Structurally, rhizoxin consists of a thiazole ring that is linked through three *trans* double bonds to a

16-membered macrolactone ring. Additional chemically interesting structural features of rhizoxin include two epoxide moieties and a six-membered ring that is fused to the macrocyclic scaffold.

1.11.3.2.6(ii) PKS/NRPS pathways and in vivo *analysis of* trans-*AT hybrid PK/NRP* The rhizoxin biosynthetic pathway (**Figure 7**) represents another example of a microbial symbiont-derived system. This pathway is colinear and consists of 11 PKS elongation modules and a single NRPS module.[235] Rhizoxin biosynthesis is initiated by the $GNAT_L$-ACP_L-KS tridomain contained on the RhiA polypeptide. Following initiation, chain elongation is continued by RhiB, which features an NRPS module and three full PKS elongation modules, and finally a split PKS module that terminates with a KR domain. The second to last module of RhiB contains a di-ACP. This split module spans to the next polypeptide, RhiC, which begins with an MT domain. RhiC continues with three PKS elongation modules, the final of which terminates with tandem ACP domains. Rhizoxin biosynthesis continues with RhiD, a polypeptide containing two full PKS elongation modules and a single partial module consisting solely of a KS domain. The remainder of this module is found on RhiE, which also includes three additional PKS modules, one of which contains a domain of unknown function. Finally, rhizoxin biosynthesis is completed by the two PKS elongation modules contained on RhiF. Additional proteins involved in rhizoxin biosynthesis include RhiG (AT_2 didomain), RhiH (P450 monooxygenase), and RhiI (*O*-methyltransferase (OM)). *In vivo* deletion of RhiG AT activity results in a rhizoxin nonproducing phenotype.[235]

1.11.3.3 Antimicrobial

1.11.3.3.1 Antibiotic TA

1.11.3.3.1(i) Biological activity and structure of trans-*AT hybrid PK/NRP* The bioactivity of antibiotic TA, also known as myxovirescin, (**Figure 6**) as well as application to human health have been fully described in the previous section. Antibiotic TA contains a number of interesting structural features with close ties to interesting biochemistry including ethyl- and methylmethoxy group inserted through β-branching.

1.11.3.3.1(ii) PKS/NRPS pathways and in vivo *analysis of* trans-*AT hybrid PK/NRP* The biosynthetic pathway responsible for the production of antibiotic TA (**Figure 7**) consists of 15 PKS elongation modules and one NRPS module, that are dispersed over five polypeptides. In addition, there are also two HMGS-type cassettes that incorporate ethyl- and methyl β-branch points.[236] Initiation of antibiotic TA biosynthesis commences at TaT, a polypeptide containing a curacin[100]-like GNAT-loading domain and a single PKS elongation module. Next in the biosynthetic sequences is the monomodular PKS polypeptide TaL, which is followed by the Ta-1-containing NRPS module and seven PKS elongation modules. Generation of antibiotic TA continues with TaO. Intriguingly, TaO begins with an ER domain at its amino terminus, followed by three PKS elongation modules before terminating with a partial PKS module consisting of a sole KS domain. The remainder of this partial module, an MT and ACP domain, is found at the amino terminus of TaP. This split module is followed by one final PKS elongation module that, in addition to including an MT domain also contains an unusual ER domain following the ACP domain. TaV is a didomain *trans*-AT_2 that services the pathway by loading CoA extension units onto the ACP domains of the PKS elongation modules. Aside from the core PKS/NRPS polypeptides, additional proteins are found within the pathway. These include TaB/E (ACPs), TaC/F (HCSs), TaD/T (unknown function), TaK (discrete KS), TaX/Y (enoyl-CoA hydratases (ECHs)), TaS/Q (MTs), and TaJ/N (oxidases).[236]

In vivo genetic disruption techniques have been utilized to verify the authenticity of the antibiotic TA biosynthetic pathway.[236] Furthermore, the role of the HMGS cassettes in β-branching was confirmed by gene deletion experiments and subsequent product analysis.[236] Interestingly, reciprocal genetic complementation studies have confirmed that the two PPTs that are present within the antibiotic TA biosynthetic gene cluster have the ability to substitute (i.e., rescue) for each.[237] Organelle separation experiments, using density-based centrifugation, demonstrated that the antibiotic TA megasynthase localizes to the inner membrane; however, these studies were not able to ascertain if the TA biosynthetic enzyme system is diffuse or forms higher-order, localized complexes.[238]

1.11.3.3.1(iii) In vitro characterization of trans-AT hybrid PKS/NRPS pathways *In vitro* radio thin-layer chromatography (TLC)/SDS-PAGE and MS assay experiments have been performed to further examine the mechanism of incorporation of the methoxymethyl (C12) or ethyl (C16) substituents at β-branching points.[239] For these experiments, the AT didomain, ACPs, HCS, and ECH domains were cloned, and expressed, and purified as discrete proteins in order to enable dissection of their individual roles in substrate transfer. Based on these experiments, the order of assembly of the β-branch was ascertained. First, the so-called 'donor' ACP (TaB or TaE) is loaded with a malonyl moiety by the AT (TaV) to give a malonyl-ACP. Next, the KS domain (TaK) catalyzes the decarboxylation of the malonyl-ACP to yield an acetyl-ACP. The acetyl-ACP group reacts with the keto group of the upstream covalent PK/NRP intermediate to form the HMG β-branch point (TaC or TaF). Subsequently, this intermediate is converted through dehydration and decarboxylation steps (TaX and TaY) to provide a methyl or ethyl group at the appropriate C12 or C16 carbon on the acceptor ACP (Ta1–T6, Ta1–T8).[239]

1.11.3.3.2 Bacillaene

1.11.3.3.2(i) Biological activity and structure of trans-AT hybrid PK/NRP Bacillaene (**Figure 6**), a hybrid PK/NRP natural product from *Bacillus subtilis*, inhibits prokaryotic growth by disrupting protein synthesis through an unknown mechanism.[240] Structurally, bacillaene is a highly conjugated linear polyene that includes a single *cis* double bond. The history of this PK/NRP natural product is particularly interesting because its biosynthetic pathway, PksX, was identified in the model Gram-positive organism *B. subtilis* prior to structural elucidation of the natural product.[241] Structural characterization of bacillaene was particularly challenging due to its sensitivity to light, acid, base, and chromatography. Ultimately, the structure of bacillaene was elucidated from a series of NMR experiments that employed semipurified compound extracts.[242] Given that the rate of pathway identification is increasing due to advances in both genomics and proteomics, this experimental approach may likely be extremely valuable in future natural product discovery efforts. Applications of sophisticated analytical methods and informatics-based structural prediction may inform more direct approaches to metabolite purification methods to increase the rate of natural product structural characterization.[242]

1.11.3.3.2(ii) PKS/NRPS pathways and in vivo analysis of trans-AT hybrid PK/NRP The bacillaene biosynthetic gene cluster (**Figure 7**) from both *B. subtilis* and the related organism *Bacillus amyloliquefaciens* incorporates isoprenoid-like and PKS/NRPS biosynthetic processes that span five polypeptide chains.[187,243,244] As with other pathways that have previously been described, bacillaene is not generated via a colinear biosynthetic architecture. PksJ initiates bacillaene biosynthesis with an AL-loading module, followed by an NRPS module and two PKS extension modules, the first of which contains a tandem ACP. PksJ terminates in a partial module that solely consists of a KS domain. The remainder of this module is found at the amino terminus of PksL. PksL also contains three PKS elongation modules, the second of which ends with a tandem ACP. Like PksJ, PksL terminates in a split PKS module consisting solely of a KS domain. The remainder of the split module is located at the amino terminus of PksM, which also contains two PKS elongation modules. Next, PksN begins with an NRPS module, followed by two PKS elongation modules. The carboxy terminus of PksN also includes a split PKS module. An MT and ACP domain are located at the amino terminus of the PksR module. These domains are followed by two PKS extension modules to complete the synthesis of the natural product.

A didomain AT-OxR (Ox), PksE, serves as one of the *trans*-AT domains in this pathway, although the Ox domain is putative and remains uncharacterized. Two monodomain *trans*-ATs are also present, PksC and PksD. In addition, several discrete enzymes are present in the pathway. These include PksB (hydrolase), PksF (KS), PksG (HMG-CoA synthase), PksH/I (ECH), and PksI (P-450 monooxygenase).[187,243,244]

Several *in vivo* experiments have been performed to better understand bacillaene biosynthesis. For example, a series of spontaneously off-loaded acyl-acid chain elongation intermediates have been detected by LC–MS, thereby confirming intermediate structures in bacillaene biosynthesis.[245] Most impressively, cellular localization studies utilizing sophisticated electron microscopy techniques have demonstrated that the bacillaene megasynthetase forms membrane-associated organelle-like structures likely due to the formation of a huge complex containing many copies of the biosynthetic proteins.[243]

1.11.3.3.2(iii) **In vitro *characterization of* trans-*AT hybrid PKS/NRPS pathways*** Efforts toward gaining a detailed understanding of bacillaene biosynthesis have broadly focused on the post-PKS/NRPS-tailoring enzymes as well as the mechanisms involved in the β-branching. To this end, the Walsh laboratory has cloned and overexpressed the tailoring enzyme PksS from the bacillaene biosynthetic pathway. *In vitro* studies have confirmed that PksS is a cytochrome P-450 that is likely involved in the conversion of dihydro-bacillaene to bacillaene.[194] It has been hypothesized that this conversion involves initial P-450-catalyzed hydroxylation of the product, followed by elimination to produce bacillaene.[194] Because of the instability of the mature natural product, it is difficult to rule out possible alternative products that might be generated *in vivo*. A reductase activity has been reconstituted from a cell-free extract that enables catalytic cycling;[194] however, it remains unclear whether or not the unknown partner is a class I ferridoxin reductase or a class II flavoreductase.

The biochemical details of β-branching in bacillaene have been explored through heterologous expression of a series of proteins that include the *trans*-AT (PksC), the PCP (AcpK), the HCS (PksG), the two ECH domains (PksH/I), as well as the 12 domain PKS protein PksL.[187] Using the combination of MS and radio-SDS-PAGE, the generation of a PksL-bound thioester and an α,β-unsaturated branched (C9′) intermediate, was observed. As in the case with antibiotic TA, the 'donor' ACP is first loaded with malonyl-CoA, followed by KS-catalyzed decarboxylation to give an acetyl-ACP species. At this point, C2 of the acetyl group reacts with the β-ketone of the acetoacetyl-ACP intermediate, followed by dehydration and decarboxylation, to yield the C9′ β-branched product.[187]

1.11.3.3.3 Lankacidin

1.11.3.3.3(i) Biological activity and structure of **trans-*AT hybrid PK/NRP*** Lankacidin (**Figure 6**) is a *trans*-AT-derived hybrid PK/NRP natural product that has demonstrated antibiotic activity in diffusion assays with *S. aureus*. Protective effects have also been observed *in vivo* with infected mice because lethal doses of the natural product were determined to be substantially higher than those required for antibiotic activity.[246] Structurally, lankacidin contains a six-membered lactone ring within a larger 17-membered macrocyclic scaffold. An unstable pyruvoyl group is installed via an amide linkage to the macrocycle.

1.11.3.3.3(ii) PKS/NRPS pathways and **in vivo *analysis of* trans-*AT hybrid PK/NRP*** The plasmid-borne lankacidin biosynthetic pathway, consisting of four PKS and one NRPS module, lacks colinear architecture.[247] This pathway (**Figure 7**) is entirely contained on a 200-kb linear, transmissible plasmid, consistent with the hypothesis that many secondary metabolite gene clusters are horizontally transferred between microorganisms. It is noteworthy that the biosynthetic pathway for the PKS macrolide lankamycin is also contained on the same plasmid. Lankacidin biosynthesis is initiated with an NRPS module located at the amino terminus of LkcA. The LkcA polypeptide terminates in a KS domain and is part of the PKS elongation module that continues with an amino terminal KR domain of LkcC, and terminates with tandem ACP domains. This module is followed next by the KS domain of yet another split module that spans to LkcF. In addition to the amino terminal partial module, LkcF bears one full PKS elongation module followed by a KS domain of unknown function. Finally, the PKS elongation module contained on LkcG completes lankacidin biosynthesis. LkcD is the *trans*-AT domain that operates along the lankacidin biosynthetic pathway. Other discrete enzymes include a DH (LkcB), and a presumed amino oxidase, LkcF.

Analysis of the pathway organization predicts the presence of five KS domains; however, generation of the mature lankacidin natural product should require eight elongation steps, and thus eight KS domains. This discrepancy suggests that some of the elongation modules are capable of iterative activity (i.e., acting more than once). Although the pathway assignment has been confirmed based on genetic disruption studies, it remains unclear which KS domains might control iterative steps.[247] In other *in vivo* studies, the role of LkcE was examined. From a combination of gene inactivation and precursor-feeding studies, LkcE was implicated in imine formation leading to ring formation from the linear intermediate.[248] Finally, the regulation of both the lankacidin and lankamycin pathways by a butyrolactone receptor system and details of the regulatory cascade have been characterized extensively, including the series of regulatory genes *srr*A–*srr*C, and *srr*X.[249,250]

1.11.3.3.4 Mycosubtilin

1.11.3.3.4(i) Biological activity and structure of trans-*AT hybrid* Mycosubtilin (**Figure 6**) is a hybrid *trans*-AT-derived fatty acid synthase (FAS)/NRP natural product. This natural product possesses antifungal activity stemming from its ability to form pores in cellular membranes.[251] Structurally, mycosubtilin is unusual in that it contains a linear fully saturated FAS-derived unit that is appended onto an NRPS-derived cyclic peptide backbone.

1.11.3.3.4(ii) PKS/NRPS pathways and in vivo *analysis of* trans-*AT hybrid PK/NRP* The mycosubtilin biosynthetic pathway (**Figure 7**) consists of one PKS extension module and seven NRPS modules that span three polypeptide chains.[252] *In vivo* mutagenesis was used to correlate natural product generation to the biosynthetic pathway.[252] Biosynthesis of mycosubtilin is initiated with a FAS-derived acyl chain serving as the starter unit. The acyl chain is subsequently elongated by the KS domain in MycA, and a unique AMT. Biosynthesis continues with the two NRPS modules located on MycA. Interestingly, the first NRPS module appears to be missing an A domain. Finally, MycA terminates with a C domain of a split module, the remainder of which is found at the amino terminus of MycB. Three additional NRPS extension modules are located in MycB to mediate chain elongation. Following these NRPS modules is a split NRPS module that is shared with MycC. The MycC polypeptide also contains a final NRPS module that completes the generation of the mycosubtilin natural product. FenF is the discrete *trans*-AT that selects and transfers malonyl-CoA to the ACP domains found on the biosynthetic pathway.[252]

1.11.3.3.4(iii) In vitro *characterization of* trans-*AT hybrid PKS/NRPS pathways* Both the Walsh and Kelleher laboratories have extensively investigated the mycosubtilin biosynthetic pathway *in vitro*. For example, using a combination of radio-SDS-PAGE and high-resolution MS, the novel PLP-dependent AMT domain of MycA was examined for its ability to convert a β-keto group to a β-amine functionality.[193] Fourier transform ion cyclotron resonance MS (FTICR-MS) studies have demonstrated that this reaction is reversible.[193] It is anticipated that further characterization of this novel PK/NRP catalytic domain might enable the development of a new tool for the introduction of amino groups into heterologous pathways.[193]

The loading of the fatty acid-derived starter unit as an acyl-adenylate has also been explored with radio-SDS-PAGE, TLC, and FTICR-MS.[195] Interestingly, these studies have demonstrated that the loading of variable length chains of both saturated and unsaturated fatty acids onto the AL-ACP didomain is relatively well tolerated.[195] *In vitro* experiments have been conducted to probe catalysis of the mycosubtilin TE domain as well as the *trans*-AT domain FenF.[253] The TE domain demonstrated the ability to catalyze cyclization or hydrolysis of a variety of synthetic peptide substrates, whereas FenF was observed to rapidly transfer malonate to relevant ACPs, as well as to PCPs with lower efficiency. Interestingly, FenF showed a preference for selecting and loading malonyl- versus methylmalonyl- or acetyl-CoA.[196]

1.11.3.3.5 Virginiamycin M

1.11.3.3.5(i) Biological activity and structure of trans-*AT hybrid PK/NRP* Virginiamycin M (**Figure 6**) is a hybrid PK/NRP natural product that is produced by *Streptomyces virginiae* through a *trans*-AT biosynthetic pathway.[254] Structurally, virginiamycin M is composed of a 23-membered macrocyclic ring that contains both oxazole and pyrrole heterocycles. This PKS/NRPS-derived natural product is active as part of a two-component antibiotic in combination with virginiamycin S.[255] Although not approved for human use, virginiamycin M has been extensively employed as an animal-feed additive to promote growth in chickens, turkeys, swine, and cattle. Opinions vary on the use of this product, and there is concern about the development of cross-resistance with the related human antibiotic Synercid that is used for treatment of resistant hospital 'superbugs.'[256,257]

1.11.3.3.5(ii) PKS/NRPS pathways and in vivo *analysis of* trans-*AT hybrid PK/NRP* The virginiamycin M biosynthetic pathway (**Figure 7**) has been sequenced and found to contain an HMGS-type cassette in addition to eight PKS extension modules and two NRPS modules.[254] Many of the PKS modules are split between two polypeptides. The biosynthesis of virginiamycin begins at VirA. This polypeptide features an unusual ACP domain that is located immediately before the two PKS modules. The second PKS module terminates with a

tandem ACP, which is followed by an NRPS extension module and then two additional PKS modules. Again, this second PKS module terminates with a tandem ACP. Next, an HMGS-type gene cassette is found as a series of four discrete enzymes that include VirB (KS), VirC (HCS), and VirD and VirE (enoyl-CoA hydratase). Subsequently, a didomain, VirF, consisting of a KS domain of unknown function begins a split module that is continued in VirG. Following completion of the split module, VirG continues with an additional PKS module and terminates with tandem cyclization domains of a split NRPS module. The remainder of the NRPS module is continued in VirH, which terminates with a KS and ACP domain sequence. Several discrete proteins are also found as part of the virginiamycin biosynthetic pathway. These include VirI (*trans*-AT), VirK (PPT), and VirJ (thioesterase). *In vivo* investigations have suggested that the expression level of VirI (*trans*-AT) may represent a rate-controlling step for virginamycin production.[254] Indeed, complementation of the *trans*-AT by an over-expression vector resulted in 1.5-fold increase in virginiamycin M production.[254] A transcriptional analysis of the biosynthetic genes has also been performed, the results of which indicate that expression of most of the genes was induced 12-h postcultivation, correlating well with virginiamycin production.[258]

1.11.3.4 Unknown Biological Activity

1.11.3.4.1 Thailandamide A

1.11.3.4.1(i) Biological activity and structure of* trans-*AT hybrid PK/NRP Thailandamide (**Figure 6**) is a complex, linear, polyene structure containing an amide bond linkage and terminating in a phenol at one end, and a free acid at the other. No biological data has been reported to date for this compound.

1.11.3.4.1(ii) PKS/NRPS pathways and* in vivo *analysis of* trans-*AT hybrid PK/NRP The thailandamide A biosynthetic pathway (**Figure 7**) was discovered as an orphan locus in *Bacillus thailandensis*. Sequence analysis indicated that this biosynthetic pathway combines 16 PKS elongation modules and a single NRPS module.[190] Briefly, four PKS modules are encoded on the first gene in the pathway. On the second, biosynthesis starts with an NRPS, followed by three PKS modules and a split module, one containing a tandem ACP. The split PKS module then advances onto the next polypeptide with four more PKS modules, two of which contain tandem ACPs. Thailandamide biosynthesis continues in the penultimate polypeptide with two PKS modules and a split module. The final split module starts with an OM and chain elongation by one more module that includes the terminal TE. Using KS-guided homology, based on a new clade-related model, it was possible to predict the structure of the new compound, which was then verified by LC–MS and NMR analysis.[190] The traditional, colinearity-based prediction of structure was significantly less accurate compared to the clade-grouping bioinformatic approach.[190] As with bacillaene, this is another example of 'pathway-first' product elucidation based on genome mining. The pathway assignment has not been verified through gene disruption or heterologous expression.

1.11.4 Engineering, Technology Development, and Other Future Directions for Hybrid PK/NRP Systems

1.11.4.1 Overview

Natural products have been the source or inspiration for numerous pharmaceutical agents utilized for human health; however, the development of many natural product leads is often hindered due to the low quantities of material that can be harvested from the microbial source. In some cases, the natural product producing organism is not amenable to laboratory culture conditions. Furthermore, the structural complexity displayed by many natural product compounds often limits the ability of chemical synthesis to access the compounds of interest and analogues thereof. Through careful, comprehensive investigations of natural product biosynthesis, it may become possible to access the biosynthetic potential of unculturable organisms by heterologous expression in culturable laboratory host. Furthermore, as our understanding of these biosynthetic pathways grows, the objective of facile combinatorial or rational engineering of these biosynthetic pathways becomes more realistic. While great efforts made in these areas have been extensively reviewed elsewhere,[10,176,259,260] developing technologies, when applied to *trans*-AT systems and hybrid

PKS/NRPS pathways, are worthy of further discussion. These developments can be divided into several categories that include heterologous expression, culture conditions, sequencing, protein analytics, and structural analysis.

1.11.4.2 Heterologous Expression and Pathway Engineering in Hybrid PK/NRP Systems

Several *trans*-AT hybrid PKS/NRPS pathways have been accessed through heterologous expression. For example, albicidin has been heterologously produced in the much faster-growing *Xanthomonas axonopodis* strain.[261] The transfer of the entire albicidin biosynthetic gene cluster to a more rapidly growing host organism resulted in an impressive sixfold increase in albicidin production.[261] In the case of leinamycin, the overexpression of the *trans*-AT domain has resulted in a three to fourfold increase in production of the desired metabolite.[191] These results suggest that the producing organism is not always optimized for maximal production under laboratory conditions. This key insight forms the basis for two classical methods of optimizing natural product yields: (1) strain improvement and (2) fermentation development.

1.11.4.3 Culture Conditions and Media Optimization in Hybrid PK/NRP Systems

Increased production of the *trans*-AT hybrid PKS/NRPS product mycosubtilin has focused on both strain improvement and fermentation development. Sophisticated fed-batch culture techniques have enabled high levels of production to be achieved, with a yield of 1.3 mg of mycosubtilin per gram of biomass per hour.[262] This represents substantially higher production of mycosubtilin than what has previously been achieved.[262] Methods have also been developed to effectively transform the producing strain, enabling the ability to monitor the mycosubtilin biosynthesis under various conditions.[263] For example, replacement of the endogenous pathway promoter with a constitutive promoter resulted in a 15-fold increase in production when compared to wild type.[264] This finding underscores the benefit of incorporating relatively straightforward genetic changes into the producing organism. Changes in simple experimental variables, such as reducing the culture temperature from 37 °C to 25 °C, resulted in a 30-fold increase in mycosubtilin production.[265] As illustrated in the mycosubtilin example, methods incorporating strain improvement and media development can have dramatic effects on production efficiency; however, not all organisms are genetically tractable to manipulation, and some even are resistant to growing in laboratory culture (i.e., nonculturable). Thus, despite their potential, these powerful techniques cannot be used to optimize metabolite production in all desirable cases.

1.11.4.4 DNA-Sequencing Strategies in Hybrid PK/NRP Systems

When searching for bioactive compounds in nature, or the genes that direct their biosynthesis, it is best to survey a relatively large pool of structural or genetic diversity. However, given that rRNA analysis suggests that <0.1% of bacterial species collected in a marine sample are amenable to traditional laboratory-culturing techniques, alternative approaches become necessary.[266] One strategy involves the collection of whole environmental DNA (eDNA) and subsequent screening for biosynthetic gene clusters based on homology to known genes. The downside of this approach is that it typically requires densely populated DNA libraries (e.g., pederin and onnamide). Even having access to large DNA libraries does not ensure successful identification of desired gene clusters, as was demonstrated in the recent search for the discodermolide biosynthetic pathway from the sponge *Discodermia disoluta*. Screening of more than 150 000 cosmids produced over 4 GB of DNA sequence data, without providing a viable lead.[267] Interestingly 90% of the DNA sequence appeared to be bacterial in nature thereby suggesting that the sponge does contain a diverse microbial community with high biosynthetic potential.[267] Because of this promise, investigators continue to pursue the development techniques to enable the efficient manipulation and screening of huge pools of DNA, such as clone pooling in semiliquid.[268]

The rise of rapid and inexpensive whole-genome DNA sequencing is also expected to have a profound impact on the ability to access data from environmental samples of microbial consortia. Early experiments have demonstrated that there are seven diverse biosynthetic pathways in a single strain of *Salinispora*

tropica.[269] Genome mining has also been applied in the search for rhizoxin pathway homologues in other source strains.[270] In the near future, the ready access to inexpensive, high-throughput DNA sequencing will undoubtedly enable direct targeting of PKS and NRPS pathways from diverse metagenomic samples.[271]

1.11.4.5 Mass Spectrometry in Hybrid PK/NRP Systems

High-performance MS, particularly FTICR-MS experiments, reported by Kelleher and coworkers,[272] has been used to characterize a wide variety of enzyme-bound intermediates in PKS, NRPS, and hybrid PKS/NRPS pathways. Application of this technology has greatly enabled analysis of enzyme kinetics using radiolabel-free approaches by identifying and characterizing intermediates with a high degree of sensitivity and selectivity.[272] FTICR-MS has also been applied toward the screening of new pathways. This work has relied on a phage display to express protein segments encoding thiolation domains, which are then identified in a high-throughput manner using loss of the phosphopantetheine prosthetic group as a specific signal.[273] New hybrid approaches, as exemplified by the Dorrestein laboratory are also highly innovative. By using matrix-assisted laser desorbtion ionization (MALDI) imaging to localize natural products to a specific location, micromanipulation can be employed to simplify the environmental sample prior to whole-genome sequencing.[274] Finally, the combination of multiple analytical techniques, such as enzyme kinetics, FTICR-MS, and X-ray crystallography, enables the enzymology of hybrid PKS/NRPS systems to be explored to new depths, as was recently demonstrated in the curacin biosynthetic pathway.[100]

1.11.4.6 Structural Biology in Hybrid PK/NRP Systems

Structural biology, through the techniques of X-ray crystallography and NMR, can enhance our fundamental understanding of how these protein machines manufacture diverse natural products. The rate-limiting step for X-ray crystallography often resides in protein crystallization; however, high-throughput techniques to rapidly clone and express diverse gene constructs offer one viable strategy for overcoming this problem.[275] Fundamental questions, such as the nature of mixed PKS/NRPS systems and whether they have a dimeric (PKS) or monomeric (NRPS) quaternary structure can be addressed with this technique, as was demonstrated with recent crystallographic work in the fatty acid synthase.[276–281] The most structurally well-characterized natural product biosynthetic system is the *cis*-PKS DEBs, which has been explored by X-ray crystallography through a series of excised catalytic domains.[282] One limiting factor in the complete structural determination of these megasynthetases appears to be the overall flexibility of the ACP domains. Thus far, structural information of ACPs has been obtained in the solution phase utilizing NMR.[283–285] Emerging approaches for studying large protein complexes such as cryo-EM may also be integrated with high-resolution X-ray structural information as a step toward full understanding of these fascinating megasynthetases.

1.11.5 Conclusions

By studying the coupled PKS and NRPS programming of hybrid PKS/NRPS pathways, a vast realm of biosynthetic space can be explored. Similarly, products of these pathways present a large pool of novel chemical entities that have been selected during evolution by providing an advantage to the producer/host. By leveraging these two pools of diversity, it is possible to access new tools to treat human health conditions within the sphere of cancer, immunomodulatory, infectious diseases, and other areas, as illustrated above. These clinically relevant PK/NRP products represent a potent and expanding source of clinical leads. The coupling of this biosynthetic and chemical diversity is enabling us to take steps toward bypassing the traditional drawbacks of natural products research by providing facile access to metabolites through fermentation, and modification of existing products through pathway engineering.

Through further investigations of the unusual biosynthetic capabilities of the emerging class of noncanonical *trans*-AT synthases, we can hope to both expand our repertoire of capabilities and further develop our fundamental understanding of the flexibility of PK/NRP biosynthesis. Other developing topics such as the role of symbiosis in natural product biosynthesis seem to be uniquely located within these unusual *trans*-AT

systems. It is encouraging to look forward with the hope of applying modern techniques in molecular biology, biochemistry, and analytical chemistry to further dissect and manipulate the mixed PK/NRP natural products through rational design of their PKS/NRPS biosynthetic machinery.

Acknowledgment

NIH support for research on PKS/NRPS systems in the Sherman laboratory is gratefully acknowledged through grants GM076477, CA108874, and ICBG U01TW007404.

Abbreviations

A	adenylation
ACP	acyl carrier protein
ADe	aspartate decarboxylase
AL	acyl ligase
AMT	aminotransferase
AT	acyltransferase
BAC	bacterial artificial chromosome
C	condensation
CDE	cyclodeaminase
Co	CoA ligase
CoA	coenzyme A
CR	crotonyl reductase
Cy	cyclization
Deh	dehydrogenase
DH	dehydratase
E	epimerization
ECH	enoyl-CoA hydratase
ER	enoylreductase
ESI	electrospray ionization
Est	esterase
FAS	fatty acid synthase
FKBP	FK506-binding protein
FTICR-MS	Fourier transform ion cyclotron resonance-mass spectrometry
GNAT	GCN5-related *N*-acetyltransferase
Hal	halogenase
HCS	HMG-CoA synthase
HDAC	histone deacetylase
HMG	3-hydroxy-3-methylglutarate
HMGS	HMG synthase
HPLC	high-performance liquid chromatography
KGS	2-ketoglutarate synthase
KR	ketoreductase
KS	ketosynthase
LC–MS	liquid chromatography–mass spectrometry
MALDI	matrix-assisted laser desorbtion ionization
MRSA	methicillin-resistant *Staphylococcus aureus*
MS	mass spectrometry
MT	methyltransferase

mTOR	mammalian target of rapamycin
NFAT	nuclear factor of activated T cells
NRP	nonribosomal peptide
NRPS	nonribosomal peptide synthase
OM	*O*-methyltransferase
ORF	open-reading frame
Ox	oxidation
OxR	oxidoreductase
PAPS	3′-phosphoadenosine-5′-phosphosulfate
PCP	peptidyl carrier protein
PCR	polymerase chain reaction
PGP	P-glycoprotein
PIKK	phosphatidylinositol kinase-related kinase
PK	polyketide
PKS	polyketide synthase
PPT	phosphopantetheinyl transferase
SAR	structure–activity relationship
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
ST	sulfotransferase
T	thiolation
TE	thioesterase
TGF-β	transforming growth factor-β
TLC	thin-layer chromatography

References

1. D. H. Williams; M. J. Stone; P. R. Hauck; S. K. Rahman, *J. Nat. Prod.* **1989**, *52* (6), 1189–1208.
2. D. D. Baker; M. Chu; U. Oza; V. Rajgarhia, *Nat. Prod. Rep.* **2007**, *24* (6), 1225–1244.
3. D. J. Newman; G. M. Cragg; K. M. Snader, *Nat. Prod. Rep.* **2000**, *17* (3), 215–234.
4. A. Ganesan, *Curr. Opin. Chem. Biol.* **2008**, *12* (3), 306–317.
5. D. J. Newman; G. M. Cragg, *J. Nat. Prod.* **2007**, *70* (3), 461–477.
6. J. L. Fortman; D. H. Sherman, *Chembiochem* **2005**, *6* (6), 960–978.
7. C. E. Salomon; N. A. Magarvey; D. H. Sherman, *Nat. Prod. Rep.* **2004**, *21* (1), 105–121.
8. H. B. Bode; R. Muller, *Angew. Chem. Int. Ed.* **2005**, *44* (42), 6828–6846.
9. G. L. Challis, *J. Med. Chem.* **2008**, *51* (9), 2618–2628. Epub 2008 Apr 5.
10. S. G. Van Lanen; B. Shen, *Curr. Opin. Microbiol.* **2006**, *9* (3), 252–260.
11. J. Kennedy, *Nat. Prod. Rep.* **2008**, *25* (1), 25–34.
12. B. N. Mijts; C. Schmidt-Dannert, *Curr. Opin. Biotechnol.* **2003**, *14* (6), 597–602.
13. J. D. Kittendorf; D. H. Sherman, *Curr. Opin. Biotechnol.* **2006**, *17* (6), 597.
14. C. D. Reeves, *Crit. Rev. Biotechnol.* **2003**, *23* (2), 95–147.
15. W. Zhang; Y. Tang, *J. Med. Chem.* **2008**, *51* (9), 2629–2633. Epub 2008 Apr 5.
16. H. G. Menzella; J. R. Carney; D. V. Santi, *Chem. Biol.* **2007**, *14* (2), 143–151.
17. R. Finking; M. A. Marahiel, *Annu. Rev. Microbiol.* **2004**, *58*, 453–488.
18. M. A. Fischbach; C. T. Walsh, *Chem. Rev.* **2006**, *106* (8), 3468–3496.
19. J. Staunton; K. J. Weissman, *Nat. Prod. Rep.* **2001**, *18* (4), 380–416.
20. K. J. Weissman; P. F. Leadlay, *Nat. Rev. Microbiol.* **2005**, *3* (12), 925–936.
21. E. A. Felnagle; E. E. Jackson; Y. A. Chan; A. M. Podevels; A. D. Berti; M. D. McMahon; M. G. Thomas, *Mol. Pharm.* **2008**, *5* (2), 191–211.
22. S. C. Wenzel; R. Muller, *Nat. Prod. Rep.* **2007**, *24* (6), 1211–1224.
23. B. Shen; L. Du; C. Sanchez; M. Chen; D. J. Edwards, *Bioorg. Chem.* **1999**, *27*, 155–171.
24. B. Shen; L. Du; C. Sanchez; D. J. Edwards; M. Chen; J. M. Murrell, *J. Nat. Prod.* **2002**, *65* (3), 422–431.
25. L. Du; Y. Q. Cheng; G. Ingenhorst; G. L. Tang; Y. Huang; B. Shen, *Genet. Eng.* **2003**, *25*, 227–267.
26. G. Hoefle; N. Bedorf; H. Steinmetz; D. Schomburg; K. Gerth; H. Reichenbach, *Angew. Chem. Int. Ed.* **1996**, *35*, 1567–1569.
27. K. Gerth; N. Bedorf; G. Hofle; H. Irschik; H. Reichenbach, *J. Antibiot. (Tokyo)* **1996**, *49* (6), 560–563.
28. D. M. Bollag; P. A. McQueney; J. Zhu; O. Hensens; L. Koupal; J. Liesch; M. Goetz; E. Lazarides; C. M. Woods, *Cancer Res.* **1995**, *55* (11), 2325–2333.

29. L. He; G. A. Orr; S. B. Horwitz, *Drug Discovery Today* **2001**, *6* (22), 1153–1164.
30. L. A. Martello; H. M. McDaid; D. L. Regl; C. P. Yang; D. Meng; T. R. Pettus; M. D. Kaufman; H. Arimoto; S. J. Danishefsky; A. B. Smith, III; S. B. Horwitz, *Clin. Cancer Res.* **2000**, *6* (5), 1978–1987.
31. R. L. Arslanian; L. Tang; S. Blough; W. Ma; R. G. Qiu; L. Katz; J. R. Carney, *J. Nat. Prod.* **2002**, *65* (7), 1061–1064.
32. I. H. Hardt; H. Steinmetz; K. Gerth; F. Sasse; H. Reichenbach; G. Hofle, *J. Nat. Prod.* **2001**, *64* (7), 847–856.
33. C. M. Starks; Y. Zhou; F. Liu; P. J. Licari, *J. Nat. Prod.* **2003**, *66* (10), 1313–1317.
34. K. C. Nicolaou, *Angew. Chem. Int. Ed.* **1998**, *37* (15), 2014.
35. K. C. Nicolaou, *Chem. Commun.* **2001**, (17), 1523.
36. E. B. Watkins, *Eur. J. Org. Chem.* **2006**, *2006* (18), 4071.
37. E. B. Watkins; A. G. Chittiboyina; J.-C. Jung; M. A. Avery, *Curr. Pharm. Des.* **2005**, *11*, 1615–1653.
38. K. H. Altmann, *Org. Biomol. Chem.* **2004**, *2* (15), 2137.
39. H. M. McDaid; S. Mani; H. J. Shen; F. Muggia; D. Sonnichsen; S. B. Horwitz, *Clin. Cancer Res.* **2002**, *8* (7), 2035–2043.
40. F. Y. Lee; R. Borzilleri; C. R. Fairchild; S. H. Kim; B. H. Long; C. Reventos-Suarez; G. D. Vite; W. C. Rose; R. A. Kramer, *Clin. Cancer Res.* **2001**, *7* (5), 1429–1437.
41. S. Okuno; W. J. Maples; M. R. Mahoney; T. Fitch; J. Stewart; P. M. Fracasso; M. Kraut; D. S. Ettinger; F. Dawkins; C. Erlichman, *J. Clin. Oncol.* **2005**, *23* (13), 3069–3073.
42. S. M. Gadgeel; A. Wozniak; R. R. Boinpally; R. Wiegand; L. K. Heilbrun; V. Jain; R. Parchment; D. Colevas; M. B. Cohen; P. M. LoRusso, *Clin. Cancer Res.* **2005**, *11* (17), 6233–6239.
43. S. Mani; H. McDaid; A. Hamilton; H. Hochster; M. B. Cohen; D. Khabelle; T. Griffin; D. E. Lebwohl; L. Liebes; F. Muggia; S. B. Horwitz, *Clin. Cancer Res.* **2004**, *10* (4), 1289–1298.
44. N. Lin; K. Brakora; M. Seiden, *Curr. Opin. Invest. Drugs* **2003**, *4* (6), 746–756.
45. S. H. Zhuang; M. Agrawal; M. Edgerly; S. Bakke; H. Kotz; P. Thambi; A. Rutt; F. M. Balis; S. Bates; T. Fojo, *Cancer* **2005**, *103* (9), 1932–1938.
46. A. Kolman, *Curr. Opin. Invest. Drugs* **2005**, *6* (6), 616–622.
47. J. J. Lee; S. M. Swain, *Semin. Oncol.* **2005**, *32* (6 Suppl 7), S22–S26.
48. S. Mani; M. Macapinlac, Jr.; S. Goel; D. Verdier-Pinard; T. Fojo; M. Rothenberg; D. Colevas, *Anticancer Drugs* **2004**, *15* (6), 553–558.
49. F. Pellegrini; D. R. Budman, *Cancer Invest.* **2005**, *23* (3), 264–273.
50. E. H. Rubin; J. Rothermel; F. Tesfaye; T. Chen; M. Hubert; Y. Y. Ho; C. H. Hsu; A. M. Oza, *J. Clin. Oncol.* **2005**, *23* (36), 9120–9129.
51. J. Rothermel; M. Wartmann; T. Chen; J. Hohneker, *Semin. Oncol.* **2003**, *30* (3 Suppl 6), 51–55.
52. T. O'Reilly; P. M. McSheehy; F. Wenger; M. Hattenberger; M. Muller; J. Vaxelaire; K. H. Altmann; M. Wartmann, *Prostate* **2005**, *65* (3), 231–240.
53. K. Pietras; M. Stumm; M. Hubert; E. Buchdunger; K. Rubin; C. H. Heldin; P. McSheehy; M. Wartmann; A. Ostman, *Clin. Cancer Res.* **2003**, *9* (10 Pt 1), 3779–3787.
54. H. Wang; Z. Wang; S. Wang; M. Li; L. Nan; J. K. Rhie; J. M. Covey; R. Zhang; D. L. Hill, *Cancer Chemother. Pharmacol.* **2005**, *56* (3), 255–260.
55. T. C. Chou; X. G. Zhang; C. R. Harris; S. D. Kuduk; A. Balog; K. A. Savin; J. R. Bertino; S. J. Danishefsky, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95* (26), 15798–15802.
56. T. C. Chou; X. G. Zhang; A. Balog; D. S. Su; D. Meng; K. Savin; J. R. Bertino; S. J. Danishefsky, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95* (16), 9642–9647.
57. A. V. Kamath; M. Chang; F. Y. Lee; Y. Zhang; P. H. Marathe, *Cancer Chemother. Pharmacol.* **2005**, *56* (2), 145–153.
58. A. Kolman, *Curr. Opin. Invest. Drugs* **2004**, *5* (12), 1292–1297.
59. U. Klar; B. Buchmann; W. Schwede; W. Skuballa; J. Hoffmann; R. B. Lichtner, *Angew. Chem. Int. Ed. Engl.* **2006**, *45* (47), 7942–7948.
60. B. Julien; S. Shah; R. Ziermann; R. Goldman; L. Katz; C. Khosla, *Gene* **2000**, *249* (1–2), 153–160.
61. I. Molnar; T. Schupp; M. Ono; R. Zirkle; M. Milnamow; B. Nowak-Thompson; N. Engel; C. Toupet; A. Stratmann; D. D. Cyr; J. Gorlach; J. M. Mayo; A. Hu; S. Goff; J. Schmid; J. M. Ligon, *Chem. Biol.* **2000**, *7* (2), 97–109.
62. L. Tang; S. Shah; L. Chung; J. Carney; L. Katz; C. Khosla; B. Julien, *Science* **2000**, *287* (5453), 640–642.
63. J. G. Olsen; A. Kadziola; P. von Wettstein-Knowles; M. Siggaard-Andersen; Y. Lindquist; S. Larsen, *FEBS Lett.* **1999**, *460* (1), 46–52.
64. M. He; M. Varoglu; D. H. Sherman, *J. Bacteriol.* **2000**, *182* (9), 2619–2623.
65. J. F. Aparicio; I. Molnar; T. Schwecke; A. Konig; S. F. Haydock; L. E. Khaw; J. Staunton; P. F. Leadlay, *Gene* **1996**, *169* (1), 9–16.
66. J. M. Jez; J. L. Ferrer; M. E. Bowman; R. A. Dixon; J. P. Noel, *Biochemistry* **2000**, *39* (5), 890–902.
67. C. Bisang; P. F. Long; J. Cortes; J. Westcott; J. Crosby; A. L. Matharu; R. J. Cox; T. J. Simpson; J. Staunton; P. F. Leadlay, *Nature* **1999**, *401* (6752), 502–505.
68. H. Ogura; C. R. Nishida; U. R. Hoch; R. Perera; J. H. Dawson; P. R. Ortiz de Montellano, *Biochemistry* **2004**, *43* (46), 14712–14721.
69. S. Nagano; H. Li; H. Shimizu; C. Nishida; H. Ogura; P. R. Ortiz de Montellano; T. L. Poulos, *J. Biol. Chem.* **2003**, *278* (45), 44886–44893.
70. H. Umezawa; K. Maeda; T. Takeuchi; Y. Okami, *J. Antibiot. (Tokyo)* **1966**, *19* (5), 200–209.
71. Y. Muraoka; A. Fuji; T. Yoshioka; T. Takita; H. Umezawa, *J. Antibiot. (Tokyo)* **1977**, *30* (2), 178–181.
72. T. Takita; Y. Muraoka; T. Yoshioka; A. Fujii; K. Maeda, *J. Antibiot. (Tokyo)* **1972**, *25* (12), 755–758.
73. T. Takita; Y. Muraoka; T. Nakatani; A. Fujii; Y. Umezawa; H. Naganawa, *J. Antibiot. (Tokyo)* **1978**, *31* (8), 801–804.
74. Y. Aoyagi; K. Katano; H. Suguna; J. Primeau; L. H. Chang; S. M. Hecht, *J. Am. Chem. Soc.* **1982**, *104*, 5537.
75. T. Takita; Y. Umezawa; S. I. Sait; H. Morishima; H. Naganawa; H. Umezawa; T. Tsuchiya; T. Miyake; S. Kageyama; S. Umezawa; Y. Muraoka; M. Suzuki; M. Otsuka; M. Narita; S. Kobayashi; M. Ohno, *Tetrahedron Lett.* **1982**, *23*, 521.
76. Y. Sugiura; T. Suzuki; M. Otsuka; S. Kobayashi; M. Ohno; T. Takita; H. Umezawa, *J. Biol. Chem.* **1983**, *258* (2), 1328–1336.
77. R. M. Burger, *Chem. Rev.* **1998**, *98* (3), 1153–1170.

78. S. M. Hecht, *J. Nat. Prod.* **2000**, *63* (1), 158–168.
79. L. Giloni; M. Takeshita; F. Johnson; C. Iden; A. P. Grollman, *J. Biol. Chem.* **1981**, *256* (16), 8608–8615.
80. H. Umezawa; T. Takita; S. Saito; Y. Muraoka; K. Takahashi; H. Ekimoto; S. Minamide; K. Nishikawa; T. Fukuoka; T. Nakatani; A. Fujii; A. Matsuda, *Bleomycin Chemotherapy*; Academic Press: Orlando, 1985; p 289.
81. M. J. Levy; S. M. Hecht, *Biochemistry* **1988**, *27* (8), 2647–2650.
82. L. M. Fisher; R. Kuroda; T. T. Sakai, *Biochemistry* **1985**, *24* (13), 3199–3207.
83. J. Kross; W. D. Henner; W. A. Haseltine; L. Rodriguez; M. D. Levin; S. M. Hecht, *Biochemistry* **1982**, *21* (15), 3711–3721.
84. J. Kross; W. D. Henner; S. M. Hecht; W. A. Haseltine, *Biochemistry* **1982**, *21* (18), 4310–4318.
85. L. F. Povirk; M. Hogan; N. Dattagupta, *Biochemistry* **1979**, *18* (1), 96–101.
86. M. Chien; A. P. Grollman; S. B. Horwitz, *Biochemistry* **1977**, *16* (16), 2641–2647.
87. P. L. Tran; J. Weinbach; P. Opolon; G. Linares-Cruz; J. P. Reynes; A. Gregoire; E. Kremer; H. Durand; M. Perricaudet, *J. Clin. Invest.* **1997**, *99* (4), 608–617.
88. W. G. Martin; K. M. Ristow; T. M. Habermann; J. P. Colgan; T. E. Witzig; S. M. Ansell, *J. Clin. Oncol.* **2005**, *23* (30), 7614–7620.
89. E. A. Tueni; R. A. Newman; F. L. Baker; J. A. Ajani; D. Fan; G. Spitzer, *Cancer Res.* **1989**, *49* (5), 1099–1102.
90. R. A. Newman; Z. H. Siddik; E. L. Travis; D. Followill; W. Ayele; T. Burditt; I. H. Krakoff, *Invest. New Drugs* **1990**, *8* (1), 33–41.
91. M. Sugiyama; C. J. Thompson; T. Kumagai; K. Suzuki; R. Deblaere; R. Villarroel; J. Davies, *Gene* **1994**, *151* (1–2), 11–16.
92. L. Du; C. Sanchez; M. Chen; D. J. Edwards; B. Shen, *Chem. Biol.* **2000**, *7* (8), 623–642.
93. P. Verdier-Pinard; J. Y. Lai; H. D. Yoo; J. Yu; B. Marquez; D. G. Nagle; M. Nambu; J. D. White; J. R. Falck; W. H. Gerwick; B. W. Day; E. Hamel, *Mol. Pharmacol.* **1998**, *53* (1), 62–76.
94. W. H. Gerwick; P. J. Proteau; D. G. Nagle; E. Hamel; A. Blokhin; D. L. Slate, *J. Org. Chem.* **1994**, *59* (6), 1243–1245.
95. A. V. Blokhin; H. D. Yoo; R. S. Geralds; D. G. Nagle; W. H. Gerwick; E. Hamel, *Mol. Pharmacol.* **1995**, *48* (3), 523–531.
96. P. Wipf; J. T. Reeves; R. Balachandran; B. W. Day, *J. Med. Chem.* **2002**, *45* (9), 1901–1917.
97. P. Wipf; J. T. Reeves; B. W. Day, *Curr. Pharm. Des.* **2004**, *10* (12), 1417–1437.
98. Z. Chang; N. Sitachitta; J. V. Rossi; M. A. Roberts; P. M. Flatt; J. Jia; D. H. Sherman; W. H. Gerwick, *J. Nat. Prod.* **2004**, *67* (8), 1356–1367.
99. J. Piel; G. Wen; M. Platzer; D. Hui, *Chembiochem* **2004**, *5* (1), 93–98.
100. L. Gu; T. W. Geders; B. Wang; W. H. Gerwick; K. Hakansson; J. L. Smith; D. H. Sherman, *Science* **2007**, *318* (5852), 970.
101. L. Gu; J. Jia; H. L. Liu; K. Hakansson; W. H. Gerwick; D. H. Sherman, *J. Am. Chem. Soc.* **2006**, *128*, 9014–9015.
102. T. W. Geders; L. Gu; J. C. Mowers; H. Liu; W. H. Gerwick; K. Hakansson; D. H. Sherman; J. L. Smith, *J. Biol. Chem.* **2007**, *282* (49), 35954–35963.
103. M. Negishi; L. G. Pedersen; E. Petrotchenko; S. Shevtsov; A. Gorokhov; Y. Kakuta; L. C. Pedersen, *Arch. Biochem. Biophys.* **2001**, *390* (2), 149–157.
104. H. Ueda; H. Nakajima; Y. Hori; T. Fujita; M. Nishimura; T. Goto; M. Okuhara, *J. Antibiot. (Tokyo)* **1994**, *47* (3), 301–310.
105. H. Ueda; H. Nakajima; Y. Hori; T. Goto; M. Okuhara, *Biosci. Biotechnol. Biochem.* **1994**, *58* (9), 1579–1583.
106. K. Garber, *Nat. Biotechnol.* **2007**, *25* (1), 17–19.
107. D. M. Vigushin, *Curr. Opin. Invest. Drugs* **2002**, *3* (9), 1396–1402.
108. H. Nakajima; Y. B. Kim; H. Terano; M. Yoshida; S. Horinouchi, *Exp. Cell Res.* **1998**, *241* (1), 126–133.
109. N. Shigematsu; H. Ueda; S. Takase; H. Tanaka; K. Yamamoto; T. Tada, *J. Antibiot. (Tokyo)* **1994**, *47* (3), 311–314.
110. K. Li; J. Wu; W. Xing; J. A. Simon, *J. Am. Chem. Soc.* **1996**, *118*, 7237–7238.
111. R. Furumai; A. Matsuyama; N. Kobashi; K. H. Lee; M. Nishiyama; H. Nakajima; A. Tanaka; Y. Komatsu; N. Nishino; M. Yoshida; S. Horinouchi, *Cancer Res.* **2002**, *62* (17), 4916–4921.
112. J. C. Byrd; C. Shinn; R. Ravi; C. R. Willis; J. K. Waselenko; I. W. Flinn; N. A. Dawson; M. R. Grever, *Blood* **1999**, *94* (4), 1401–1408.
113. Y. Q. Cheng; M. Yang; A. M. Matter, *Appl. Environ. Microbiol.* **2007**, *73* (11), 3460–3469.
114. R. E. Schwartz; C. F. Hirsch; D. F. Sesin; J. E. Flor; M. Chartrain; R. E. Fromtling; G. H. Harris; M. J. Salvatore; J. M. Liesch; K. Yudin, *J. Ind. Microbiol.* **1990**, *5*, 113–124.
115. G. Trimurtulu; I. Ohtani; G. Patterson; R. E. Moore; T. H. Corbett; F. A. Valeriote; L. Demchik, *J. Am. Chem. Soc.* **1994**, *116*, 4729–4737.
116. T. H. Corbett; F. A. Valeriote; L. Demchik; L. Polin; C. Panchapor; S. Pugh; K. White; J. Knight; J. Jones; L. Jones; P. LoRusso; B. Foster; R. A. Wiegand; L. Lisow; T. Golakoti; C. E. Heltzel; J. Ogino; G. M. Patterson; R. E. Moore, *J. Exp. Ther. Oncol.* **1996**, *1* (2), 95–108.
117. D. Panda; R. H. Himes; R. E. Moore; L. Wilson; M. A. Jordan, *Biochemistry* **1997**, *36* (42), 12948–12953.
118. C. D. Smith; X. Zhang; S. L. Mooberry; G. M. Patterson; R. E. Moore, *Cancer Res.* **1994**, *54* (14), 3779–3784.
119. G. V. Subbaraju; T. Golakoti; G. M. Patterson; R. E. Moore, *J. Nat. Prod.* **1997**, *60* (3), 302–305.
120. S. Chaganty; T. Golakoti; C. Heltzel; R. E. Moore; W. Y. Yoshida, *J. Nat. Prod.* **2004**, *67* (8), 1403–1406.
121. K. Lu; J. Dempsey; R. M. Schultz; C. Shih; B. A. Teicher, *Cancer Chemother. Pharmacol.* **2001**, *47* (2), 170–178.
122. M. J. Edelman; D. R. Gandara; P. Hausner; V. Israel; D. Thornton; J. DeSanto; L. A. Doyle, *Lung Cancer* **2003**, *39* (2), 197–199.
123. G. D'Agostino; J. del Campo; B. Mellado; M. A. Izquierdo; T. Minarik; L. Cirri; L. Marini; J. L. Perez-Gracia; G. Scambia, *Int. J. Gynecol. Cancer* **2006**, *16* (1), 71–76.
124. J. Liang; R. E. Moore; E. D. Moher; J. E. Munroe; R. S. Al-awar; D. A. Hay; D. L. Varie; T. Y. Zhang; J. A. Aikins; M. J. Martinelli; C. Shih; J. E. Ray; L. L. Gibson; V. Vasudevan; L. Polin; K. White; J. Kushner; C. Simpson; S. Pugh; T. H. Corbett, *Invest. New Drugs* **2005**, *23* (3), 213–224.
125. N. A. Magarvey; Z. Q. Beck; T. Golakoti; Y. Ding; U. Huber; T. K. Hemscheidt; D. Abelson; R. E. Moore; D. H. Sherman, *ACS Chem. Biol.* **2006**, *1* (12), 766–779.
126. W. Seufert; Z. Q. Beck; D. H. Sherman, *Angew. Chem. Int. Ed.* **2007**, *46* (48), 9298–9300.
127. Z. Q. Beck; C. C. Aldrich; N. A. Magarvey; G. I. Georg; D. H. Sherman, *Biochemistry* **2005**, *44* (41), 13457–13466.
128. Z. Q. Beck; D. A. Burr; D. H. Sherman, *Chembiochem* **2007**, *8* (12), 1373–1375.
129. Y. Ding; W. H. Seufert; Z. Q. Beck; D. H. Sherman, *J. Am. Chem. Soc.* **2008**, *130* (16), 5492–5498.
130. S. N. Sehgal; H. Baker; C. Vezina, *J. Antibiot. (Tokyo)* **1975**, *28* (10), 727–732.

131. C. Vezina; A. Kudelski; S. N. Sehgal, *J. Antibiot. (Tokyo)* **1975**, *28* (10), 721–726.
132. N. Swindells; P. S. White; J. A. Findlay, *Can. J. Chem.* **1978**, *56*, 2491.
133. J. A. Findlay; L. Radics, *Can. J. Chem.* **1980**, *58*, 579–590.
134. A. B. Smith; S. M. Condon; J. A. McCauley; J. L. Leazer; J. W. Leahy; R. E. Maleczka, *J. Am. Chem. Soc.* **1997**, *119*, 947–961.
135. C. M. Hayward; D. Yohannes; S. J. Danishefsky, *J. Am. Chem. Soc.* **1993**, *115* (20), 9345–9346.
136. K. C. Nicolaou; T. K. Chakraborty; A. D. Piscopio; N. Minowa; P. Bertinato, *J. Am. Chem. Soc.* **1993**, *115* (10), 4419–4420.
137. D. Romo; S. D. Meyer; D. D. Johnson; S. L. Schreiber, *J. Am. Chem. Soc.* **1993**, *115* (17), 7906–7907.
138. B. K. Law, *Crit. Rev. Oncol. Hematol.* **2005**, *56* (1), 47–60.
139. J. Kunz; M. N. Hall, *Trends Biochem. Sci.* **1993**, *18* (9), 334–338.
140. S. N. Sehgal, *Ther. Drug Monit.* **1995**, *17* (6), 660–665.
141. N. R. Brook; J. R. Waller; G. R. Bicknell; M. L. Nicholson, *Transplant Proc.* **2005**, *37* (2), 837–838.
142. N. L. Paiva; A. L. Demain; M. F. Roberts, *J. Nat. Prod.* **1991**, *54* (1), 167–177.
143. T. Schwecke; J. F. Aparicio; I. Molnar; A. Konig; L. E. Khaw; S. F. Haycock; M. Oliynyk; P. Caffrey; J. Cortes; J. B. Lester, *Proc. Natl. Acad. Sci.* **1995**, *92* (17), 7839–7843.
144. A. Konig; T. Schwecke; I. Molnar; G. A. Bohm; P. A. Lowden; J. Staunton; P. F. Leadlay, *Eur. J. Biochem.* **1997**, *247* (2), 526–534.
145. N. Lomovskaya; L. Fonstein; X. Ruan; D. Stassi; L. Katz; C. R. Hutchinson, *Microbiology* **1997**, *143* (Pt 3), 875–883.
146. T. Kino; H. Hatanaka; M. Hashimoto; M. Nishiyama; T. Goto; M. Okuhara; M. Kohsaka; H. Aoki; H. Imanaka, *J. Antibiot. (Tokyo)* **1987**, *40* (9), 1249–1255.
147. T. Kino; H. Hatanaka; S. Miyata; N. Inamura; M. Nishiyama; T. Yajima; T. Goto; M. Okuhara; M. Kohsaka; H. Aoki, *J. Antibiot. (Tokyo)* **1987**, *40* (9), 1256–1265.
148. H. Tanaka; A. Kuroda; H. Marusawa; H. Hatanaka; T. Kino; T. Goto; M. Hashimoto; T. Taga, *J. Am. Chem. Soc.* **1987**, *109* (16), 5031–5033.
149. S. Ho; N. Clipstone; L. Timmermann; J. Northrop; I. Graef; D. Fiorentino; J. Nourse; G. R. Crabtree, *Clin. Immunol. Immunopathol.* **1996**, *80* (3 Pt 2), S40–S45.
150. F. J. Dumont; M. J. Staruch; S. L. Koprak; J. J. Siekierka; C. S. Lin; R. Harrison; T. Sewell; V. M. Kindt; T. R. Beattie; M. Wyvratt, *J. Exp. Med.* **1992**, *176* (3), 751–760.
151. H. Motamedi; A. Shafiee, *Eur. J. Biochem.* **1998**, *256* (3), 528–534.
152. H. Motamedi; S. J. Cai; A. Shafiee; K. O. Elliston, *Eur. J. Biochem.* **1997**, *244* (1), 74–80.
153. M. Arai; K. Fukuhara; S. Nakamura; Y. Sakagami; H. Yonehara, *J. Antibiot. (Tokyo)* **1956**, *9* (5), 193.
154. C. Cocito, *Microbiol. Rev.* **1979**, *43* (2), 145–192.
155. M. Aumercier; S. Bouhallab; M. L. Capmau; L. F. Goffic, *J. Antimicrob. Chemother.* **1992**, *30* (Suppl A), 9–14.
156. B. Ronan; E. Bacqué; J. C. Barrière; S. Sablé, *Tetrahedron* **2003**, *59* (16), 2929–2937.
157. M. B. Purvis; J. W. LeFevre; V. L. Jones; D. G. I. Kingston; A. M. Biot; F. Gossele, *J. Am. Chem. Soc.* **1989**, *111* (15), 5931–5935.
158. J. W. LeFevre; D. G. I. Kingston, *J. Org. Chem.* **1984**, *49* (14), 2588–2593.
159. D. G. I. Kingston; M. X. Kolpak, *J. Am. Chem. Soc.* **1980**, *102* (18), 5964–5966.
160. D. G. I. Kingston; M. X. Kolpak; J. W. LeFevre; I. Borup-Grochtmann, *J. Am. Chem. Soc.* **1983**, *105* (15), 5106–5110.
161. V. Blanc; D. Lagneaux; P. Didier; P. Gil; P. Lacroix; J. Crouzet, *J. Bacteriol.* **1995**, *177* (18), 5206–5214.
162. D. Thibaut; D. Bisch; N. Ratet; L. Maton; M. Couder; L. Debussche; F. Blanche, *J. Bacteriol.* **1997**, *179* (3), 697–704.
163. V. de Crecy-Lagard; W. Saurin; D. Thibaut; P. Gil; L. Naudin; J. Crouzet; V. Blanc, *Antimicrob. Agents Chemother.* **1997**, *41* (9), 1904–1909.
164. V. de Crecy-Lagard; V. Blanc; P. Gil; L. Naudin; S. Lorenzon; A. Famechon; N. Bamas-Jacques; J. Crouzet; D. Thibaut, *J. Bacteriol.* **1997**, *179* (3), 705–713.
165. D. Thibaut; N. Ratet; D. Bisch; D. Faucher; L. Debussche; F. Blanche, *J. Bacteriol.* **1995**, *177* (18), 5199–5205.
166. G. Sezonov; V. Blanc; N. Bamas-Jacques; A. Friedmann; J. L. Pernodet; M. Guerineau, *Nat. Biotechnol.* **1997**, *15* (4), 349–353.
167. W. Trowitzsch; V. Wray; K. Gerth; G. Hofle, *J. Chem. Soc. Chem. Commun.***1982**, (23),1340–1342.
168. S. Fytlovitch; P. D. Nathan; D. Zafriri; E. Rosenberg, *J. Antibiot. (Tokyo)* **1983**, *36* (11), 1525–1530.
169. E. Rosenberg; S. Fytlovitch; S. Carmeli; Y. Kashman, *J. Antibiot. (Tokyo)* **1982**, *35* (7), 788–793.
170. E. Rosenberg; B. Vaks; A. Zuckerberg, *Antimicrob. Agents Chemother.* **1973**, *4* (5), 507–513.
171. A. Manor; M. Varon; E. Rosenberg, *J. Dent. Res.* **1985**, *64* (12), 1371–1373.
172. A. Manor; I. Eli; M. Varon; H. Judes; E. Rosenberg, *J. Clin. Periodontol* **1989**, *16* (10), 621–624.
173. E. Simhi; H. C. van der Mei; E. Z. Ron; E. Rosenberg; H. J. Busscher, *FEMS Microbiol. Lett.* **2000**, *192* (1), 97–100.
174. V. Simunovic; J. Zapp; S. Rachid; D. Krug; P. Meiser; R. Muller, *Chembiochem* **2006**, *7* (8), 1206–1220.
175. G.-L. Tang; Y.-Q. Cheng; B. Shen, *Chem. Biol.* **2004**, *11* (1), 33–45.
176. L. Du; B. Shen, *Curr. Opin. Drug Discovery* **2001**, *4* (2), 215–228.
177. C. T. Walsh, *Science* **2004**, *303* (5665), 1805–1810.
178. J. Piel, *Nat. Prod. Rep.* **2004**, *21* (4), 519.
179. E. W. Schmidt, *Nat. Chem. Biol.* **2008**, *4* (8), 466.
180. S. J. Moss; C. J. Martin; B. Wilkinson, *Nat. Prod. Rep.* **2004**, *21* (5), 575.
181. G. Zhu; M. J. LaGier; F. Stejskal; J. J. Millership; X. Cai; J. S. Keithly, *Gene* **2002**, *298* (1), 79–89.
182. A. K. El-Sayed; J. Hothersall; S. M. Cooper; E. Stephens; T. J. Simpson; C. M. Thomas, *Chem. Biol.* **2003**, *10* (5), 419–430.
183. Y. Ogasawara; K. Katayama; A. Minami; M. Otsuka; T. Eguchi; K. Kakinuma, *Chem. Biol.* **2004**, *11*, 79–86.
184. X.-H. Chen; J. Vater; J. Piel; P. Franke; R. Scholz; K. Schneider; A. Koumoutsi; G. Hitzeroth; N. Grammel; A. W. Strittmatter; G. Gottschalk; R. D. Sussmuth; R. Borriss, *J. Bacteriol.* **2006**, *188* (11), 4024–4036.
185. S. Sudek; N. B. Lopanik; L. E. Waggoner; M. Hildebrand; C. Anderson; H. Liu; A. Patel; D. H. Sherman; M. G. Haygood, *J. Nat. Prod.* **2007**, *70* (1), 67–74.
186. N. B. Lopanik; J. A. Shields; T. J. Buchholz; C. M. Rath; J. Hothersall; M. G. Haygood; K. Hakansson; C. T. Thomas; D. H. Sherman, *Chem. Biol.* **2008**, *15* (11), 1175.
187. C. T. Calderone; W. E. Kowtoniuk; N. L. Kelleher; C. T. Walsh; P. C. Dorrestein, *Proc. Natl. Acad. Sci.* **2006**, *103* (24), 8977.
188. G. Yadav; R. S. Gokhale; D. Mohanty, *J. Mol. Biol.* **2003**, *328* (2), 335–363.

189. H. von Dohren; R. Dieckmann; M. P. Pavela-Vrancic, *Chem. Biol.* **1999**, *6* (10), R273.
190. T. Nguyen; K. Ishida; H. Jenke-Kodama; E. Dittmann; C. Gurgui; T. Hochmuth; S. Taudien; M. Platzer; C. Hertweck; J. Piel, *Nat. Biotechnol.* **2008**, *26* (2), 225–233.
191. Y. Q. Cheng, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (6), 3149.
192. G. L. Tang; Y. Q. Cheng; B. Shen, *J. Nat. Prod.* **2006**, *69* (3), 387–393.
193. Z. D. Aron; P. C. Dorrestein; J. R. Blackhall; N. L. Kelleher; C. T. Walsh, *J. Am. Chem. Soc.* **2005**, *127* (43), 14986–14987.
194. J. J. Reddick; S. A. Antolak; G. M. Raner, *Biochem. Biophys. Res. Commun.* **2007**, *358* (1), 363–367.
195. D. B. Hansen; S. B. Bumpus; Z. D. Aron; N. L. Kelleher; C. T. Walsh, *J. Am. Chem. Soc.* **2007**, *129* (20), 6366–6367.
196. Z. D. Aron; P. D. Fortin; C. T. Calderone; C. T. Walsh, *Chembiochem* **2007**, *8* (6), 613.
197. P. C. Dorrestein; S. B. Bumpus; C. T. Calderone; S. Garneau-Tsodikova; Z. D. Aron; P. D. Straight; R. Kolter; C. T. Walsh; N. L. Kelleher, *Biochemistry* **2006**, *45* (42), 12756–12766.
198. M. A. Fischbach; C. T. Walsh; J. Clardy, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (12), 4601–4608.
199. C. A. Bewley; D. J. Faulkner, *Angew. Chem. Int. Ed.* **1998**, *37* (16), 2162–2178.
200. M. G. Haygood; S. K. Davidson; D. J. Faulkner, *J. Microbiol Biotechnol.* **1999**, *1* (1), 33–43.
201. S. M. Hashimi; M. K. Wall; A. B. Smith; A. Maxwell; R. G. Birch, *Antimicrob. Agents Chemother.* **2007**, *51* (1), 181–187.
202. P. Champoiseau; J. H. Daugrois; J. C. Girard; M. Royer; P. C. Rott, *Phytopathology* **2006**, *96* (1), 33.
203. M. Royer; L. Costet; E. Vivien; M. Bes; A. Cousin; A. Damais; I. Pieretti; A. Savin; S. Megessier; M. Viard; R. Frutos; D. W. Gabriel; P. C. Rott, *Mol. Plant Microbe Interact.* **2004**, *17* (4), 414–427.
204. G. Huang; L. Zhang; R. G. Birch, *Microbiology* **2001**, *147* (Pt 3), 631–642.
206. H. Irschik; R. Jansen; K. Gerth; G. Holfe; H. Reichenbach, *J. Antibiot. (Tokyo)* **1995**, *48* (9), 962–966.
207. O. Perlova; K. Gerth; O. Kaiser; A. Hans; R. Müller, *J. Biotechnol.* **2006**, *121* (2), 174–191.
208. R. Müller; K. Gerth, *J. Biotechnol.* **2006**, *121* (2), 192–200.
209. C. Kegler; K. Gerth; R. Müller, *J. Biotechnol.* **2006**, *121* (2), 201–212.
210. S. Schneiker; O. Perlova; O. Kaiser; K. Gerth; A. Alici; M. O. Altmeyer; D. Bartels; T. Bekel; S. Beyer; E. Bode; H. B. Bode; C. J. Bolten; J. V. Choudhuri; S. Doss; Y. A. Elnakady; B. Frank; L. Gaigalat; A. Goesmann; C. Groeger; F. Gross; L. Jelsbak; L. Jelsbak; J. Kalinowski; C. Kegler; T. Knauber; S. Konietzny; M. Kopp; L. Krause; D. Krug; B. Linke; T. Mahmud; R. Martinez-Arias; A. C. McHardy; M. Merai; F. Meyer; S. Mormann; J. Muñoz-Dorado; J. Perez; S. Pradella; S. Rachid; G. Raddatz; F. Rosenau; C. Rückert; F. Sasse; M. Scharfe; S. C. Schuster; G. Suen; A. Treuner-Lange; G. J. Velicer; F.-J. Vorhölter; K. J. Weissman; R. D. Welch; S. C. Wenzel; D. E. Whitworth; S. Wilhelm; C. Wittmann; H. Blöcker; A. Pühler; R. Müller, *Nat. Biotechnol.* **2007**, *25* (11), 1281.
211. H. Irschik; R. Jansen; K. Gerth; G. Hofle; H. Reichenback, *J. Antibiot. (Tokyo)* **1995**, *48* (1), 31.
212. Y. A. Elnakady; F. Sasse; H. Lünsdorf; H. Reichenbach, *Biochem. Pharmacol.* **2004**, *67* (5), 927–935.
213. M. Kopp; H. Irschik; S. Pradella; R. Müller, *Chembiochem* **2005**, *6* (7), 1277–1286.
214. M. Kara; K. Asano; I. Kawamoto; T. Takiouchi; S. Katsumate; K. Takahashi; H. Nakano, *J. Antibiot. (Tokyo)* **1989**, *42* (12), 1768.
215. K. S. Gates, *Chem. Res Toxicol.* **2000**, *13* (10), 953–956.
216. Y. Q. Cheng; G. L. Tang; B. Shen, *J. Bacteriol.* **2002**, *184* (24), 7013.
217. G. L. Tang; Y. Q. Cheng; B. Shen, *J. Biol. Chem.* **2007**, *282* (28), 20273–20282.
218. K. H. Lee; S. Nishimura; S. Matsunaga; N. Fusetani; S. Horinouchi; M. Yoshida, *Cancer Sci.* **2005**, *96* (6), 357–364.
219. N. B. Perry; J. W. Blunt; M. H. G. Munro; L. K. Pannell, *J. Am. Chem. Soc.* **1988**, *110* (14), 4850.
220. S. Sakemi; T. Ichiba; S. Kohmoto; G. Saucy; T. Higa, *J. Am. Chem. Soc.* **1988**, *110* (14), 4851–4853.
221. J. Piel, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (22), 14002.
222. J. Piel; D. Hui; G. Wen; D. Butzke; M. PLatzer; N. Fusetani; S. Matsunaga, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (46), 16222.
223. J. Piel; D. Butzke; N. Fusetani; D. Hui; M. Platzer; G. Wen; S. Matsunaga, *J. Nat. Prod.* **2005**, *68* (3), 472–479.
224. J. Piel; D. Hui; G. Wen; D. Butzke; M. Platzer; N. Fusetani; S. Matsunaga, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (46), 16222–16227.
225. J. Piel; I. Hofer; D. Hui, *J. Bacteriol.* **2004**, *186* (5), 1280–1286.
226. Z. F. Li; J. Y. Zhao; Z. J. Xia; J. Shi; H. Liu; Z. H. Wu; W. Hu; W. F. Liu; Y. Z. Li, *Syst. Appl. Microbiol.* **2007**, *30* (3), 189–196.
227. M. Takahashi; S. Iwasaki; H. Kobayashi; S. Okuda; T. Murai; Y. Sato, *Biochim. Biophys. Acta Biomembr.* **1987**, *926* (3), 215.
228. L. P. Partida-Martinez, *Nature* **2005**, *437* (7060), 884.
229. K. Scherlach; L. P. Partida-Martinez; H.-M. Dahse; C. Hertweck, *J. Am. Chem. Soc.* **2006**, *128* (35), 11529–11536.
230. L. P. Partida-Martinez; I. Groth; I. Schmitt; W. Richter; M. Roth; C. Hertweck, *Int. J. Syst. Evol. Microbiol.* **2007**, *57* (Pt 11), 2583–2590.
231. L. P. Partida-Martinez; C. F. de Loob; K. Ishida; M. Ishida; M. Roth; K. Buder; C. Hertwick, *Appl. Environ. Microbiol.* **2006**, *73* (3), 793.
232. D. Bissett; M. A. Graham; A. Setanoians; G. A. Chadwick; P. Wilson; I. Koier; R. Henrar; G. Schwartsmann; J. Cassidy; S. B. Kaye, *Cancer Res.* **1992**, *52* (10), 2894–2898.
233. H. L. McLeod; L. S. Murray; J. Wanders; A. Setanoians; M. A. Graham; N. Pavlidis; B. Heinrich; W. W. ten Bokkel Huinink; D. J. Wagener; S. Aamdal; J. Verweij, *Br. J. Cancer* **1996**, *74* (12), 1944–1948.
234. A. W. A. C. Tolcher; J. Rizzo; E. Izbicka; E. Campbell; J. Kuhn; G. Weiss; D. D. Von Hoff; E. K. Rowinsky, *Ann. Oncol.* **2000**, *11* (3), 333–338.
235. N. Brendel; L. P. Partida-Martinez; K. Scherlach; C. Hertweck, *Org. Biomol. Chem.* **2007**, *5* (14), 2211–2213.
236. V. Simunovic; J. Zapp; S. Rachid; D. Krug; P. Meiser; R. Muller, *Chembiochem* **2006**, *7* (8), 1206.
237. M. Peter; R. Muller; *Chembiochem* **2008**, *9* (10), 1549–1553.
238. V. Simunovic; F. C. Gherardini; L. J. Shimkets, *J. Bacteriol.* **2003**, *185* (17), 5066–5075.
239. C. T. Calderone; D. F. Iwig; P. C. Dorrestein; N. L. Kelleher; C. T. Walsh, *Chem. Biol.* **2007**, *14* (7), 835–846.
240. P. S. Patel; S. Huang; S. Fisher; D. Pirnik; C. Aklonis; L. Dean; E. Meyers; P. Fernandes; F. Mayerl, *J. Antibiot. (Tokyo)* **1995**, *48* (9), 997–1003.
241. A. M. Albertini; T. Caramori; F. Scoffone; C. Scotti; A. Galizzi, *Microbiology* **1995**, *141* (2), 299–309.

242. R. A. Butcher; F. C. Schroeder; M. A. Fischback; P. D. Straight; R. Kolter; C. T. Walsh; J. Clardy, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (5), 1506.
243. P. D. Straight; M. A. Fishback; C. T. Walsh; D. Z. Rudner; R. Kolter, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (1), 305.
244. C. Scotti; M. Piatti; A. Cuzzoni; P. Perani; A. Tognoni; G. Grandi; A. Galizzi; A. M. Albertini, *Gene* **1993**, *130* (1), 65–71.
245. J. Moldenhauer; X.-H. Chen; R. Borriss; J. Piel, *Angew. Chem. Int. Ed.* **2007**, *46* (43), 8195–8197.
246. S. Harada; T. Yamazaki; K. Hatano; K. Tsuchiya; T. Kishi, *J. Antibiot. (Tokyo)* **1973**, *26* (11), 647–657.
247. S. Mochizuki; K. Hiratsu; M. Suwa; T. Ishii; F. Sugino; K. Yamada; H. Kinashii, *Mol. Microbiol.* **2003**, *48* (6), 1501.
248. K. Arakawa; F. Sugino; K. Kodama; T. Ishii; H. Kinashi, *Chem. Biol.* **2005**, *12* (2), 249–256.
249. K. Arakawa; S. Mochizuki; K. Yamada; T. Noma; H. Kinashi, *Microbiology* **2007**, *153* (Pt 6), 1817–1827.
250. S. Yamamoto; Y. He; K. Arakawa; H. Kinashi, *J. Bacteriol.* **2008**, *190* (4), 1308–1316.
251. R. Maget-Dana; M. Ptak, *Biochim. Biophys. Acta* **1990**, *1023* (1), 34–40.
252. E. H. Duitman, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (23), 13294.
253. S. A. Sieber; J. Tao; C. T. Walsh; M. A. Marahiel, *Angew. Chem. Int. Ed.* **2004**, *43* (4), 493–498.
254. W. Namwat; Y. Kamioka; H. Kinoshita; Y. Yamada; T. Nihira, *Gene* **2002**, *286* (2), 283–290.
255. C. Cocito, *Microbiol. Mol. Biol. Rev.* **1979**, *43* (2), 145.
256. J. Acar; M. Casewell; J. Freeman; C. Friis; H. Goossens, *Clin. Microbiol. Infect.* **2000**, *6* (9), 477.
257. L. A. Cox, *Environ. Int.* **2005**, *31* (4), 549–563.
258. N. Pulsawat; S. Kitani; T. Nihira, *Gene* **2007**, *393* (1–2), 31–42.
259. J. Clardy, *Gen. Biol.* **2005**, *6* (9), 232.
260. B. Wilkinson; J. Micklefield, *Nat. Chem. Biol.* **2007**, *3* (7), 379–386.
261. E. Vivien; D. Pitorre; S. Cociancich; I. Pieretti; D. W. Gabriel; P. C. Rott; M. Royer, *Antimicrob. Agents Chemother.* **2007**, *51* (4), 1549–1552.
262. J. S. Guez; S. Chenikher; J. P. Cassar; P. Jacques, *J. Biotechnol.* **2007**, *131* (1), 67–75.
263. E. H. Duitman; D. Wyczawski; L. G. Boven; G. Venema; O. P. Kuipers; L. W. Hameon, *Appl. Environ. Microbiol.* **2007**, *73* (11), 3490.
264. V. Leclere; M. Bechet; A. Adam; J. S. Guez; B. Wathelet; M. Ongena; P. Thonart; F. Gancel; M. Chollet-Imbart; P. Jacques, *Appl. Environ. Microbiol.* **2005**, *71* (8), 4577.
265. P. Fickers; V. Leclère; J.-S. Guez; M. Béchet; F. Coucheney; B. Joris; P. Jacques, *Res. Microbiol.* **2008**, *159* (6), 449.
266. N. S. Webster; K. J. Wilson; L. L. Blackall; R. T. Hill, *Appl. Environ. Microbiol.* **2001**, *67* (1), 434–444.
267. A. Schirmer; R. Gadjarum; C. D. Reeves; F. Ibrahim; E. F. DeLong; C. R. Hutchinson, *Appl. Environ. Microbiol.* **2005**, *71* (8), 4840.
268. S. Hrvatin; J. Piel, *J. Microbiol. Methods* **2007**, *68* (2), 434.
269. D. W. Udwary; L. Zeigler; R. N. Asolkar; V. Singan; A. Lapidus; W. Fenical; P. R. Jensen; B. S. Moore, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (25), 10376.
270. J. E. Loper; M. D. Henkels; B. T. Shaffer; F. A. Valeriote; H. Gross, *Appl. Environ. Microbiol.* **2008**, *74* (10), 3085–3093.
271. J. Piel; D. Hui; N. Fusetani; S. Matsunaga, *Environ. Microbiol.* **2004**, *6* (9), 921–927.
272. P. C. Dorrestein; N. L. Kelleher, *Nat. Prod. Rep.* **2006**, *23* (6), 893–918.
273. J. Yin; P. D. Straight; S. Hrvatin; P. C. Dorrestein; S. B. Bumpus; C. Jao; N. L. Kelleher; R. Kolter; C. T. Walsh, *Chem. Biol.* **2007**, *14* (3), 303–312.
274. E. Esquenazi, *Mol. Biosystems* **2008**, *4* (6), 562.
275. W. C. Dunlap; M. Jaspars; D. Hranueli; C. N. Battershill; N. Perić-Concha; J. Zucko; S. H. Wright; P. F. Long, *Curr. Med. Chem.* **2006**, *13* (6), 697–710.
276. S. Smith, *Chem. Biol.* **2002**, *9* (9), 955.
277. T. Maier; S. Jenni; N. Ban, *Science* **2006**, *311* (5765), 1258–1262.
278. S. Jenni; M. Leibundgut; D. Boehringer; C. Frick; B. Mikolasek; N. Ban, *Science* **2007**, *316* (5822), 254–261.
279. M. Leibundgut; S. Jenni; C. Frick; N. Ban, *Science* **2007**, *316* (5822), 288–290.
280. S. Smith; S. C. Tsai, *Nat. Prod. Rep.* **2007**, *24* (5), 1041.
281. T. Maier; M. Leibundgut; N. Ban, *Science* **2008**, *321* (5894), 1315.
282. C. Khosla; Y. Tang; A. Y. Chen; N. A. Schnarr; D. E. Cane, *Annu. Rev. Biochem.* **2007**, *76*, 195–221.
283. V. Y. Alekseyev; C. W. Liu; D. E. Cane; J. D. Puglisi; C. Khosla, *Protein Sci.* **2007**, *16* (10), 2093–2107.
284. A. C. Mercer; M. D. Burkart, *Nat. Prod. Rep.* **2007**, *24* (4), 750–773.
285. Z. Zhou; J. R. Lai; C. T. Walsh, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (28), 11621.

Biographical Sketches

Christopher M. Rath is a graduate student in the Chemical Biology Ph.D. program at the University of Michigan. His research is focused on understanding polyketide natural product biosynthesis using modern analytical methodologies such as Fourier transform ion cyclotron resonance mass spectrometry. His Ph.D. research project is a collaboration between the Sherman and Hakansson laboratories. Chris obtained his B.S. at the University of California at Santa Cruz, where he also studied natural products isolated from sponge-derived fungi with Professor Phil Crews. He then worked in this industry for 3 years at PDL Biopharma, characterizing biotherapeutics, before beginning his graduate work at Michigan.

Jamie B. Scaglione was born in DuQuoin, Illinois, in 1979. She received a B.S. with honors in biology and chemistry from the University of Evansville in 2001. After completing a Ph.D. in chemical biology at Washington University School of Medicine in 2007 under the direction of Professor Douglas F. Covey, she joined Professor David H. Sherman's group as a postdoctoral research fellow in the Life Sciences Institute at the University of Michigan.

Jeffrey D. Kittendorf received his B.S. degree in biochemistry at Eastern Michigan University and completed his Ph.D. in medicinal chemistry at the University of Michigan where he studied tRNA-modifying enzymes in the laboratory of Professor George A. Garcia. In 2004, he joined the laboratory of Professor David H. Sherman in the Life Sciences Institute at the University of Michigan, where he was an NRSA postdoctoral fellow. His postdoctoral research was focused on understanding mechanistic details of macrolactone production by the pikromycin polyketide synthase. Currently, Dr. Kittendorf is a research investigator in Professor Sherman's laboratory and also a founding scientist of Alluvium Biosciences. His research interests include the biosynthesis of polyketides and nonribosomal peptides, with particular emphasis on the mechanistic enzymology and engineering of modular biosynthetic systems.

David H. Sherman, Hans W. Vahlteich Professor of Medicinal Chemistry in the College of Pharmacy at the University of Michigan, also holds appointments in the Department of Chemistry and in the University of Michigan Medical School Department of Microbiology & Immunology. Dr. Sherman currently focuses on the isolation and analysis of new natural products from diverse source material including terrestrial and marine actinomycetes, cyanobacteria, and fungi. Dr. David Sherman has distinguished himself in many scientific arenas including natural product chemistry and microbiology in the pursuit of promising anticancer and anti-infective drugs. He combines his scientific pursuits and fascination with biodiversity of fragile tropical reef habitats by conducting field studies in the waters near Panama, the Caribbean, Papua New Guinea, and the Red Sea. In addition, Dr. Sherman is the Director of the Center for Chemical Genomics at the University of Michigan Life Science Institute (LSI). LSI maintains core facilities covering the areas of chemical genomics, structural biology, and protein production with resources to support cross-disciplinary science including genetics; genomics and proteomics; molecular and cellular biology; and structural, chemical, and computational biology.

1.12 Mevalonate Pathway in Bacteria and Archaea

Tomohisa Kuzuyama, The University of Tokyo, Tokyo, Japan

Hisashi Hemmi, Nagoya University, Nagoya, Japan

Shunji Takahashi, RIKEN, Advanced Science Institute, Saitama, Japan

© 2010 Elsevier Ltd. All rights reserved.

1.12.1	**Introduction**	493
1.12.2	**Overview of the Mevalonate Pathway**	494
1.12.2.1	Mevalonate Pathway	494
1.12.2.2	Gene Cluster of the Mevalonate Pathway	494
1.12.3	**Two Classes of HMG-CoA Reductase**	497
1.12.3.1	Class I HMG-CoA Reductase	497
1.12.3.1.1	Archaeal class I HMGR	497
1.12.3.1.2	Bacterial class I HMGR	498
1.12.3.1.3	Regulation of class I HMGR catalytic activity	500
1.12.3.2	Class II HMG-CoA Reductase	500
1.12.3.2.1	Bacterial class II HMGR	500
1.12.3.2.2	Archaeal class II HMGR	501
1.12.3.2.3	Class II HMGR as a new molecular target for drug development	501
1.12.3.3	Crystallographic Studies of Two Classes of HMGR	502
1.12.4	**Two Types of IPP Isomerase**	502
1.12.4.1	Type 1 IPP Isomerase	503
1.12.4.2	Type 2 IPP Isomerase	506
1.12.5	**Biosynthetic Gene Clusters of Secondary Metabolites from Actinomycetes**	510
1.12.6	**Mevalonate Fermentation**	511
1.12.7	**Conclusions**	512
References		513

1.12.1 Introduction

This chapter focuses on the mevalonate pathway of isoprenoid biosynthesis in bacteria and archaea. As described in Section 1.12.2, the mevalonate pathway is a metabolic pathway that provides isopentenyl diphosphate (IPP) **1** and its isomer dimethylallyl diphosphate (DMAPP) **2**. Both compounds are common precursors for the biosynthesis of all isoprenoids. Isoprenoids are ubiquitous in living organisms and are diverse in structure and biological function.[1] For example, they function as steroid hormones in mammals, carotenoids in plants, ubiquinone or menaquinone in bacteria, and membrane lipids in archaea. In addition, many isoprenoids have been isolated as secondary metabolites with diverse structures and activities, predominantly from higher plants. Those isoprenoids are described in detail in other chapters.

In addition to the mevalonate pathway, an alternative biosynthetic pathway has been recently described for the biosynthesis of IPP and DMAPP. This alternative pathway is generally called the 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway, as elaborated in Chapter 12. Most bacteria utilize the MEP pathway of isoprenoid biosynthesis, whereas all eukaryotes, archaea, and some bacteria such as Gram-positive bacteria including *Staphylococcus aureus*, *Enterococcus faecalis*, and some Actinobacteria, utilize the mevalonate pathway.

3-Hydroxy-3-methylglutaryl (HMG)-coenzyme A (CoA) reductases, the rate-limiting enzymes of the mevalonate pathway, are differentiated into two classes based on the amino acid sequences: class I for the eukaryotic type of enzymes and class II for the others. Among bacteria using the mevalonate pathway, most

bacteria have a class II HMG-CoA reductase, but some high GC Gram-positive members of the Actinobacteria, such as *Streptomyces* strains, have a class I enzyme. Both classes of enzyme are found in archaea. Properties and distributions of the two classes of HMGR are described in Section 1.12.3.

In the last step in the mevalonate pathway, IPP isomerase catalyzes isomerization of IPP **1** to give its isomer, DMAPP **2**. Based on the amino acid sequences and the requirements for coenzymes, IPP isomerases are also classified into two groups: type 1 and type 2. The distribution, properties, and reaction mechanism of IPP isomerases are elaborated in Section 1.12.4.

1.12.2 Overview of the Mevalonate Pathway

1.12.2.1 Mevalonate Pathway

In this chapter, the mevalonate pathway is considered to consist of seven enzymes: acetoacetyl-CoA thiolase, HMG-CoA synthase, HMGR, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and IPP isomerase. The biosynthetic pathway of IPP and DMAPP biosynthesis is summarized in **Scheme 1**. Eukaryotes and bacteria share a common pathway, but the pathway in archaea differs slightly from that of eukaryotes and bacteria in the final steps.

The mevalonate pathway begins with a thioester-dependent Claisen condensation reaction between two molecules of acetyl-CoA **3**, catalyzed by acetoacetyl-CoA thiolase (EC 2.3.1.9), to give acetoacetyl-CoA **4**. A second condensation reaction with a third molecule of acetyl-CoA then yields the six-carbon compound, (3*S*)-HMG-CoA **5**. This condensation reaction, which is an aldol-like process, is catalyzed by HMG-CoA synthase (EC 2.3.3.10). HMG-CoA **5** is then reduced to mevalonate **6** by HMGR (EC 1.1.1.34), the rate-limiting enzyme of the mevalonate pathway. This enzyme is classified into either class I or class II, based on its amino acid sequence (see Section 1.12.3).

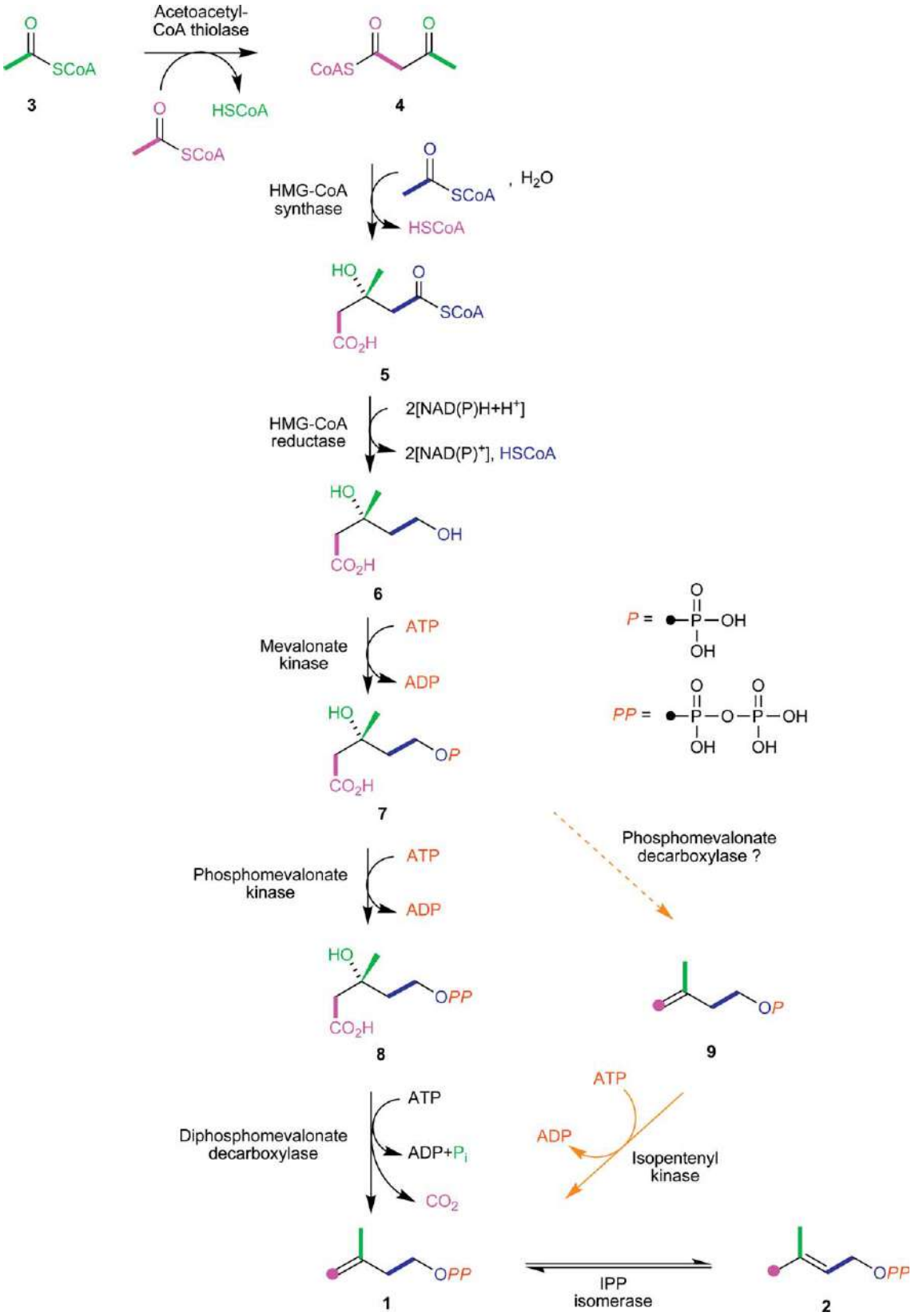
Mevalonate **6** is converted into diphosphomevalonate **8** via phosphomevalonate **7** by two consecutive phosphorylation reactions involving adenosine 5′-triphosphate (ATP) hydrolysis catalyzed by mevalonate kinase (EC 2.7.1.36) and phosphomevalonate kinase (EC 2.7.4.2). Diphosphomevalonate **8** is then converted into IPP **1** by diphosphomevalonate decarboxylase (EC 4.1.1.33). In this step, the release of CO_2 from diphosphomevalonate **8** occurs together with the hydrolysis of ATP to adenosine 5′-diphosphate (ADP) and inorganic phosphate. Finally, IPP **1** is converted into its isomer DMAPP **2** by the action of IPP isomerase. As mentioned, there are two types of IPP isomerases, and they have no amino acid sequence similarity (see Section 1.12.4).

Archaea like *Methanocaldococcus jannaschii* have a slightly different pathway from those of eukaryotes and bacteria in the late steps of the mevalonate pathway (**Scheme 1**). In 'the archaeal mevalonate pathway,' IPP **1** is synthesized from isopentenyl phosphate (IP) **9** by the action of IP kinase.[2] This kinase is also found in several other archaea including *Methanobacterium thermautotrophicum*, *Methanosarcina mazei*, and *Aeropyrum pernix*. However, the enzyme responsible for the synthesis of IP remains unknown. White and coworkers proposed that an unidentified phosphomevalonate decarboxylase may catalyze the decarboxylation of mevalonate phosphate to give IP (**Scheme 1**).[2]

1.12.2.2 Gene Cluster of the Mevalonate Pathway

Figure 1 shows the organization of the genes responsible for the mevalonate pathway of eight bacteria and the archaeon, *M. jannaschii*.

Interestingly, in high-GC Gram-positive members of the actinobacteria, including two *Streptomyces* species,[3,4] *Actinoplanes* species,[5] and *Nocardia farcinica*,[6] the six mevalonate pathway genes (except the acetoacetyl-CoA thiolase gene) are clustered in a single locus on their genomes, and the order and extent of the genes are identical. Also, in *Borrelia garinii*,[7] a Gram-negative spirochete bacterium that is a causative agent of Lyme borreliosis, and *Paracoccus zeaxanthinifaciens*,[8] a Gram-negative marine bacterium that produces the yellow carotenoid zeaxanthin, the six mevalonate pathway genes are clustered in a single locus. However, the structures of these clusters are different from those of high-GC Gram-positive



Scheme 1 Mevalonate pathway in bacteria and archaea. Orange arrows represent the pathway characteristic of archaea. Phosphomevalonate decarboxylase has not yet been identified.

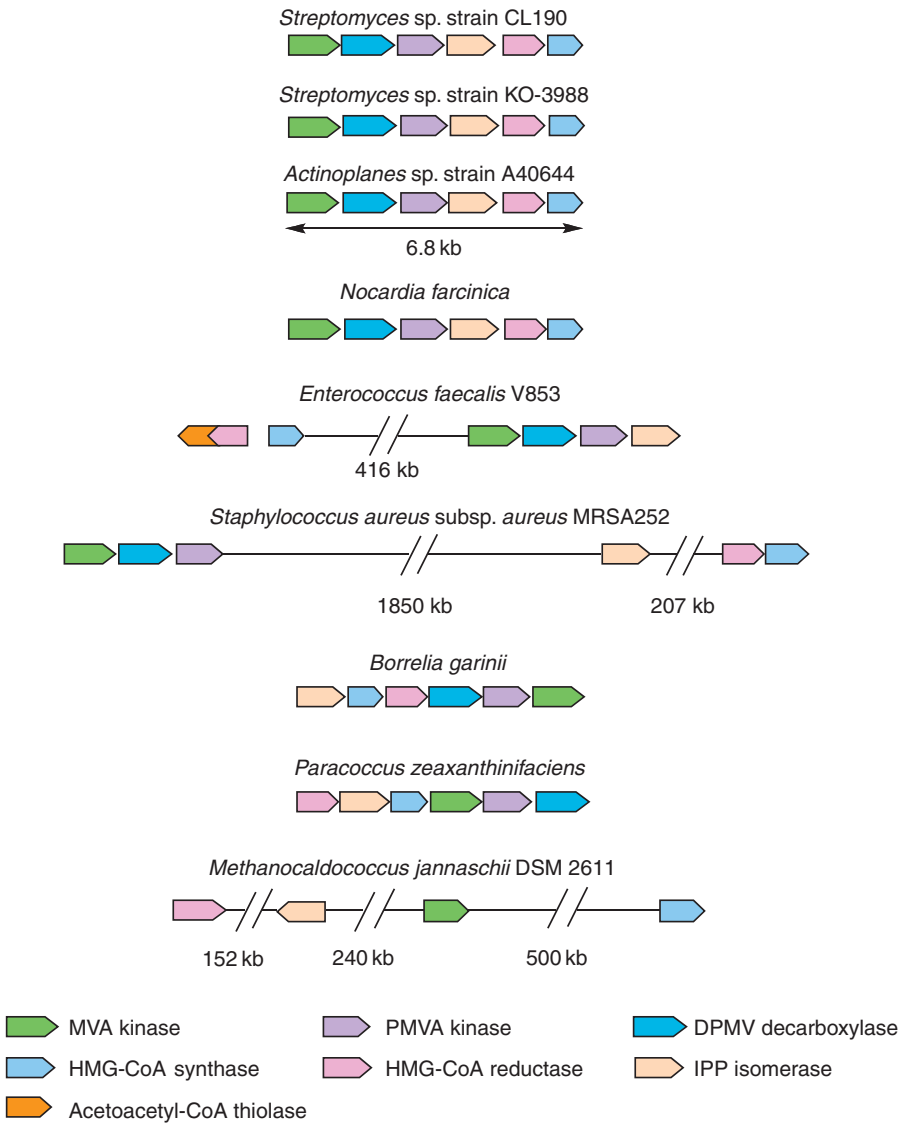


Figure 1 Mevalonate pathway gene clusters.

members of the Actinobacteria. *E. faecalis* has two loci with two kinase genes, the diphosphomevalonate decarboxylase gene and the IPP isomerase gene grouped in one locus, and HMG-CoA synthase and HMGR reductase genes on the second locus.[9,10] In addition, the *E. faecalis* HMG-CoA reductase gene is fused with the acetoacetyl-CoA thiolase gene to encode a dual-function protein (MvaE) that catalyzes both the acetoacetyl-CoA thiolase and the HMGR reactions.[9] In *S. aureus*, the six mevalonate pathway genes are encoded in three loci.[11] In the archaea *M. jannaschii,* the four mevalonate pathway genes are scattered on the genome and the phosphomevalonate kinase and the diphosphomevalonate decarboxylase genes are missing.[2] As described, *M. jannaschii* utilizes a slightly different pathway from those of eukaryotes and bacteria in the late steps of the mevalonate pathway (**Scheme 1**). In the archaeal pathway, IPP **1** is synthesized from IP **9** by the action of IP kinase, whose gene is located 586 kb upstream of the HMGR gene. However, a gene for the IP synthesizing enzyme has not yet been identified.

1.12.3 Two Classes of HMG-CoA Reductase

HMGR (EC 1.1.1.34) is the rate-limiting enzyme of the mevalonate pathway. This enzyme is the target of statin drugs such as lovastatin (mevinolin) **10**, mevastatin (compactin) **11**, and simvastatin **12**, which lower blood cholesterol levels (**Figure 2**).[12,13] Recent progress in genome analysis unveiled an extensive amount of information regarding the HMGR sequence. Comparison of amino acid sequences and phylogenetic analysis of representative HMGRs from eukaryotes, archaea, and bacteria revealed two distinct classes.[14–16] **Figure 3** shows the evolutionary divergence in the sequences between the two classes.

1.12.3.1 Class I HMG-CoA Reductase

Class I HMGRs are found in eukaryotes, archaea, and high-GC Gram-positive members of the Actinobacteria. Class I HMGRs that have been characterized include human, rat, hamster, yeast, plants, *Haloferax volcanii*,[18] *Sulfolobus solfataricus*,[19,20] and some *Streptomyces* sp.[21,22] Class I HMGRs utilize two NADPH equivalents to reductively deacylate the thioester group of HMG-CoA to the primary alcohol of mevalonate:

$$(S)\text{-HMG-CoA}\ \ \mathbf{5} + 2\text{NADPH} + 2\text{H}^+ \rightarrow (R)\text{-mevalonate}\ \ \mathbf{6} + \text{CoASH} + 2\text{NADP}^+$$

The reaction proceeds in three stages. The first and third stages are reductive, and involve the formation of enzyme-bound mevaldyl-CoA **13** and mevaldehyde **14** as probable intermediates (**Figure 4**).

Stage 1: HMG-CoA **5** + NADPH + $\text{H}^+ \rightarrow$ [Mevaldyl-CoA] **13** + NADP^+
Stage 2: [Mevaldyl-CoA] **13** $\rightarrow$ [Mevaldehyde] **14** + CoASH
Stage 3: [Mevaldehyde] **14** + NADPH+$\text{H}^+ \rightarrow$ Mevalonate **6** + NADP^+

1.12.3.1.1 Archaeal class I HMGR

In all archaea, HMGR is an essential enzyme for the synthesis of isoprenoid glycerol ethers, the main component of cell membranes.[23,24] Early studies suggested that *Halobacterium halobium* synthesizes mevalonate through NADPH-dependent HMGR, which is sensitive to inhibition by lovastatin **10** (**Figure 2**).[25] A study of lovastatin-resistant mutations using the halophilic archaeon *H. volcanii* identified the first archaeal HMGR gene.[26] The gene responsible for lovastatin resistance contained an up-promoter mutation, which resulted in an increase of the HMGR gene product. The HMGR gene of *H. volcanii* has been expressed in *Escherichia coli* and its encoded protein (40% amino acid identity to mammalian HMGR) has been purified to homogeneity and characterized.[18] Except for its optimal enzyme activity under high ionic strength conditions, all the catalytic properties of *H. volcanii* HMGR were comparable to those of HMGR from higher eukaryotes. The archaeal HMGR showed competitive inhibition to lovastatin with a K_i of 15 nmol l^{-1}. Consistent with cell-free extract assay results,[25] *H. volcanii* HMGR used NADPH as a coenzyme and exhibited stereospecificity for (*S*)-HMG-CoA.[18] The HMGR gene was also cloned from the thermophilic archaeon *S. solfataricus* P2.[19] *S. solfataricus*

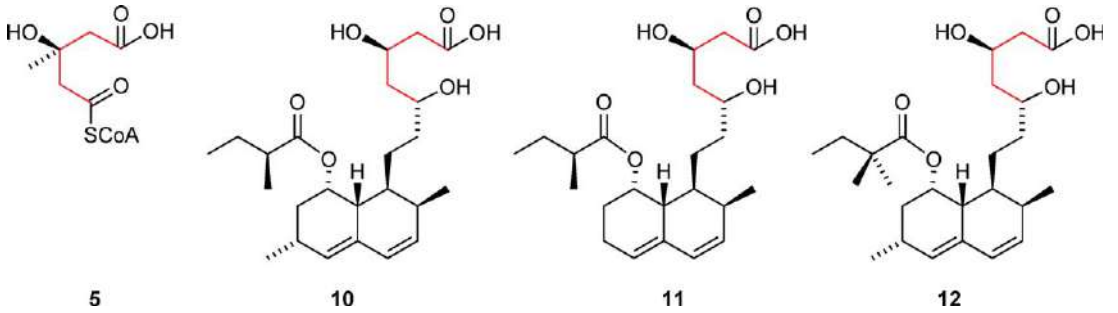


Figure 2 HMG-CoA and statin inhibitors, lovastatin, mevastatin, and simvastatin. The acid forms of each statin are illustrated.

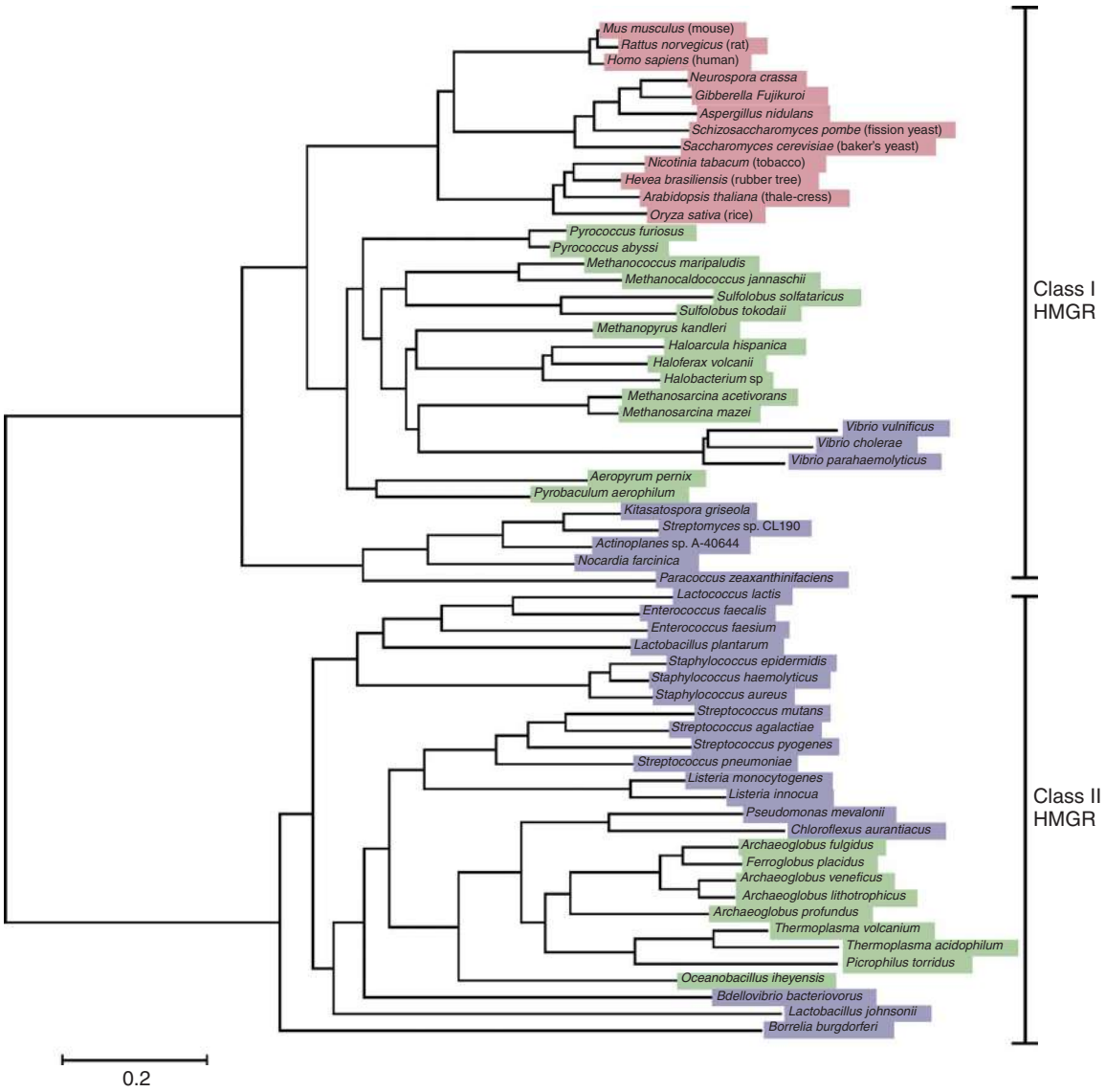


Figure 3 Phylogenic tree of class I and class II HMGRs. The tree includes 61 selected organisms that possess the HMGR gene. Phylogenic analysis was performed using aligned amino acid sequences of HMGR catalytic domains. The phylogenic tree was constructed with a program Clustal W.[17] Red, eukaryotes; violet, bacteria; green, archaea.

HMGR, which exhibits 47% sequence identity to *H. volcanii* HMGR, was purified to homogeneity and its enzymatic properties were characterized.[19,20] Although *S. solfataricus* HMGR showed optimal activity at 85 °C and pH 5.5,[19] the enzyme exhibited substrate specificities typical for eukaryotic HMGR.

1.12.3.1.2 Bacterial class I HMGR

Bacterial class I HMGR was first characterized from *Streptomyces* sp. strain CL190,[21] a high-GC Gram-positive member of the Actinobacteria. This biosynthetic enzyme utilizes NADPH and the inhibition profile for statin resembles those of eukaryotic enzymes. Interestingly, the existence of the mevalonate pathway in *Streptomyces* is unusual and is associated with the production of isoprenoids categorized as secondary metabolites.[27,28] Research on isoprenoid biosynthesis in *Streptomyces* suggested that the MEP pathway is used by all *Streptomyces* species for the production of primary metabolites such as menaquinones, and that a few

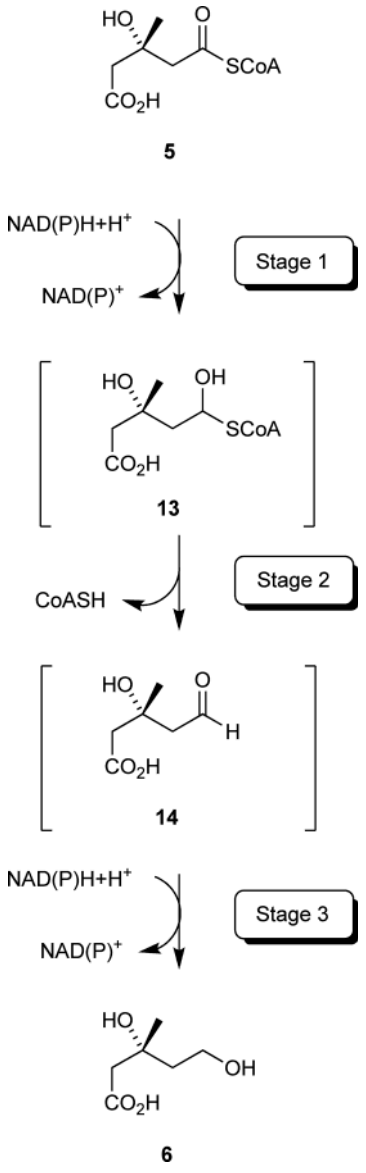


Figure 4 Substrates and products of the reaction catalyzed by HMGR. The putative HMGR-bound intermediates, mevaldyl-CoA **13** and mevaldehyde **14**, are also illustrated in brackets.

Streptomyces strains also utilize the mevalonate pathway for the production of secondary metabolites with isoprenoid moieties. Southern hybridization experiments using the HMGR gene as a probe demonstrated that the presence of the HMGR gene is associated with the production of isoprenoid secondary metabolites.[27,28] Currently, the discovery of novel compounds by random screening of microorganisms is time-consuming. Thus, utilization of the HMGR gene as a probe may be a promising approach for isolating new producers of structurally diverse, biologically active isoprenoids.

Curiously, class I HMGR genes are also found in pathogenic bacteria including *Vibrio cholerae*, *Vibrio vulnificus*, and *Vibrio parahaemolyticus*. No other genes for the mevalonate pathway other than the HMGR gene are found on their genomes. Instead, these bacteria carry all the genes required for the MEP pathway. Thus, these bacteria must exclusively utilize the MEP pathway for isoprenoid biosynthesis. The question why these pathogenic bacteria carry the HMGR gene remains unknown.

1.12.3.1.3 Regulation of class I HMGR catalytic activity

Eukaryotic HMGRs are regulated at several steps, including transcription, translation, and enzyme degradation.[29] Eukaryotic HMGRs have a highly conserved C-terminal domain and an N-terminal membrane-anchor domain composed of two to eight transmembrane helices. The membrane domain of eukaryotic HMGRs is closely related to enzyme degradation. There have been no reports of regulation by protein degradation in archaeal and bacterial HMGRs, which lack an N-terminal membrane domain.

In eukaryotes, phosphorylation attenuates and dephosphorylation restores the catalytic activity of HMGR.[30,31] The phosphorylation site of serine is located precisely six residues on the C-terminal side of the active site histidine.[32] Because of the ubiquitous presence of the phosphoacceptor serine and protein kinase motif, the regulation of HMGR by phosphorylation is general for all higher eukaryotes. The substitution of aspartate for the regulatory serine decreases its activity by 90%, suggesting that the negative charge introduced by phosphorylation may play a major role in regulating catalytic activity.[32] The mechanism by which the negative charge on this aspartate regulates catalysis is believed to be due to the carboxylate anion of the aspartate competing for the proton of the imidazolium ion on the active site histidine, making the histidinium proton unavailable for protonating the inhibitory $CoAS^-$ thioanion.[32] In contrast to higher eukaryotic HMGRs, there is no phosphoacceptor serine residue in the aligned sequences of archaeal and bacterial HMGRs. Thus, no phosphorylation regulation has been reported *in vivo*. Interestingly, protein engineering of *S. solfataricus* HMGR with the cyclic adenosine monophosphate (cAMP)-dependent protein kinase recognition motif enabled the regulation of enzyme activity by phosphorylation and dephosphorylation.[33] Although a similar domain-swapping approach was not successful for *H. volcanii* HMGR,[34] site-directed mutagenesis of the active site histidine to glutamine, and the introduction of a negatively charged aspartate at the putative C-terminal regulatory region, demonstrated that the enzymatic properties of *H. volcanii* HMGR are closely similar to those of eukaryotic HMGRs.

1.12.3.2 Class II HMG-CoA Reductase

Class II HMGR is found in bacteria and archaea. Class II HMGRs that have been characterized include *Pseudomonas mevalonii*,[35–38] *S. aureus*,[11] *E. faecalis*,[9] *Archaeoglobus fulgidus*,[39] and *Listeria monocytogenes*.[40] The inhibition of class I and class II HMGR by statins has been extensively investigated and has revealed different sensitivities to statin drugs. Although class I HMGR is strongly inhibited by statins at nanomolar concentrations, class II HMGRs are over 4 orders of magnitude less sensitive to statins.[12,13,16]

1.12.3.2.1 Bacterial class II HMGR

A bacterial class II HMGR was first obtained from *P. mevalonii*, which can grow on mevalonate as the sole carbon source.[36] *Pseudomonas mevalonii* HMGR plays a biodegradative role *in vivo*, converting (*R*)-mevalonate to (*S*)-HMG-CoA. Unlike class I HMGR, which utilizes NADPH as coenzyme exclusively, class II HMGR from *P. mevalonii* utilizes NADH as coenzyme.[37,38] *P. mevalonii* HMGR is one of the best-characterized enzymes among the HMGRs reported. Comparison of class I human HMGR[41,42] and class II *P. mevalonii* HMGR structures,[43–45] and kinetic analyses by site-directed mutagenesis have identified the catalytic role of conserved glutamate,[46] aspartate,[47] histidine,[47–49] and lysine residues (**Figure 5**).[42,44,50,51]

Although *P. mevalonii* HMGR has a biodegradative role, other bacterial class II HMGRs have biosynthetic roles for the production of isoprenoid compounds. A biosynthetic class II HMGR was first characterized from the Gram-positive pathogen *S. aureus*.[11] The *S. aureus* HMGR gene was expressed in *E. coli* and the protein was purified to homogeneity. The kinetic parameters of *S. aureus* HMGR for substrate, the K_i for statin, and the optimal pH of the enzyme were similar to that of *P. mevalonii*. Unlike *P. mevalonii* HMGR, which utilizes only NADH, *S. aureus* HMGR can utilize either NADH or NADPH as coenzyme, although NADPH is the preferred coenzyme.

Gram-positive cocci use the mevalonate pathway, which is essential for growth. Analysis of genome sequences of enterococci revealed the presence of the *mvaE* gene that encodes both acetoacetyl-CoA thiolase and class II HMGR in a single gene.[52] The MvaE protein was the first example of an HMGR fused to another enzyme. Western blotting analysis using enterococcal extracts indicated that the *mvaE* gene product is expressed in *Enterococcus faecium*, *E. faecalis*, and *Enterococcus hirae*.[9] The *mvaE* gene from *E. faecalis* was expressed

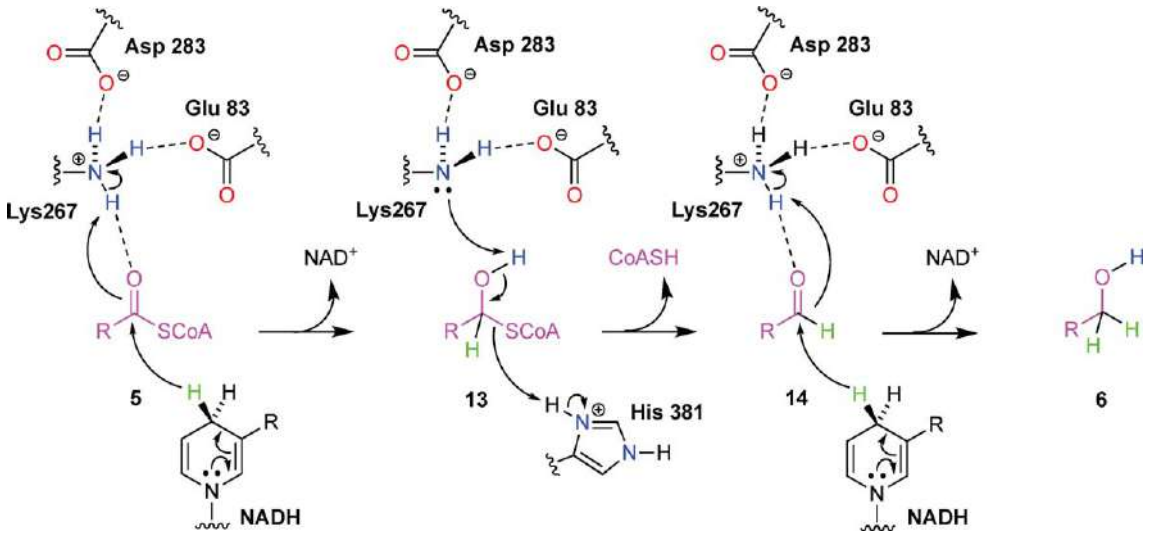


Figure 5 Reaction mechanism proposed for *P. mevalonii* HMGR. Substrates and products of the reaction catalyzed by HMGR are colored in magenta. The three stages of the reaction and the roles for the catalytic residues Glu83, Lys267, Asp283, and His381 are illustrated.

in *E. coli* and purified to homogeneity as a 90-kDa protein. The fusion protein catalyzed both the acetoacetyl-CoA thiolase and HMGR reactions.[9] The K_m values and pH profiles, and the K_i values of *E. faecalis* HMGR, were comparable to those of other class II HMGRs. In addition, the enzyme exclusively used NADPH as coenzyme.

The class II HMGR gene was also isolated from *L. monocytogenes*, a Gram-positive rod-shaped bacterium that causes listeriosis.[40] The HMGR gene was expressed in *E. coli* and purified to homogeneity as a 48-kDa protein. Although the purified enzyme utilized both NADH and NADPH for catalysis, NADPH was the preferred coenzyme. *L. monocytogenes* HMGR required micromolar concentrations of lovastatin and mevastatin for inhibition.

1.12.3.2.2 Archaeal class II HMGR

The hyperthermophilic archaeon *A. fulgidus*, which grows in hydrothermal environments, harbors class II HMGR instead of class I HMGR found in other archaea. The first example of an archaeal class II HMGR gene from *A. fulgidus* was expressed in *E. coli* and the enzymatic properties of the purified protein were characterized.[39] Except for its optimum activity at 85 °C, the kinetic parameters of *A. fulgidus* HMGR for substrates, its K_i for statin, and its optimal pH are similar to that of *P. mevalonii*. Unlike *P. mevalonii* HMGR, *A. fulgidus* HMGR exhibited dual coenzyme specificity. Although significant activity was observed using NADPH, NADH was the preferred coenzyme.

1.12.3.2.3 Class II HMGR as a new molecular target for drug development

Class II HMGRs are present in clinically important human pathogens, including *Borrelia burgdorferi*, *Streptococcus pyogenes*, *Streptococcus pneumoniae*, *S. aureus*,[11] *E. faecalis*,[9] and *L. monocytogenes*.[40] Gene disruption in the mevalonate gene cluster of *S. pneumoniae* showed that the mevalonate pathway is essential for the survival of Gram-positive cocci. The virulence of the *mvaS* (HMG-CoA synthase gene) null mutant of *S. pneumoniae* was severely attenuated.[52] In addition, disruption of the *mvaA* gene that encodes *S. aureus* HMGR demonstrated that the gene is essential for the growth of this organism, and growth of the *mvaA* null mutant was severely attenuated in the mouse hematogenous pyelonephritis infection model.[11] Since the enterococci are a major cause of nosocomial infections, the eubacterial class II HMGR is a promising molecular target for the development of antibiotics against multidrug-resistant strains.

1.12.3.3 Crystallographic Studies of Two Classes of HMGR

The structures of the catalytic portion of human HMGR complexed with six different statins have been determined by crystallography (**Figure 6**).[41,42] The study characterized the active site and the HMGR–statin interactions (**Figure 7**). The rational design of improved human HMGR inhibitors will be based on such structural data. However, no crystal structure of a class I HMGR from archaea or bacteria has been solved to date.

Statins are excellent inhibitors of the class I HMGR but relatively poor inhibitors of the class II HMGR of important bacterial pathogens mentioned above. To investigate the molecular basis for this difference, the X-ray structure of the class II *P. mevalonii* HMGR in complex with the statin drug lovastatin **10** has been determined (**Figures 6** and **7**).[45] Comparison of the lovastatin **10**-bound structures of the class II *P. mevalonii* HMGR and mevastatin **11**- or simvastatin **12**-bound structure of the class I human HMGR revealed an overall similar manner of binding for these statins but significant differences in the specific interactions around the carboxylic acid group of the statins (**Figure 7**). For example, in the human HMGR complexes, additional hydrogen bond interactions were observed between the carboxylic acid group of the statins and Ser684. It is proposed that these differences might be thus exploited to develop class II-specific inhibitors against clinically important human pathogens.[45]

1.12.4 Two Types of IPP Isomerase

IPP **1** produced through the mevalonate pathway must be converted into DMAPP **2**, because DMAPP **2** is required as a priming allylic substrate (by generating an allylic carbocation intermediate) to start the prenyl-chain elongation reactions (**Figure 8**). IPP isomerase can catalyze the interconversion reaction between IPP **1** and DMAPP **2**. Based on their amino acid sequences and coenzyme requirements, IPP isomerases are classified as either type 1 or type 2 (**Figure 8**).

DMAPP **2** generated from IPP **1** by IPP isomerase condenses with other molecules of IPP **1** to give prenyl diphosphates such as geranyl diphosphate (GDP, C_{10}) **15**, farnesyl diphosphate (FDP, C_{15}) **16**, and geranylgeranyl diphosphate (GGDP, C_{20}) **17**. These prenyl diphosphates are produced by prenyl-chain elongation reactions. The prenyl-chain length is catalytically determined by polyprenyl diphosphate synthases such as

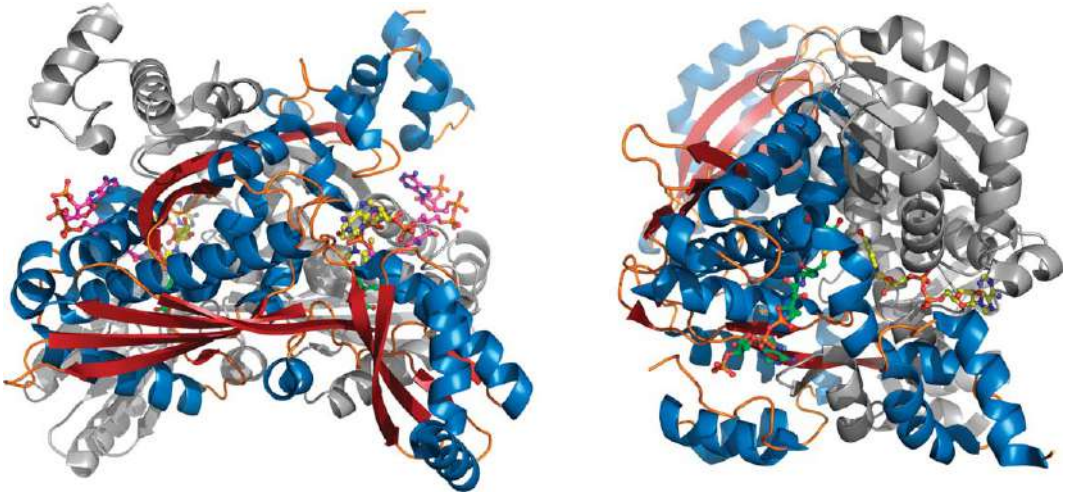


Figure 6 Crystal structures of two classes of HMGR. Ribbon diagrams of HMGRs. For clarity, only one monomer is colored. *Left*, human class I HMGR binding with HMG (C = green, O = red), CoA (C = magenta, N = blue, O = red, P = orange, S = gold), and $NADP^+$ (C = yellow, N = blue, O = red, P = orange) (PDB ID, 1dqa); *right*, *P. mevalonii* class II HMGR binding with HMG-CoA (C = green, N = blue, O = red, P = orange, S = gold) and NAD^+ (C = yellow, N = blue, O = red, P = orange) (PDB ID, 1QAX).

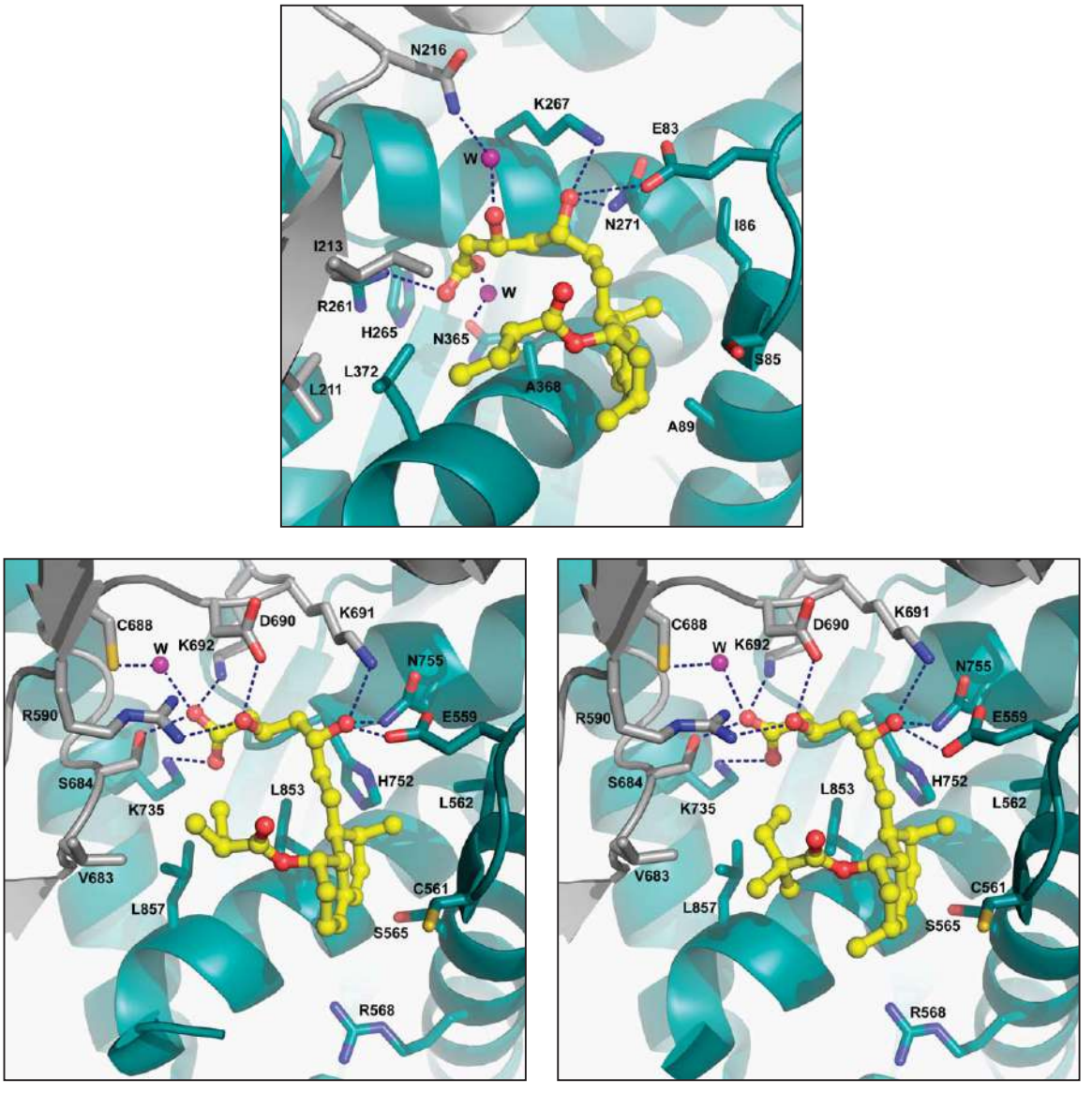


Figure 7 Interaction of the statin inhibitors with two classes of HMGR. Upper, *P. mevalonii* HMGR binding with lovastatin (PDB ID, 1t02); lower left, human HMGR binding with mevastatin (PDB ID, 1hw8); lower right, human HMGR binding with simvastatin (PDB ID, 1hw9). The inhibitors are shown in yellow with O = red. The chain A of the proteins are shown in teal whereas chain B in gray with N = blue, O = red.

GDP synthase, FDP synthase, and GGDP synthase. The involvement of these multiple enzymes results in the observed diversity of isoprenoid compounds (**Figure 8**).

1.12.4.1 Type 1 IPP Isomerase

IPP isomerase activity was first detected in crude extracts from *Saccharomyces cerevisiae* in 1959.[53] Consequently, the basic properties of this enzyme have been elucidated by studying the yeast version of the enzyme. The monomeric 33-kDa enzyme catalyzes the isomerase reaction in the presence of a divalent metal cation such as Mg^{2+} or Mn^{2+}.[54,55] Yeast IPP isomerase is completely inhibited by thiol-blocking reagents such as iodoacetamide, suggesting that a cysteine residue is involved in the catalytic mechanism.[56] Similar IPP isomerase

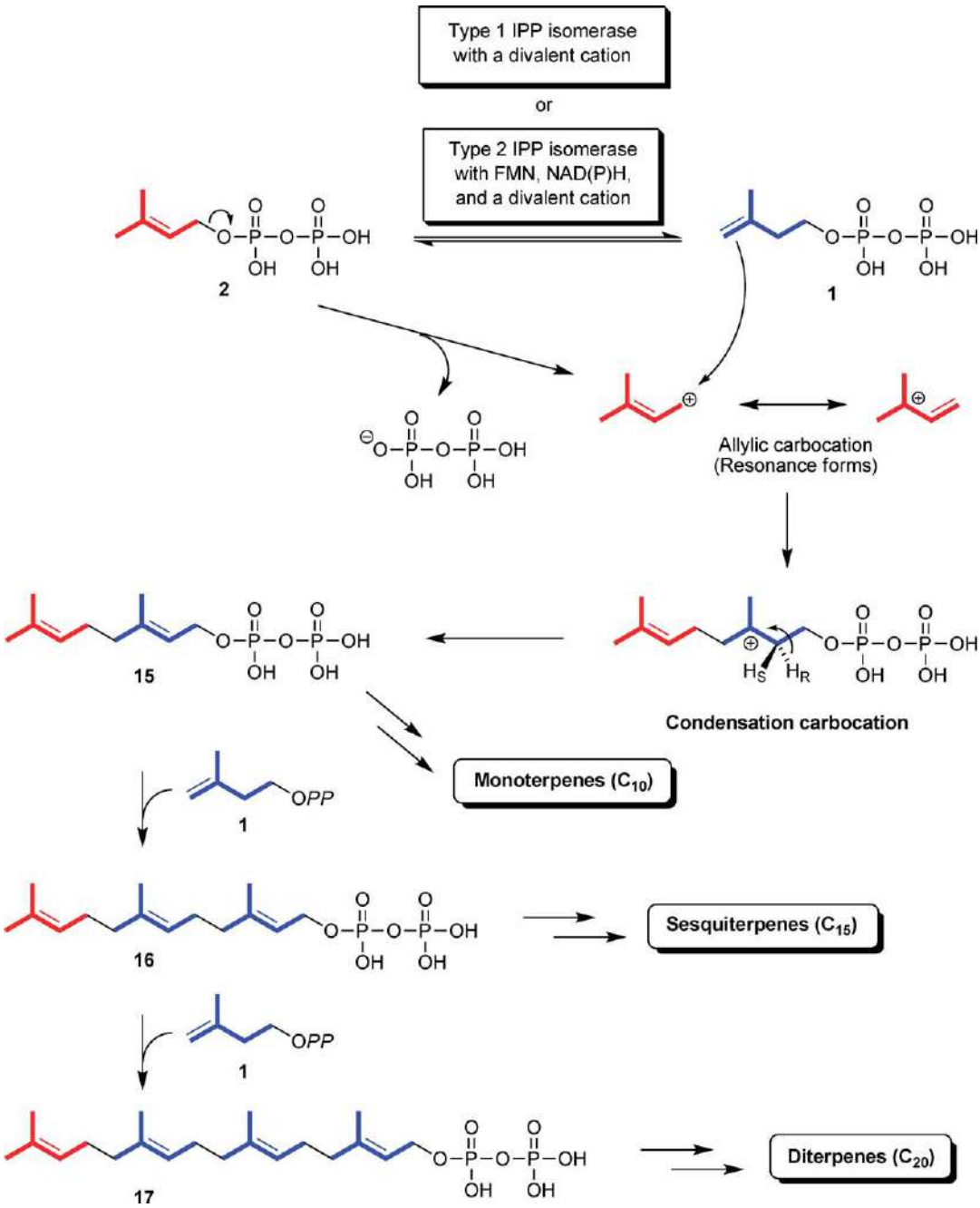


Figure 8 Reactions of IPP isomerase and polyprenyl diphosphate synthase. Bold red and blue lines represent isoprene units.

activity in *E. coli*, which is also dependent on a metal divalent cation and is inhibited by iodoacetamide, was the first reported example of the enzyme from bacteria.[57] These enzymes are now called type 1 IPP isomerases because of the later discovery of evolutionarily independent type 2 enzymes, as described in the next section.

The stereochemistry of type 1 IPP isomerase was determined by detailed labeling studies.[58,59] Using extracts from rat or pig liver, mevalonate, which was stereospecifically labeled with deuterium or tritium, was enzymatically converted into isoprenoid compounds such as FDP **16** and a sesquiterpene, squalene.

Deuterated water was used as required. The labeled isoprenoid compounds were extracted, purified, and analyzed to determine the stereospecificity of IPP isomerase. The enzyme was shown to catalyze antarafacial isomerization, in which the pro-*R* hydrogen at C-2 is specifically eliminated and a hydrogen is added to C-4 of IPP (*E*-methyl in DMAPP) from the *re* face of the double bond by a two-base mechanism (**Figure 9**). The same stereospecificity at C-2 was also confirmed with bacterial type 1 IPP isomerase from *E. coli*.[60] However, a more detailed hydrogen exchange experiment performed later using recombinant yeast enzyme showed that the *E*-methyl specificity of type 1 IPP isomerase is not as strict, since hydrogen exchange was also observed at the *Z*-methyl of DMAPP (C-5 of IPP) and at the pro-*S* hydrogen at C-2 of IPP, although this latter reaction happens very slowly.[61]

Chemically synthesized substrate analogues are powerful tools for elucidating the catalytic mechanism of type 1 IPP isomerase. In particular, the fact that 2-(dimethylamino)ethyl diphosphate (NIPP) **18** (**Figure 10**) strongly inhibits type 1 IPP isomerases ($K_i < 1.4 \times 10^{-11}$ and $< 10^{-10}$ for the enzyme from *S. cerevisiae*[54] and the ergot fungus *Claviceps purpurea*,[62] respectively) provided proof that the isomerizing reaction proceeds via the carbocation intermediate. In addition, irreversible inhibitor analogues such as (*E*)-3-(fluoromethyl)-3-butenyl diphosphate (fIPP) **19**[63] and 3-methyl-3,4-epoxybutyl diphosphate (eIPP) **20**[64] have been used in labeling experiments (**Figure 11**). The radiolabeled analogues inhibited *S. cerevisiae* IPP isomerase by forming irreversible complexes. The amino acid sequence of the labeled peptide was determined after proteolytic digestion, resulting in the assignment of a cysteine residue as the labeled site. These experiments were performed using recombinant enzyme prepared by the overexpression of the IPP isomerase gene from *S. cerevisiae*, *IDI1*, which was cloned at approximately the same time[55] and was confirmed to be essential for the growth of yeast.[65] Moreover, site-directed mutagenesis at the cysteine was shown to completely inactivate the enzyme, which suggested that the residue is an active site that acts as a catalytic base.[66] A glutamate residue was identified by mutagenesis as another possible catalytic residue.[67] The glutamate was assigned as the labeled site in a labeling experiment using fIPP **19** and a mutant IPP isomerase in which the active cysteine residue was replaced with serine. Bacterial genes of IPP isomerase were identified later by a database homology search, showing that both the cysteine and glutamate residues are highly conserved among the homologues of type 1 IPP isomerase.[68]

Crystal structures of type 1 IPP isomerases from *E. coli*[69,70] (**Figure 12**) and human[71,72] fit well with the proposed catalytic mechanism, in which two catalytic bases catalyze antarafacial proton addition/elimination.

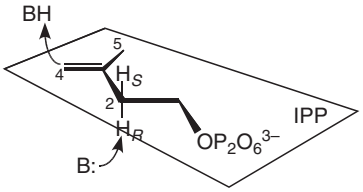
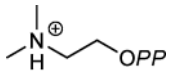
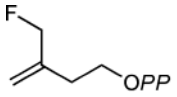


Figure 9 Antarafacial stereochemistry of the isomerization of IPP catalyzed by type 1 IPP isomerase.

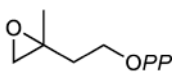


18

Figure 10 Transition-state analogue, NIPP.



19



20

Figure 11 Irreversible inhibitors, fIPP and eIPP.

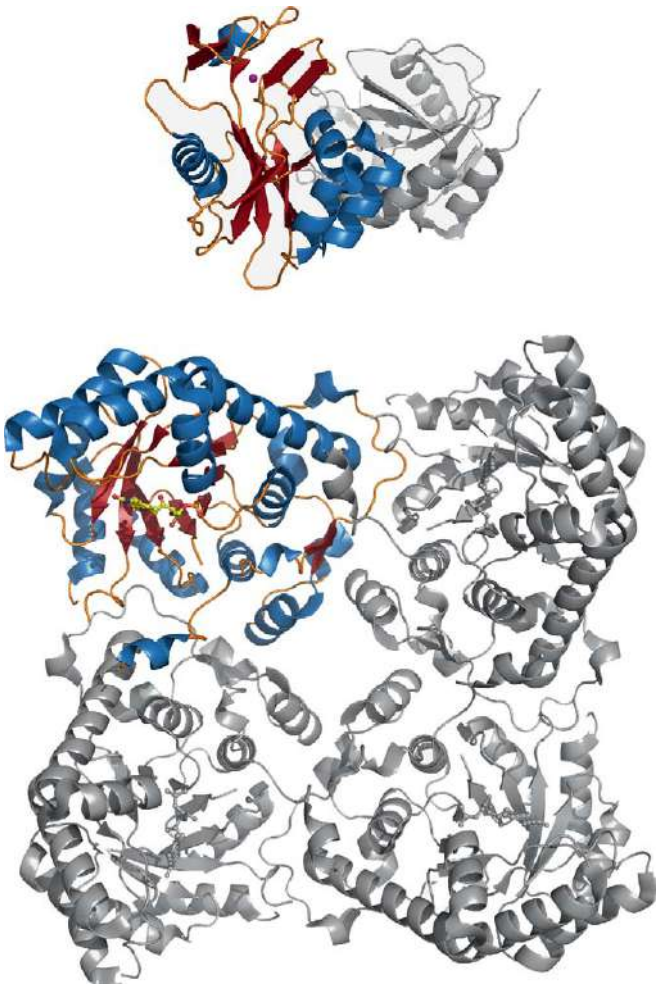


Figure 12 Crystal structures of two types of IPP isomerase. Ribbon diagrams of IPP isomerase. For clarity, only one monomer is colored. Upper, type 1 IPP isomerase from *E. coli* binding with Mn^{2+} (purple) (PDB ID, 1hx3); lower, type 2 IPP isomerase from *B. subtilis* binding with FMN (C = yellow, N = blue, O = red, P = orange) (PDB ID, 1p0n).

In particular, the structures of human IPP isomerase complexes with the substrate[72] or its analogues[71,73–75] clearly showed that the active cysteine and glutamate residues contact the substrate from opposite sides (**Figure 13**). These structures suggest that protonation of IPP by the cysteine residue occurs at C-4 to give a tertiary carbocation intermediate, followed by deprotonation of the pro-*R* hydrogen at C-2 by the glutamate residue to yield DMAPP (**Figure 14**).

1.12.4.2 Type 2 IPP Isomerase

In contrast to the newly discovered MEP pathway, which produces both IPP **1** and DMAPP **2**, the mevalonate pathway only yields IPP **1** as the active isoprene unit for isoprenoid biosynthesis, indicating that IPP isomerase is essential for organisms that possess only the mevalonate pathway. In fact, the IPP isomerase gene is dispensable for *E. coli*,[68] which utilizes the MEP pathway for isoprenoid biosynthesis, whereas the IPP isomerase gene is essential for *S. cerevisiae*,[65] which utilizes the mevalonate pathway. It has been suggested that IPP isomerase in some microorganisms is not homologous with the type 1 enzyme, since genes of type 1 IPP isomerase homologues have not been found in the whole genome sequences of several bacteria and archaea that exclusively utilize the mevalonate pathway.[76]

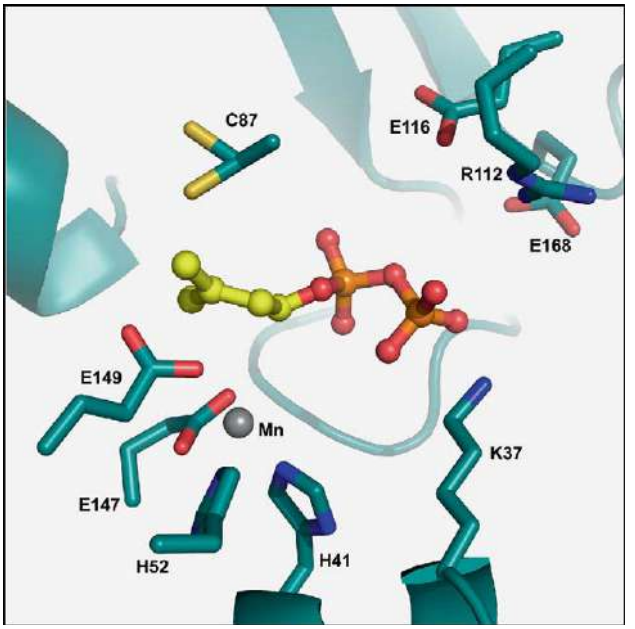


Figure 13 IPP located in the active site of type 1 IPP isomerase. In the crystal structure of human IPP isomerase (PDB ID, 2ick), Cys87 and Glu149 are located on opposite faces of the allyl moiety in IPP. IPP is shown in yellow with O = red, P = orange. The Mn^{2+} is in gray. The protein is shown in teal with N = blue, O = red.

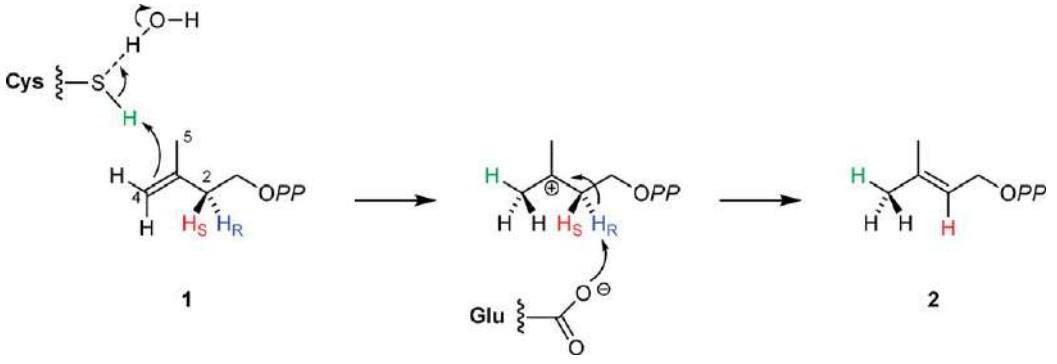


Figure 14 Reaction mechanism of the isomerization of IPP to DMAPP catalyzed by type 1 IPP isomerase. Type 1 IPP isomerase reaction occurs through a tertiary carbocation intermediate.

The first report of a new type of isomerase, type 2 IPP isomerase, was published in 2001, more than 40 years after the identification of type 1 isomerase.[77] The enzyme was isolated from *Streptomyces* sp. strain CL190, which possesses the MEP pathway and type 1 IPP isomerase.[78] Surprisingly, the organism also has a gene cluster of the mevalonate pathway of the biosynthesis of an isoprenoid secondary metabolite (see Section 1.12.2.2), and the cluster contains the gene for type 2 IPP isomerase (**Figure 1**). Type 2 IPP isomerase from CL190, which is a homotetramer of 37-kDa subunits, has no sequence homology with type 1 enzymes, suggesting that the two types of IPP isomerase are evolutionally independent. Intriguingly, the enzyme requires flavin mononucleotide (FMN) **21** and NAD(P)H, in addition to a metal divalent cation, for activity (**Figure 8**). This curious cofactor requirement is considered to have made the *in vitro* detection of this enzyme difficult, and is the main reason why this type of IPP isomerase was overlooked for so long. The pro-*R* stereospecificity of the enzyme at C-2 of IPP is the same as that of type 1 IPP isomerase,[79,80] although the reaction specificity of hydrogen exchange at C-2 and C-4 of IPP is much more stringent.[81]

Type 2 IPP isomerase is distributed among a wide range of bacteria, including pathogenic *Staphylococcus*, *Enterococcus*, and *Streptococcus*, and almost all archaea.[78,82] It should be noted that *Bacillus* [82,83] and *Synechocystis*[84] also possess type 2 IPP isomerase, although these bacteria utilize the MEP pathway and thus can produce both IPP and DMAPP. Interestingly, the species distribution of the mevalonate and MEP pathways are not associated with those of type 1 and type 2 IPP isomerases (**Figure 15**). Although most bacteria possessing the MEP pathway do not have IPP isomerase, some have type 1 enzyme, some have type 2, and others have both. On the contrary, bacteria using only the mevalonate pathway usually possess type 2 IPP isomerase. This is the case for most archaea. However, the Q-fever pathogen *Coxiella burnetii* has the mevalonate pathway genes and type 1 IPP isomerase, and the halophilic archaeon *Halobacterium* sp. NRC-1 possesses both functional type 1 and type 2 IPP isomerases, in addition to the mevalonate pathway.[85]

The amino acid sequences of type 2 IPP isomerases are homologous with those of the FMN-dependent (*S*)-α-hydroxyacid dehydrogenase family of enzymes, including flavocytochrome b_2, hydroxyacid oxidase, glycolate oxidase, and L-lactate dehydrogenase. All these enzymes have a triosephosphate isomerase (TIM)-barrel fold.[82] This sequence similarity with redox enzymes and their requirement for redox coenzymes (i.e., FMN and NAD(P)H) suggest the redox-catalyzing nature of type 2 IPP isomerase. Actually, type 2 IPP isomerase from *Bacillus subtilis* was reported to show weak catalysis of pyruvate to form lactate in the presence of NADH.[82] Moreover, an archaeal homologue from *Sulfolobus shibatae*, which binds FMN, catalyzes the NADH oxidase reaction in the absence of the substrates.[86] In fact, the crystal structures of type 2 IPP isomerases reported to date, from *B. subtilis*[87] and *Thermus thermophilus*,[88,89] also have the TIM-barrel (or β/α-barrel) fold, which is assembled into a homotetramer or a homooctamer (**Figure 12**). An FMN molecule is noncovalently bound inside each barrel in type 2 IPP isomerase, in a similar manner to that observed in the crystal structures of FMN-dependent oxidoreductases.

Because the reaction catalyzed by type 2 IPP isomerase involves no net redox change, the roles of FMN and NAD(P)H in the reaction have attracted the attention of researchers.[90,91] The fact that some type 2 IPP isomerases can catalyze redox reactions[82,86] clearly indicates that the enzyme can mediate electron transfer through FMN. In fact, the UV-visible spectrum of FMN bound in type 2 IPP isomerase showed it to be in the reduced state in the presence of NAD(P)H.[92,93] However, NADH is used only for the initial reduction of FMN and is not consumed in the accompanying isomerase reaction, suggesting that only reduced FMN ($FMNH_2$ **22**) is required for activity, and is not reoxidized at the end of the reaction.[86] This hypothesis was confirmed by the fact that type 2 IPP isomerase could catalyze the isomerizing reaction when NAD(P)H was replaced with a reducing agent, sodium dithionite.

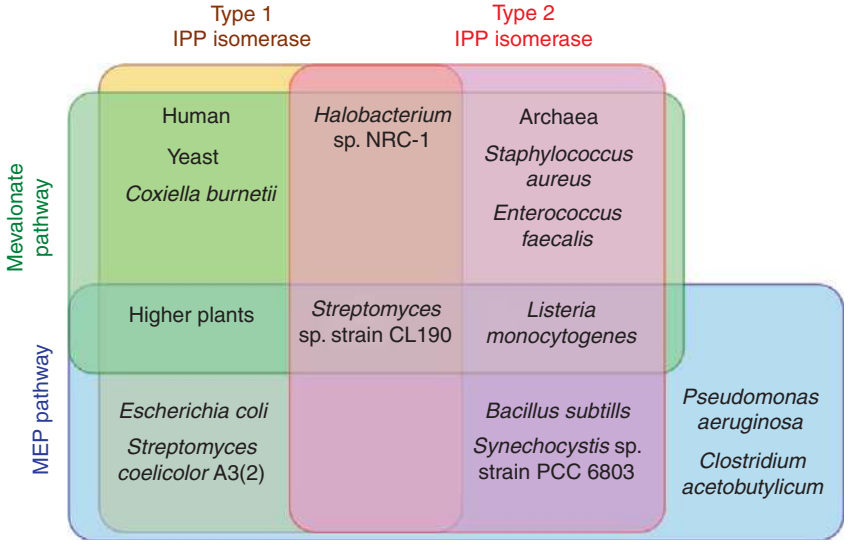


Figure 15 Distributions of two types of IPP isomerase and two distinct pathways for isoprenoid biosynthesis.

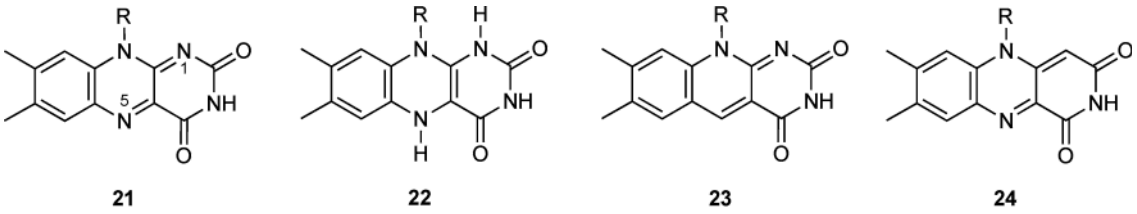


Figure 16 FMN and its analogues, 5-deazaFMN and 1-deazaFMN.

Further information about the role of FMN was obtained from experiments with flavin analogues (**Figure 16**). Reconstruction of *apo*-type 2 IPP isomerase with 5-deazaFMN **23** completely inactivated the enzyme,[92,94] whereas the replacement of FMN with 1-deazaFMN **24** had little effect on enzyme activity. Based on the knowledge of traditional flavin chemistry, these findings were regarded as proof that the enzyme catalyzes a radical reaction. In contrast, experiments using substrate analogues led to a totally different conclusion. When eIPP **20** was reacted with type 2 IPP isomerase in the presence of FMN and a reducing agent, it irreversibly inactivated the enzyme by forming an adduct **25** with FMN, probably at the N5 nitrogen (**Figure 17**).[95,96] Similar adduct formation with FMN was reported with other substrate analogues such as fIPP **19**, 3-methylene-4-penten-1-yl diphosphate (vIPP) **26**, and 3-oxiranyl-but-3-en-1-yl diphosphate (oIPP) **27**, suggesting nucleophilic attack by the N5 nitrogen of $FMNH_2$ **22** (**Figure 17**).[96] Given these results, the inactivation of type 2 IPP isomerase reconstituted with 5-deazaFMN **23** might be explained by the loss of the lone pair on FMN N5. Moreover, the fact that a 'radical clock' substrate analogue, 3-cyclopropyl-but-3-en-1-yl diphosphate (cIPP) **28**, underwent isomerization to yield 3-cyclopropyl-but-2-en-1-yl diphosphate (cDMAPP) **29** without breaking the cyclopropyl ring revealed that no radical intermediate is formed in the reaction of type 2 IPP isomerase (**Figure 18**).[97] No radical intermediate was detected by electron paramagnetic resonance (EPR), even when the substrate was coupled with the semiquinone form of FMN.[93,98] Rejection of the radical mechanism suggested an alternative proton addition/elimination mechanism, resembling that of type 1 IPP isomerase. The proton addition/elimination mechanism was supported by the fact that deuterium kinetic isotopic effects of ~2 and 1.5 were observed when IPP had deuterium at the C-2 *pro-R* position and when the reaction was performed in D_2O, respectively.[98] Given our understanding of acid–base

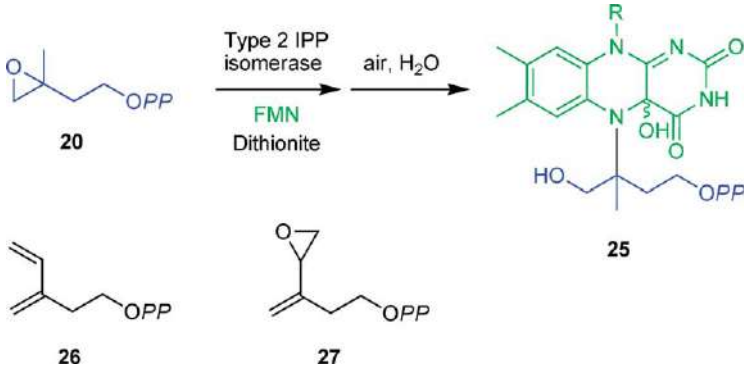


Figure 17 Irreversible inhibitors, eIPP, vIPP, and oIPP, that form adducts with FMN.

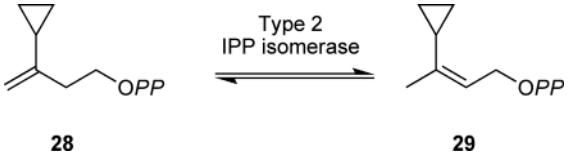
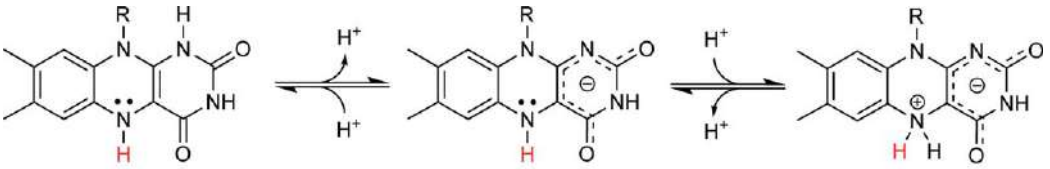


Figure 18 Radical clock substrate analogue, cIPP.



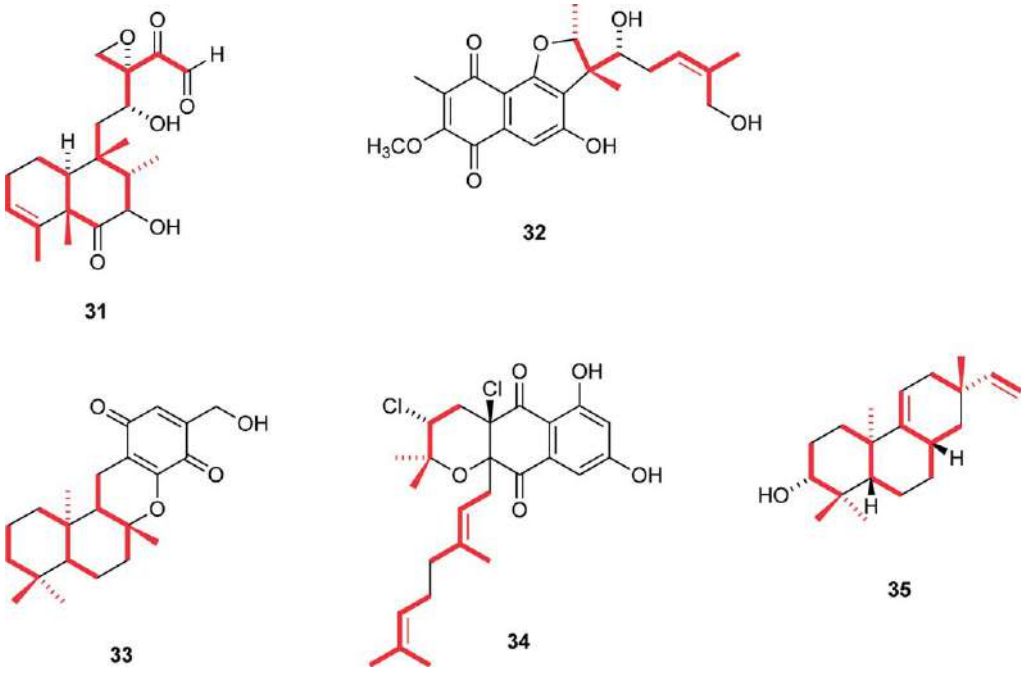
22 **30**

Figure 19 5,5-Zwitterionic species of $FMNH_2$ as an acid–base catalyst.

chemistry, some explanations for the importance of the N5 nitrogen of FMN in the reaction of type 2 IPP isomerase were proposed.[98] In one of the hypothetical mechanisms, N5 of $FMNH_2$ **22** acts as the acid–base catalyst through the generation of the 5,5-zwitterionic species **30** (**Figure 19**).[99] However, the reaction mechanism of type 2 IPP isomerase and the role of FMN remain controversial. The crystal structures reported to date show no bound substrate.[87–89] The structures show only FMN or both FMN and diphosphate, and so provide limited information about the catalytic mechanism. It is anticipated that definitive data will be obtained from crystal structures of the enzyme complexed with both FMN and the substrate (or an analogue).

1.12.5 Biosynthetic Gene Clusters of Secondary Metabolites from Actinomycetes

Terpentecin **31**,[100,101] furaquinocin A **32**,[102] BE-40644 **33**,[103] and napyradiomycin A **34**[104] are isoprenoid secondary metabolites produced by actinomycetes (**Figure 20**). To date, the isoprenoid biosynthetic gene clusters have been cloned from four actinomycetes strains that can utilize the mevalonate pathway of isoprenoid biosynthesis (**Figure 21**): *Kitasatospora griseola*, a terpentecin-producer;[105] *Streptomyces* sp. KO-3988, a furaquinocin producer;[106] *Actinoplanes* sp. A40644, a BE-40644 producer;[5] and *Streptomyces* sp. CNQ525, a napyradiomycin producer.[107] Each gene cluster is located adjacent to the mevalonate pathway gene cluster



31 **32** **33** **34** **35**

Figure 20 Isoprenoids produced by actinomycetes. Bold red lines represent isoprene units.

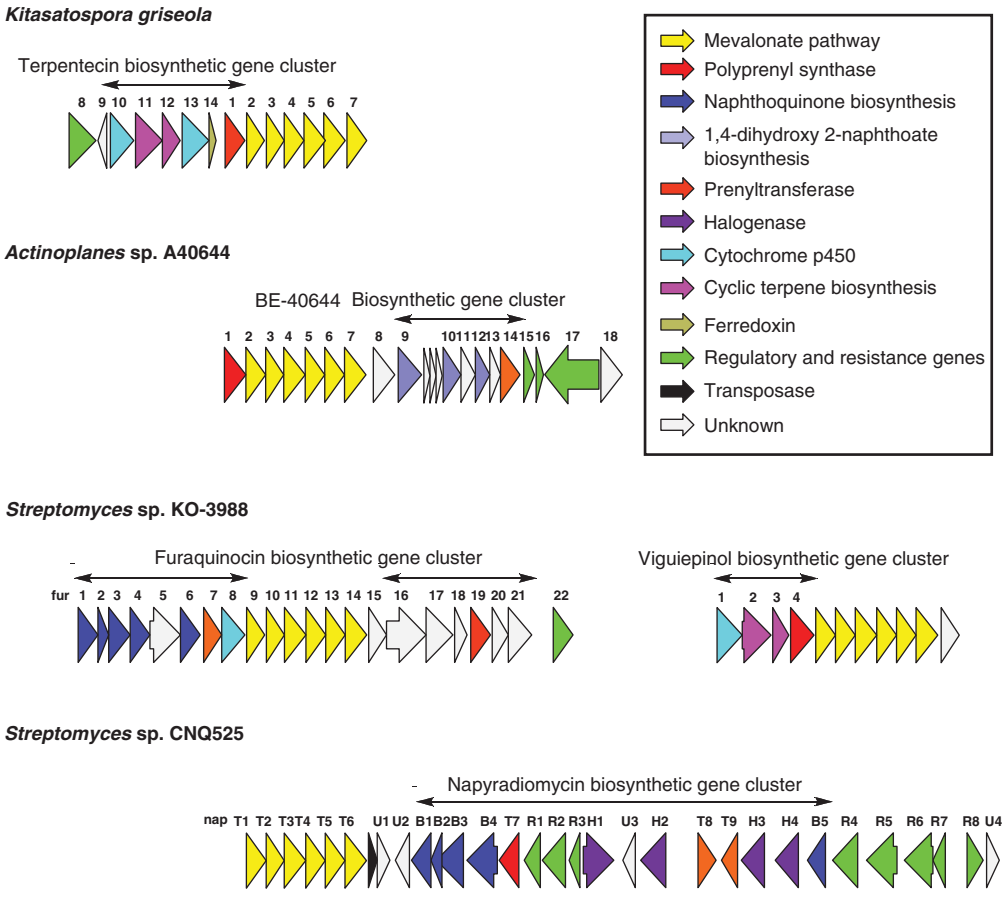


Figure 21 Isoprenoid biosynthetic gene clusters from actinomycetes.

(Figure 21). This suggests that the mevalonate pathway genes and isoprenoid biosynthetic genes are usually clustered in isoprenoid-producing actinomycetes, and that the presence of the mevalonate pathway might be a reliable marker for detecting the production of isoprenoids by actinomycetes. In fact, Dairi and coworkers identified the biosynthetic gene cluster for viguiepinol **35** by analyzing flanking regions of an additional mevalonate pathway gene cluster in *Streptomyces* sp. KO-3988 (Figure 21).[4,106]

1.12.6 Mevalonate Fermentation

Mevalonate, a ubiquitous biosynthetic intermediate of isoprenoids, is utilized as a moisturizer in cosmetics and a chemical in biochemical research. The current supply of optically active mevalonate relies on fermentation production by the yeast, *Saccharomycopsis fibuligera* ADK 8107, whose mevalonate productivity is 19 g l^{-1} for 12 days.[108] However, this fermentation process has a drawback, namely, that the yeast consumes mevalonate for growth, especially in the late log-phase stage, causing a decrease in the production yield of mevalonate. To overcome this problem, two research groups have attempted to develop an alternative method to produce mevalonate in heterologous hosts that neither require nor consume mevalonate for growth. As the heterologous host, one group chose *Streptomyces lividans* TK23[109] and the other chose *E. coli*.[110] Both organisms exclusively use only the MEP pathway for growth. Heterologous production of mevalonate was achieved by both the metabolically engineered *S. lividans* TK23 and *E. coli*.

An *S. lividans* TK23 transformant heterologously expressing acetoacetyl-CoA thiolase, HMG-CoA synthase, and HMGR genes from *Streptomyces* sp. strain CL190 was constructed. The resulting transformant

heterologously produced a peak concentration of 0.66 g l^{-1} mevalonate at 8 days.[109] However, this productivity is much lower than that of *S. fibuligera* ADK 8107 (19 g l^{-1} at 12 days).[108]

An *E. coli* transformant heterologously expressing acetoacetyl-CoA thiolase/HMG-CoA reductase gene and HMG-CoA synthase gene from *E. faecalis* was constructed. Surprisingly, the recombinant *E. coli* produced 47 g l^{-1} mevalonate in 50 h of fed-batch cultivation in a 2-l fermenter.[110] This heterologous production is the highest titer reported to date.

1.12.7 Conclusions

In this chapter, the authors have attempted to provide an overview of the mevalonate pathway for isoprenoid biosynthesis in bacteria and archaea. Bacteria and eukaryotes share a common pathway, but the mevalonate pathway in archaea differs slightly from that of bacteria and eukaryotes in the final steps that are not fully elucidated. Two distinct classes of HMG-CoA reductase and two types of IPP isomerase are distributed in bacteria and archaea. Further analysis of the mevalonate pathway in bacteria and archaea may illustrate the interplay of functional convergence and divergence in the evolution of metabolic pathways.

Abbreviations

ADP	adenosine 5′-diphosphate
ATP	adenosine 5′-triphosphate
cDMAPP	3-cyclopropyl-but-2-en-1-yl diphosphate
cIPP	3-cyclopropyl-but-3-en-1-yl diphosphate
DMAPP	dimethylallyl diphosphate
eIPP	3-methyl-3,4-epoxybutyl diphosphate
EPR	electron paramagnetic resonance
FDP	farnesyl diphosphate
fIPP	(*E*)-3-(fluoromethyl)-3-butenyl diphosphate
FMN	flavin mononucleotide
$FMNH_2$	reduced form of FMN
GDP	geranyl diphosphate
GGDP	geranylgeranyl diphosphate
HMG-CoA	3-hydroxy-3-methylglutaryl-coenzyme A
HMGR	HMG-CoA reductase
IP	isopentenyl phosphate
IPP	isopentenyl diphosphate
MEP	2-*C*-methyl-D-erythritol 4-phosphate
NADH	reduced form of nicotinamide adenine dinucleotide
NADPH	reduced form of nicotinamide adenine dinucleotide phosphate
NIPP	2-(dimethylamino)ethyl diphosphate
oIPP	3-oxiranyl-but-3-en-1-yl diphosphate
TIM	triosephosphate isomerase
vIPP	3-methylene-4-penten-1-yl diphosphate

Nomenclature

K_i	dissociation constant of an inhibitor for binding to an enzyme
K_m	Michaelis-Menten constant

References

1. D. E. Cane, Ed.; *Isoprenoids Including Carotenoids and Steroids*; Amsterdam, Netherlands: Elsevier, 1999; Vol. 2.
2. L. L. Grochowski; H. Xu; R. H. White, *J. Bacterial.* **2006**, *188*, 3192–3198.
3. M. Takagi; T. Kuzuyama; S. Takahashi; H. Seto, *J. Bacteriol.* **2000**, *182*, 4153–4157.
4. T. Kawasaki; T. Kuzuyama; Y. Kuwamori; N. Matsuura; N. Itoh; K. Furihata; H. Seto; T. Dairi, *J. Antibiot. (Tokyo)* **2004**, *57*, 739–747.
5. T. Kawasaki; T. Kuzuyama; K. Furihata; N. Itoh; H. Seto; T. Dairi, *J. Antibiot. (Tokyo)* **2003**, *56*, 957–966.
6. J. Ishikawa; A. Yamashita; Y. Mikami; Y. Hoshino; H. Kurita; K. Hotta; T. Shiba; M. Hattori, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14925–14930.
7. G. Glockner; R. Lehmann; A. Romualdi; S. Pradella; U. Schulte-Spechtel; M. Schilhabel; B. Wilske; J. Suhnel; M. Platzer, *Nucleic Acids Res.* **2004**, *32*, 6038–6046.
8. M. Humbelin; A. Thomas; J. Lin; J. Li; J. Jore; A. Berry, *Gene* **2002**, *297*, 129–139.
9. M. Hedl; A. Sutherlin; E. I. Wilding; M. Mazzulla; D. McDevitt; P. Lane; J. W. Burgner, II; K. R. Lehnbeuter; C. V. Stauffacher; M. N. Gwynn; V. W. Rodwell, *J. Bacteriol.* **2002**, *184*, 2116–2122.
10. A. Sutherlin; M. Hedl; B. Sanchez-Neri; J. W. Burgner, II; C. V. Stauffacher; V. W. Rodwell, *J. Bacterial.* **2002**, *184*, 4065–4070.
11. E. I. Wilding; D. Y. Kim; A. P. Bryant; M. N. Gwynn; R. D. Lunsford; D. McDevitt; J. E. Myers, Jr.; M. Rosenberg; D. Sylvester; C. V. Stauffacher; V. W. Rodwell, *J. Bacterial.* **2000**, *182*, 5147–5152.
12. A. W. Alberts; J. Chen; G. Kuron; V. Hunt; J. Huff; C. Hoffman; J. Rothrock; M. Lopez; H. Joshua; E. Harris; A. Patchett; R. Monaghan; S. Currie; E. Stapley; G. Albers-Schonberg; O. Hensens; J. Hirshfield; K. Hoogsteen; J. Liesch; J. Springer, *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 3957–3961.
13. A. Endo, *J. Lipid Res.* **1992**, *33*, 1569–1582.
14. D. A. Bochar; C. V. Stauffacher; V. W. Rodwell, *Mol. Genet. Metab.* **1999**, *66*, 122–127.
15. J. A. Friesen; V. W. Rodwell, *Genome Biol.* **2004**, *5*, 248.
16. M. Hedl; V. W. Rodwell, *Protein Sci.* **2004**, *13*, 1693–1697.
17. J. D. Thompson; D. G. Higgins; T. J. Gibson, *Nucleic Acids Res.* **1994**, *22*, 4673–4680.
18. K. M. Bischoff; V. W. Rodwell, *J. Bacteriol.* **1996**, *178*, 19–23.
19. D. A. Bochar; J. R. Brown; W. F. Doolittle; H. P. Klenk; W. Lam; M. E. Schenk; C. V. Stauffacher; V. W. Rodwell, *J. Bacteriol.* **1997**, *179*, 3632–3638.
20. D. Y. Kim; D. A. Bochar; C. V. Stauffacher; V. W. Rodwell, *Protein Exp. Purif.* **1999**, *17*, 435–442.
21. S. Takahashi; T. Kuzuyama; H. Seto, *J. Bacteriol.* **1999**, *181*, 1256–1263.
22. T. Dairi; Y. Motohira; T. Kuzuyama; S. Takahashi; N. Itoh; H. Seto, *Mol. Gen. Genet.* **2000**, *262*, 957–964.
23. M. De Rosa; A. Gambacorta; A. Gliozzi, *Microbiol. Rev.* **1986**, *50*, 70–80.
24. M. Kates; M. K. Wassef; D. J. Kushner, *Can. J. Biochem.* **1968**, *46*, 971–977.
25. J. A. Cabrera; J. Bolds; P. E. Shields; C. M. Havel; J. A. Watson, *J. Biol. Chem.* **1986**, *261*, 3578–3583.
26. W. L. Lam; W. F. Doolittle, *J. Biol. Chem.* **1992**, *267*, 5829–5834.
27. T. Kuzuyama; S. Takahashi; T. Dairi; H. Seto, *J. Antibiot. (Tokyo)* **2002**, *55*, 919–923.
28. J. M. Sigmund; D. C. Clark; F. A. Rainey; A. S. Anderson, *Microb. Ecol.* **2003**, *46*, 106–112.
29. J. L. Goldstein; M. S. Brown, *Nature* **1990**, *343*, 425–430.
30. C. F. Hunter; V. W. Rodwell, *J. Lipid Res.* **1980**, *21*, 399–405.
31. S. Dale; M. Arro; B. Becerra; N. G. Morrice; A. Boronat; D. G. Hardie; A. Ferrer, *Eur. J. Biochem.* **1995**, *233*, 506–513.
32. R. V. Omkumar; V. W. Rodwell, *J. Biol. Chem.* **1994**, *269*, 16862–16866.
33. D. Y. Kim; C. V. Stauffacher; V. W. Rodwell, *Biochemistry* **2000**, *39*, 2269–2275.
34. K. M. Bischoff; V. W. Rodwell, *Protein Sci.* **1997**, *6*, 156–161.
35. V. W. Rodwell; W. R. Bensch, *Methods Enzymol.* **1981**, *71* (Pt C), 480–486.
36. J. F. Gill, Jr.; M. J. Beach; V. W. Rodwell, *J. Biol. Chem.* **1985**, *260*, 9393–9398.
37. T. C. Jordan-Starck; V. W. Rodwell, *J. Biol. Chem.* **1989**, *264*, 17913–17918.
38. M. J. Beach; V. W. Rodwell, *J. Bacteriol.* **1989**, *171*, 2994–3001.
39. D. Y. Kim; C. V. Stauffacher; V. W. Rodwell, *Protein Sci.* **2000**, *9*, 1226–1234.
40. A. E. Theivagt; E. N. Amanti; N. J. Beresford; L. Tabernero; J. A. Friesen, *Biochemistry* **2006**, *45*, 14397–14406.
41. E. S. Istvan; J. Deisenhofer, *Science* **2001**, *292*, 1160–1164.
42. E. S. Istvan; M. Palnitkar; S. K. Buchanan; J. Deisenhofer, *EMBO J.* **2000**, *19*, 819–830.
43. C. M. Lawrence; V. W. Rodwell; C. V. Stauffacher, *Science* **1995**, *268*, 1758–1762.
44. L. Tabernero; D. A. Bochar; V. W. Rodwell; C. V. Stauffacher, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 7167–7171.
45. L. Tabernero; V. W. Rodwell; C. V. Stauffacher, *J. Biol. Chem.* **2003**, *278*, 19933–19938.
46. Y. Wang; B. G. Darnay; V. W. Rodwell, *J. Biol. Chem.* **1990**, *265*, 21634–21641.
47. K. Frimpong; V. W. Rodwell, *J. Biol. Chem.* **1994**, *269*, 1217–1221.
48. B. G. Darnay; V. W. Rodwell, *J. Biol. Chem.* **1993**, *268*, 8429–8435.
49. B. G. Darnay; Y. Wang; V. W. Rodwell, *J. Biol. Chem.* **1992**, *267*, 15064–15070.
50. D. A. Bochar; L. Tabernero; C. V. Stauffacher; V. W. Rodwell, *Biochemistry* **1999**, *38*, 8879–8883.
51. D. A. Bochar; C. V. Stauffacher; V. W. Rodwell, *Biochemistry* **1999**, *38*, 15848–15852.
52. E. I. Wilding; J. R. Brown; A. P. Bryant; A. F. Chalker; D. J. Holmes; K. A. Ingraham; S. Iordanescu; C. Y. So; M. Rosenberg; M. N. Gwynn, *J. Bacteriol.* **2000**, *182*, 4319–4327.
53. B. W. Agranoff; H. Eggerer; U. Henning; F. Lynen, *J. Am. Chem. Soc.* **1959**, *81*, 1254–1255.
54. J. E. Reardon; R. H. Abeles, *Biochemistry* **1986**, *25*, 5609–5616.
55. M. S. Anderson; M. Muehlbacher; I. P. Street; J. Proffitt; C. D. Poulter, *J. Biol. Chem.* **1989**, *264*, 19169–19175.
56. B. W. Agranoff; H. Eggerer; U. Henning; F. Lynen, *J. Biol. Chem.* **1960**, *235*, 326–332.
57. S. Fujisaki; T. Nishino; H. Katsuki, *J. Biochem.* **1986**, *99*, 1327–1337.

58. J. W. Cornforth; R. H. Cornforth; C. Donninger; G. Popjak, *Proc. R. Soc. Lond. B: Biol. Sci.* **1966**, *163*, 492–514.
59. J. W. Cornforth; K. Clifford; R. Mallaby; G. T. Phillips, *Proc. R. Soc. Lond. B: Biol. Sci.* **1972**, *182*, 277–295.
60. A. E. Leyes; J. A. Baker; F. M. Hahn; C. D. Poulter, *Chem. Commun.* **1999**, 717–718.
61. I. P. Street; D. J. Christensen; C. D. Poulter, *J. Am. Chem. Soc.* **1990**, *112*, 8577–8578.
62. M. Muehlbacher; C. D. Poulter, *Biochemistry* **1988**, *27*, 7315–7328.
63. I. P. Street; C. D. Poulter, *Biochemistry* **1990**, *29*, 7531–7538.
64. X. J. Lu; D. J. Christensen; C. D. Poulter, *Biochemistry* **1992**, *31*, 9955–9960.
65. M. P. Mayer; F. M. Hahn; D. J. Stillman; C. D. Poulter, *Yeast* **1992**, *8*, 743–748.
66. I. P. Street; H. R. Coffman; C. D. Poulter, *Tetrahedron* **1991**, *47*, 5919–5924.
67. I. P. Street; H. R. Coffman; J. A. Baker; C. D. Poulter, *Biochemistry* **1994**, *33*, 4212–4217.
68. F. M. Hahn; J. A. Baker; C. D. Poulter, *J. Bacteriol.* **1996**, *178*, 619–624.
69. J. B. Bonanno; C. Edo; N. Eswar; U. Pieper; M. J. Romanowski; V. Ilyin; S. E. Gerchman; H. Kycia; F. W. Studier; A. Sali; S. K. Burley, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12896–12901.
70. V. Durbecq; G. Sainz; Y. Oudjama; B. Clantin; C. Bompard-Gilles; C. Tricot; J. Caillet; V. Stalon; L. Droogmans; V. Villeret, *EMBO J.* **2001**, *20*, 1530–1537.
71. C. Zhang; L. Liu; H. Xu; Z. Wei; Y. Wang; Y. Lin; W. Gong, *J. Mol. Biol.* **2007**, *366*, 1437–1446.
72. W. Zheng; F. Sun; M. Bartlam; X. Li; R. Li; Z. Rao, *J. Mol. Biol.* **2007**, *366*, 1447–1458.
73. J. Wouters; Y. Oudjama; S. J. Barkley; C. Tricot; V. Stalon; L. Droogmans; C. D. Poulter, *J. Biol. Chem.* **2003**, *278*, 11903–11908.
74. J. Wouters; Y. Oudjama; S. Ghosh; V. Stalon; L. Droogmans; E. Oldfield, *J. Am. Chem. Soc.* **2003**, *125*, 3198–3199.
75. J. de Ruyck; V. Durisotti; Y. Oudjama; J. Wouters, *J. Biol. Chem.* **2006**, *281*, 17864–17869.
76. A. Smit; A. Mushegian, *Genome Res.* **2000**, *10*, 1468–1484.
77. K. Kaneda; T. Kuzuyama; M. Takagi; Y. Hayakawa; H. Seto, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 932–937.
78. T. Kuzuyama; H. Seto, *Nat. Prod. Rep.* **2003**, *20*, 171–183.
79. C. L. Kao; W. Kittleman; H. Zhang; H. Seto; H. W. Liu, *Org. Lett.* **2005**, *7*, 5677–5680.
80. R. Laupitz; T. Grawert; C. Rieder; F. Zepeck; A. Bacher; D. Arigoni; F. Rohdich; W. Eisenreich, *Chem. Biodivers.* **2004**, *1*, 1367–1376.
81. S. J. Barkley; S. B. Desai; C. D. Poulter, *Org. Lett.* **2004**, *6*, 5019–5021.
82. R. Laupitz; S. Hecht; S. Amslinger; F. Zepeck; J. Kaiser; G. Richter; N. Schramek; S. Steinbacher; R. Huber; D. Arigoni; A. Bacher; W. Eisenreich; F. Rohdich, *Eur. J. Biochem.* **2004**, *271*, 2658–2669.
83. M. Takagi; K. Kaneda; T. Shimizu; Y. Hayakawa; H. Seto; T. Kuzuyama, *Biosci. Biotechnol. Biochem.* **2004**, *68*, 132–137.
84. S. J. Barkley; S. B. Desai; C. D. Poulter, *J. Bacteriol.* **2004**, *186*, 8156–8158.
85. T. Hoshino; T. Eguchi, *Biosci. Biotechnol. Biochem.* **2007**, *71*, 2588–2591.
86. S. Yamashita; H. Hemmi; Y. Ikeda; T. Nakayama; T. Nishino, *Eur. J. Biochem.* **2004**, *271*, 1087–1093.
87. S. Steinbacher; J. Kaiser; S. Gerhardt; W. Eisenreich; R. Huber; A. Bacher; F. Rohdich, *J. Mol. Biol.* **2003**, *329*, 973–982.
88. J. de Ruyck; S. C. Rothman; C. D. Poulter; J. Wouters, *Biochem. Biophys. Res. Commun.* **2005**, *338*, 1515–1518.
89. J. de Ruyck; J. Pouyez; S. C. Rothman; D. Poulter; J. Wouters, *Biochemistry* **2008**, *47*, 9051–9053.
90. S. Bornemann, *Nat. Prod. Rep.* **2002**, *19*, 761–772.
91. S. O. Mansoorabadi; C. J. Thibodeaux; H. W. Liu, *J. Org. Chem.* **2007**, *72*, 6329–6342.
92. H. Hemmi; Y. Ikeda; S. Yamashita; T. Nakayama; T. Nishino, *Biochem. Biophys. Res. Commun.* **2004**, *322*, 905–910.
93. S. C. Rothman; T. R. Helm; C. D. Poulter, *Biochemistry* **2007**, *46*, 5437–5445.
94. W. Kittleman; C. J. Thibodeaux; Y. N. Liu; H. Zhang; H. W. Liu, *Biochemistry* **2007**, *46*, 8401–8413.
95. T. Hoshino; H. Tamegai; K. Kakinuma; T. Eguchi, *Bioorg. Med. Chem.* **2006**, *14*, 6555–6559.
96. S. C. Rothman; J. B. Johnston; S. Lee; J. R. Walker; C. D. Poulter, *J. Am. Chem. Soc.* **2008**, *130*, 4906–4913.
97. J. B. Johnston; J. R. Walker; S. C. Rothman; C. D. Poulter, *J. Am. Chem. Soc.* **2007**, *129*, 7740–7741.
98. C. J. Thibodeaux; S. O. Mansoorabadi; W. Kittleman; W. C. Chang; H. W. Liu, *Biochemistry* **2008**, *47*, 2547–2558.
99. P. Macheroux; S. Ghisla; C. Sanner; H. Ruterjans; F. Muller, *BMC Biochem.* **2005**, *6*, 26.
100. T. Tamamura; T. Sawa; K. Isshiki; T. Masuda; Y. Homma; H. Inuma; H. Naganawa; M. Hamada; T. Takeuchi; H. Umezawa, *J. Antibiot. (Tokyo)* **1985**, *38*, 1664–1669.
101. K. Isshiki; T. Tamamura; Y. Takahashi; T. Sawa; H. Naganawa; T. Takeuchi; H. Umezawa, *J. Antibiot. (Tokyo)* **1985**, *38*, 1819–1821.
102. K. Komiyama; S. Funayama; Y. Anraku; M. Ishibashi; Y. Takahashi; S. Omura, *J. Antibiot. (Tokyo)* **1990**, *43*, 247–252.
103. K. Torigoe; N. Wakasugi; N. Sakaizumi; T. Ikejima; H. Suzuki; K. Kojiri; H. Suda, *J. Antibiot. (Tokyo)* **1996**, *49*, 314–317.
104. K. Shiomi; H. Nakamura; H. Iinuma; H. Naganawa; K. Isshiki; T. Takeuchi; H. Umezawa; Y. Iitaka, *J. Antibiot. (Tokyo)* **1986**, *39*, 494–501.
105. T. Dairi; Y. Hamano; T. Kuzuyama; N. Itoh; K. Furihata; H. Seto, *J. Bacteriol.* **2001**, *183*, 6085–6094.
106. T. Kawasaki; Y. Hayashi; T. Kuzuyama; K. Furihata; N. Itoh; H. Seto; T. Dairi, *J. Bacteriol.* **2006**, *188*, 1236–1244.
107. J. M. Winter; M. C. Moffitt; E. Zazopoulos; J. B. McAlpine; P. C. Dorrestein; B. S. Moore, *J. Biol. Chem.* **2007**, *282*, 16362–16368.
108. H. Yamashita; T. Shimada; H. E. P. Sugiyama 392,346 (Oct. 17, 1990), *Chem. Abstr.* **1991**, *114*, 205566h.
109. T. Kuzuyama; T. Dairi; H. Yamashita; Y. Shoji; H. Seto, *Biosci. Biotechnol. Biochem.* **2004**, *68*, 931–934.
110. K. Tabata; S. Hashimoto, *Biotechnol. Lett.* **2004**, *26*, 1487–1491.

Biographical Sketches

Tomohisa Kuzuyama was born in Yokosuka, Japan, in 1966. He completed his Ph.D. on a thesis entitled 'Studies on the Fosfomycin Biosynthesis' under the supervision of Professor Haruo Seto at the Graduate School of Agricultural and Life Sciences, The University of Tokyo, in 1995. In the same year, he was appointed as an assistant professor in the Institute of Molecular and Cellular Biosciences (formerly known as the Institute of Applied Microbiology), The University of Tokyo, where he began to study the mevalonate pathway and the MEP pathway with Professor Haruo Seto. He is currently an associate professor at the Biotechnology Research Center, The University of Tokyo, since 2004. His current research interests aim at understanding the biosyntheses of hybrid terpenoid-polyketid compounds and the chemoenzymatic syntheses of bioactive compounds.

Hisashi Hemmi was born in Miyagi, Japan, in 1970. He received his B.S. (1993), M.S. (1995), and Ph.D. (1998) from Tohoku University. He worked as a research associate at the Graduate School of Engineering, Tohoku University from 1998 to 2005. He was also a visiting scientist at the University of Nebraska, Lincoln from 2004 to 2005. He has been appointed as an associate professor at the Graduate School of Bioagricultural Sciences, Nagoya University from 2005. His work focuses on archaeal enzymes for the biosynthesis of isoprenoid compounds such as archaeal membrane lipid, respiratory quinone, and glycosyl-carrier lipid. His current interest is in the enzymology and crystallography of type 2 isopentenyl diphosphate isomerase. Currently, he serves as a member of the editorial board of the *Journal of Bioscience and Bioengineering*, published by the Society for Biotechnology, Japan.

Shunji Takahashi was born in Yokohama, Japan, in 1967. He completed his Ph.D. on a thesis entitled 'Studies on Sulfate Assimilation Pathway in Plants' under the supervision of Professor Goro Tamura at the Department of Bioresources Chemistry, Chiba University in 1997. He did his postdoctoral study on the mevalonate pathway and the MEP pathway under the supervision of Professor Haruo Seto. In 1998, he moved to the Graduate School of Medicine, Chiba University as an assistant professor. In 2002, he developed a technique on sesquiterpene hydroxylase and sesquiterpene production in yeast under the supervision of Professor Joe Chappell at the University of Kentucky. In 2005, he moved to the Antibiotic Laboratory organized by Professor Hiroyuki Osada and started the biosynthetic study of polyketide compounds in *Streptomyces*. He is currently a senior scientist of the Chemical Biology Department, Advanced Science Institute, RIKEN. His research interest is focused on understanding polyketide chain organization and modification process to create new natural products.

1.13 Methylerythritol Phosphate Pathway

Michel Rohmer, Université de Strasbourg/CNRS, Strasbourg, France

© 2010 Elsevier Ltd. All rights reserved.

1.13.1	**Introduction**	518
1.13.1.1	Isoprenoids	518
1.13.1.2	Formation of Isoprene Units: Mevalonate or Methylerythritol Phosphate Pathway	518
1.13.1.3	The First Two C_5 Intermediates: 1-Deoxy-D-Xylulose 5-Phosphate and 2-*C*-Methyl-D-Erythritol 4-Phosphate	522
1.13.2	**Identification of the Genes and the Enzymes**	522
1.13.2.1	1-Deoxy-D-Xylulose 5-Phosphate and 1-Deoxy-D-Xylulose 5-Phosphate Synthase	522
1.13.2.2	2-*C*-Methyl-D-Erythritol 4-Phosphate and 2-*C*-methyl-D-Erythritol 4-Phosphate Reducto-Isomerase	524
1.13.2.2.1	*dxr* gene discovery	524
1.13.2.2.2	Reaction mechanism	525
1.13.2.2.3	Enzyme structure	527
1.13.2.2.4	Deoxyxylulose phosphate reducto-isomerase inhibitors and substrate analogues	527
1.13.2.3	From 2-*C*-Methyl-D-Erythritol 4-Phosphate to 2-*C*-Methyl-D-Erythritol 2,4-Cyclodiphosphate	530
1.13.2.3.1	4-Diphosphocytidyl-2-*C*-methyl-D-erythritol and 4-diphosphocytidyl-2-*C*-methyl-D-erythritol synthase	531
1.13.2.3.2	4-D-Diphosphocytidyl-2-*C*-methyl-D-erythritol 2-phosphate and 4-diphosphocytidyl-2-*C*-methyl-D-erythritol kinase	532
1.13.2.3.3	2-*C*-m,ethyl-D-erythritol 2,4-cyclodiphosphate and 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate synthase	533
1.13.2.3.4	Bacterial bifunctional 4-diphosphocytidyl-2-*C*-methyl-D-erythritol synthase/ 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate synthase	534
1.13.2.4	4-Hydroxy-2-Methylbut-2-Enyl Diphosphate and 2-*C*-Methyl-D-Erythritol 2,4-Cyclodiphosphate Reductase	535
1.13.2.5	The Formation of Isopentenyl Diphosphate and Dimethylallyl Diphosphate and 4-Hydroxy-2-Methylbut-2-Enyl Diphosphate Reductase	537
1.13.2.5.1	Presence of a branching in the MEP pathway producing IPP and DMAPP	537
1.13.2.5.2	4-Hydroxy-2-methylbut-2-enyl diphosphate reductase	538
1.13.3	**Distribution of the MEP Pathway**	542
1.13.3.1	Distribution of the MEP Pathway in Prokaryotes	542
1.13.3.2	Presence of the MEP Pathway in Phototrophic Phyla	542
1.13.3.2.1	Higher plants, ferns, and mosses	542
1.13.3.2.2	Algae and related nonphototrophic phyla	543
1.13.4	**Openings and Further Developments**	544
1.13.4.1	The MEP Pathway: A Target for Antimicrobial Drugs	544
1.13.4.2	Overproduction of MEP Pathway-Derived Isoprenoids in Bacteria and Plants	545
1.13.4.3	Isoprenoid Biosynthesis in Plants: MVA Versus MEP Pathway: Cross-Talk between the Cytosolic and the Plastidial Compartments	546
1.13.4.4	IPP and DMAPP Production in Bacteria and Plants: HMBPP Reductase Versus IPP/DMAPP Isomerase	547
1.13.4.5	Light Influence on the MEP Pathway in Plants	548
References		549

1.13.1 Introduction

1.13.1.1 Isoprenoids

Isoprenoids represent the most diverse family of natural products (**Figures 1** and **2**).[1,2] Their carbon skeleton is formally derived from the branched C_5 skeleton of isoprene **1** (**Figure 1**). This isoprene motif is readily recognized in the acyclic members of the series and can be found in the more complex compounds if one considers in addition cyclization reactions, Wagner–Meerwein rearrangements, and further modifications of the carbon skeleton by oxidative removal of carbon atoms or methylation reactions.

Isoprenoids include metabolites that are essential for cell life: for example, sterols **11** (**Figure 2**) modulating the permeability and the fluidity of the plasma membrane in eukaryotic cells and acting as precursors of steroid hormones in vertebrates, the acyclic prenyl chains of the quinones (ubiquinone **16**, menaquinone **17**, plastoquinone **18**, **Figure 2**) in electron transport chains or involved in the protein prenylation in eukaryotes, carotenoids, and the phytyl **13** chain of chlorophylls (**Figure 2**) in the photosynthetic apparatus of phototrophic organisms. The vast majority of isoprenoids has, however, a less obvious biological significance. This includes most members of series especially found in plants, in invertebrates, and in some microorganisms: for example, monoterpenes (C_{10}) derived from geranyl diphosphate **4** (**Figure 1**) and sesquiterpenes (C_{15}) derived from farnesyl diphosphate **5** from plant essential oils, diterpenes (C_{20}) derived from geranylgeranyl diphosphate **6** and triterpenes (C_{30}) derived from squalene **9** from plants and bacteria, and finally, the prenyl moiety from composite natural products (meroterpenoids such as prenylated aromatic rings in plants or in microorganisms) (**Figure 1**). The presence of such secondary metabolites is thought to infer some ecological and adaptative advantages to the producing organisms.

1.13.1.2 Formation of Isoprene Units: Mevalonate or Methylerythritol Phosphate Pathway

Isopentenyl diphosphate (IPP) **2** and dimethylallyl diphosphate (DMAPP) **3** (**Figure 1**) correspond to the biological equivalents of isoprene and are the universal precursors of the isoprenoids in all living organisms. The formation of isoprene units was first investigated in the 1950s in liver tissues and in yeast and led to the elucidation of the well-known mevalonate **22** (MVA) pathway (**Figure 3**).[3,4] This pathway is present in all animals and fungi. It was also rapidly obvious that this pathway is involved in the biosynthesis of plant sterols and triterpenes. The MVA pathway was rapidly accepted as the unique pathway for the formation of IPP and DMAPP in all living organisms, and this despite early contradictory results. Several labeling experiments essentially performed with ^{14}C-labeled MVA or acetate on isoprenoids of the photosynthetic apparatus of plant chloroplasts (carotenoids, phytol from chlorophyll) or of some bacterial isoprenoids (prenyl chain of ubiquinone in *Escherichia coli*, sesquiterpene lactones from a *Streptomyces* sp.) could not be interpreted in the frame of the unanimously accepted MVA pathway.[5,6]

The discovery of an alternative route toward IPP and DMAPP was in fact fortuitous. It was merely an unexpected side product of our activity in the field of chemistry and biochemistry of the bacterial triterpenoids of the hopane series (**12**, **Figure 2**). These natural products are characterized by a unique feature, that is, by an additional polyfunctionalized *n*-alkyl side chain that is linked by a carbon–carbon bond to the triterpenic hopane moiety. Attempts to elucidate the origin of this unusual peculiarity were performed by the incorporation of ^{13}C-labeled acetate. The origin of the side chain could be deciphered: it corresponds to a D-pentose issued from nonoxidative pentose phosphate pathway and linked through its C-5 carbon atom to the hopane isopropyl side chain. In contrast, the labeling pattern observed in the isoprene units did not fit well with a direct incorporation of acetate into the MVA pathway.[7] These first experiments pointed out a crucial problem concerning the formation of isoprene units in several bacteria that could not be solved by the sole acetate incorporation experiments. Incorporation of a whole series of ^{13}C-labeled glucose isotopomers into the hopanoids of the bacterium *Zymomonas mobilis* permitted to identify the origin of all carbon atoms of the isoprene units and suggested that its carbon skeleton is derived from a C_2 subunit formed from pyruvate by decarboxylation and a C_3 subunit derived from a triose phosphate derivative. Additional incorporations of glucose isotopomers with multiple ^{13}C-labeling excluded definitively the intervention of the MVA pathway for the biosynthesis of the isoprene units in the investigated bacteria. Incorporation of doubly labeled

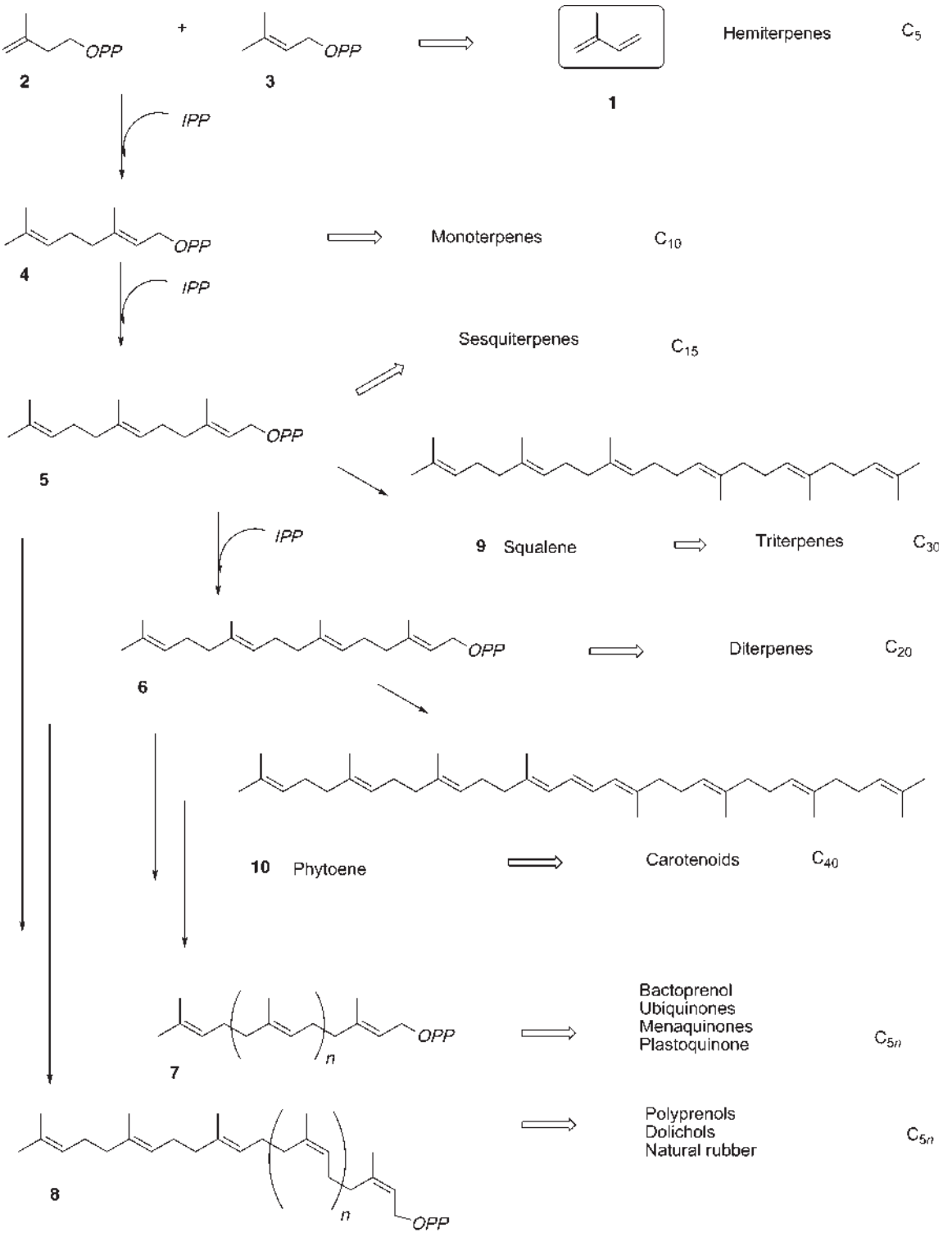


Figure 1 Biogenetic scheme for the formation of the main isoprenoid series. **1**, Isoprene; **2**, isopentenyl diphosphate; **3**, dimethylallyl diphosphate; **4**, geranyl diphosphate; **5**, farnesyl diphosphate; **6**, geranylgeranyl diphosphate; **7**, all-*trans* polyprenyl diphosphates; **8**, *cis*-polyprenyl diphosphates; **9**, squalene; **10**, phytoene. Reproduced from M. Rohmer; J. P. Vandecasteele, Natural Hydrocarbons in the Environment. In *Petroleum Microbiology*; J. P. Vandecasteele, Ed.; Editions Technip: Paris, 2008; Vol. 1, pp 37–77.

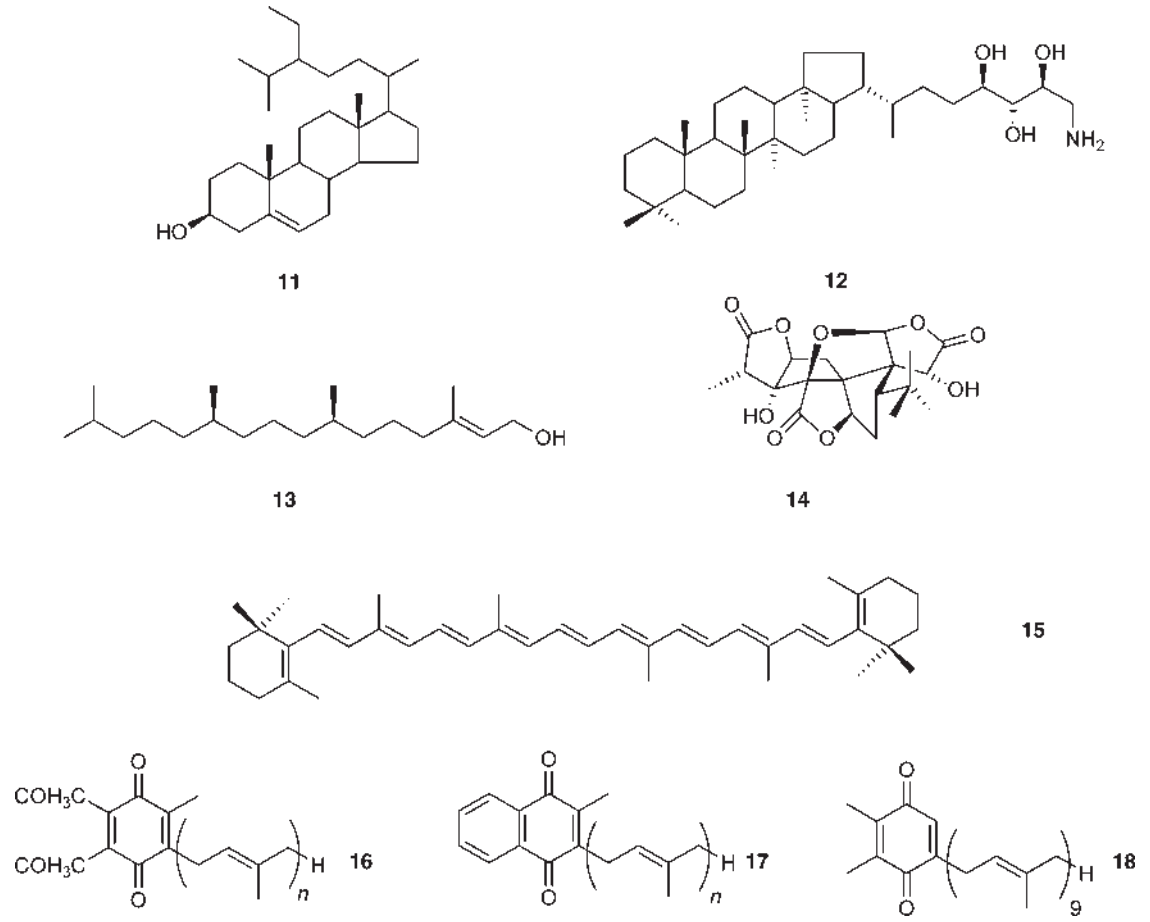


Figure 2 Key isoprenoids for the elucidation of the MEP pathway. **11**, Sitosterol; **12**, aminobacteriohopanetriol; **13**, phytol; **14**, ginkgolide; **15**, β-carotene; **16**, ubiquinone; **17**, menaquinone; **18**, plastoquinone.

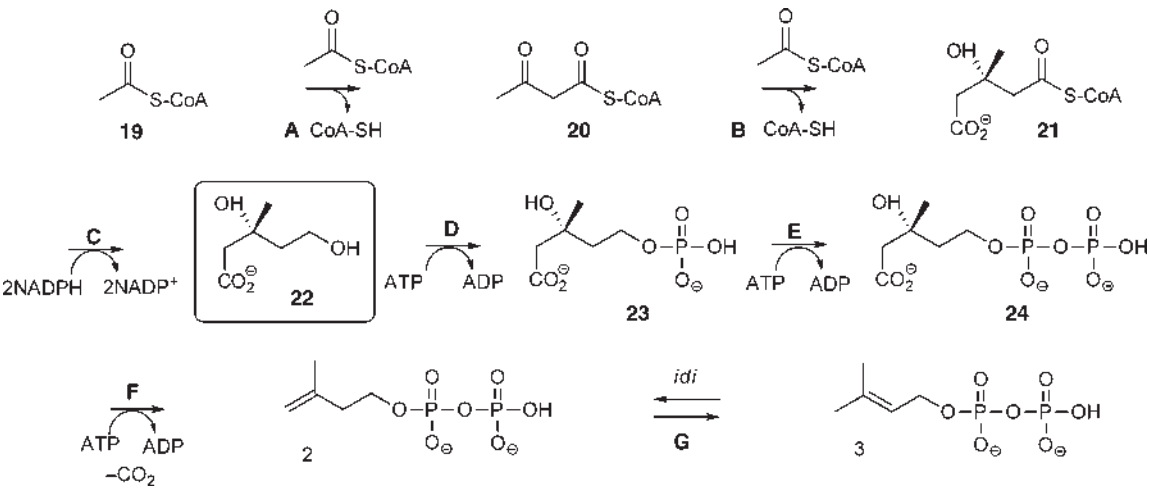


Figure 3 Mevalonate (MVA) pathway. **2**, Isopentenyl diphosphate (IPP); **3**, dimethylallyl diphosphate (DMAPP); **19**, acetyl-CoA; **20**, acetoacetyl-CoA; **21**, 3-hydroxymethylglutaryl-CoA; **22**, mevalonate (MVA); **23**, phosphoMVA; **24**, diphosphoMVA. **A**, β-Keto thiolase; **B**, hydroxymethylglutaryl-CoA synthase; **C**, hydroxymethylglutaryl-CoA reductase; **D**, MVA kinase; **E**, phosphoMVA kinase; **F**, diphosphoMVA decarboxylase; **G**, IPP isomerase.

[4,5-$^{13}C_2$]glucose into the hopanoids and the ubiquinone prenyl chain of *Methylobacterium fujisawaense* and of uniformly labeled [U-$^{13}C_6$]glucose into the hopanoids of *Z. mobilis* confirmed the abovementioned conclusions and showed in addition from the conserved $^{13}C/^{13}C$ coupling constants that an intramolecular rearrangement was involved in the formation of the isoprene units. This feature is fully incompatible with the MVA pathway.[8,9]

Two series of incorporation of ^{13}C-labeled pyruvate in the presence of natural abundance glycerol on the one hand or of nonlabeled pyruvate with ^{13}C-labeled glycerol with *E. coli* mutants, each lacking a single enzyme of the triose phosphate metabolism on the other hand showed that D-glyceraldehyde 3-phosphate **25** (GAP) together with pyruvate **24** are the precursors of the isoprene units and suggested 1-deoxy-D-xylulose (DX) 5-phosphate **26** as the first C_5 intermediate in this novel pathway (**Figure 4**). From this straight-chain intermediate, an acid-catalyzed rearrangement of an α-ketol would yield the branched isoprene skeleton. This hypothetical reaction was written by analogy with the formation of the carbon skeleton of branched chain amino acids, which is followed by the concomitant NADPH-dependent reduction of the resulting carbonyl compound, affording 2-*C*-methyl-D-erythritol 4-phosphate **27** (MEP) (**Figure 4**).[5,9]

Independently of our work on the biosynthesis of bacterial isoprenoids and nearly simultaneously, the biosynthesis of diterpenoids of the ginkgolide **14** (**Figure 2**) and bilobalide series was investigated in *Ginkgo biloba* embryos by the group of Duilio Arigoni. Incorporation of ^{13}C-labeled glucose into these plant diterpenes led to the same conclusions as those performed on the bacterial isoprenoids. The bacterial MVA-independent pathway is involved in the formation of the *G. biloba* diterpenes, whereas the MVA pathway contributes to the formation of sterols in the cytoplasm.[10,11]

Similar ^{13}C-labeled glucose-feeding experiments were carried out with other plants and allowed to generalize the same dichotomy as that observed at the level of the isoprenoid biosynthesis in the cytoplasm and the plastids of *G. biloba*: sitosterol **11** (**Figure 2**) is synthesized in the cytoplasm from barley (*Hordeum sativum*), a duck weed (*Lemna gibba*), and a carrot tissue culture (*Daucus carota*) through MVA, whereas all investigated

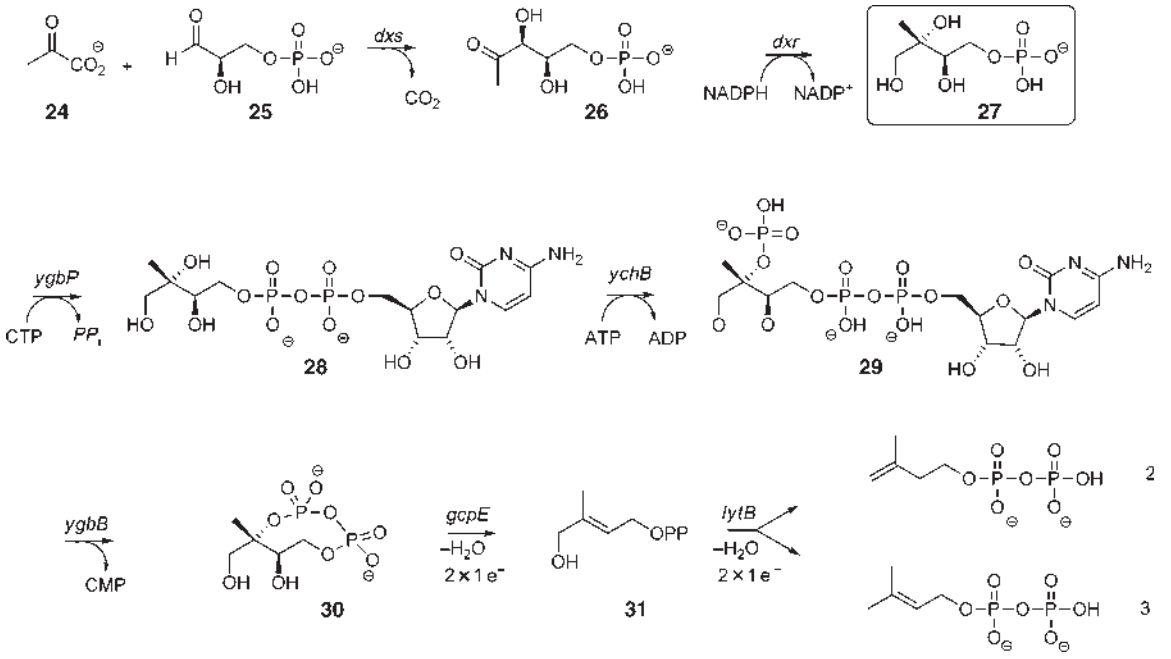


Figure 4 Methylerythritol phosphate (MEP) pathway. **2**, Isopentenyl diphosphate (IPP); **3**, dimethylallyl diphosphate (DMAPP); **24**, pyruvate; **25**, D-glyceraldehyde phosphate (GAP); **26**, 1-deoxy-D-xylulose 5-phosphate (DXP); **27**, 2-*C*-methyl-D-erythritol 4-phosphate (MEP); **28**, 4-diphosphocytidyl-2-*C*-methyl-D-erythritol (ME-CDP); **29**, 4-diphosphocytidyl-2-*C*-methyl-D-erythritol 2-phosphate; **30**, 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate (MEcDP); **31**, 4-hydroxy-3-methylbut-2-enyl diphosphate (HMBPP). *dxs*, 1-Deoxy-D-xylulose 5-phosphate synthase; *dxr*, 1-deoxy-D-xylulose 5-phosphate synthase reducto-isomerase; *ygbP*, 4-diphosphocytidyl-2-*C*-methyl-D-erythritol; *ychB*, 4-diphosphocytidyl-2-*C*-methyl-D-erythritol kinase; *ygbB*, 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate synthase; *gcpE*, 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate reductase; *lytB*, 4-hydroxy-3-methylbut-2-enyl diphosphate.

chloroplast isoprenoids (carotenoids such as β-carotene **15** and lutein, phytol **13**, and the prenyl chain of plastoquinone **9**, **Figure 2**) are derived from the MVA-independent pathway.[12] The MEP pathway was also found in other phototrophic phyla. It is the only biosynthetic pathway found in unicellular green algae. In pluricellular green algae and in all other investigated algae taxa, the same dichotomy between cytoplasmic and plastidial isoprenoids as seen in higher plants occurred.

1.13.1.3 The First Two C_5 Intermediates: 1-Deoxy-D-Xylulose 5-Phosphate and 2-*C*-Methyl-D-Erythritol 4-Phosphate

The labeling experiments that led to the discovery of the alternative MVA-independent route for isoprenoid biosynthesis have been extensively described in Chapters 3 and 14 of the first edition of this encyclopedia.[5,11] They showed that pyruvate **24** and glyceraldehyde phosphate **25** (GAP) are the starting metabolites and resulted in a hypothetical biogenetic scheme in which 1-deoxy-D-xylulose 5-phosphate (DXP) **26** and 2-*C*-methyl-D-erythritol 4-phosphate **27** (MEP) are the first two C_5 intermediates (**Figure 4**). These two compounds were already known; free 1-deoxy-D-xylulose (DX) was identified in the fermentation broth of a *Streptomyces* sp. and was known as a precursor of thiamin diphosphate and pyridoxal phosphate in *E. coli*, whereas free 2-*C*-methyl-D-erythritol (ME) was reported from many higher plants and the corresponding 2,4-cyclodiphosphate **30** (**Figure 4**) was accumulated in oxidative stress conditions by several Gram-positive bacteria.[9] Incorporation of deuterium-labeled free DX or ME into the prenyl chains of ubiquinone and menaquinone from *E. coli* were the first proofs that these carbohydrate derivatives were involved in a biosynthetic pathway toward isoprenoids.[5] It was most likely that both DX and ME were phosphorylated *in vivo* by nonspecific kinases, which were later identified as a nonspecific D-xylulose kinase in *E. coli*[13] and *Arabidopsis thaliana*[14] for DX and a sorbitol phosphotransferase in *Salmonella enterica* for ME.[15] These early findings, already described in the first issue of this encyclopedia, represented the starting point for the further identification of all genes and enzymes of the pathway.

1.13.2 Identification of the Genes and the Enzymes

1.13.2.1 1-Deoxy-D-Xylulose 5-Phosphate and 1-Deoxy-D-Xylulose 5-Phosphate Synthase

The formation of the carbon skeleton of DX/DXP from GAP **25** and pyruvate **24** could be deduced by analogy with the mechanisms of known enzymatic reactions. DXP synthase is a thiamin diphosphate **32** enzyme and catalyzes an acyloin-type condensation (**Figure 5**). (Hydroxyethylidene)thiamin diphosphate **33**, formed by the decarboxylation of pyruvate **24** as found in the reactions catalyzed by pyruvate decarboxylase or pyruvate dehydrogenase, is a nucleophile that can be added onto the carbonyl group of GAP in a manner similar to that of the second reaction step catalyzed by the transketolase of the nonoxidative pentose phosphate pathway. This hypothesis suggested a similarity in the nucleotide sequence of the transketolases and the enzyme responsible for the synthesis of the C_5 skeleton of DX/DXP. This hypothesis proved the right one. An unannotated gene with similarities with the sequences of transketolases was found in the genomes of *E. coli*[16,17] and peppermint.[18] Cloning and overexpression in *E. coli* allowed identifying the bacterial enzyme catalyzing the predicted reaction.

No direct characterization of the plant enzyme is available. Proof of activity was only obtained indirectly. Deletion of the DXP synthase in *E. coli* is rescued by insertion of the corresponding plant gene.[18] Two types of DXP synthase isogenes have been repeatedly recorded in plants. In *Medicago truncatula*, two different classes of DXP synthase-like cDNAs have been found. Both correspond to a functional enzyme as shown by heterologous expression in *E. coli*. Type I is preferentially expressed in many developing plant tissues with the exception of roots. In contrast, type II transcript levels are low in most tissues except in roots, especially when they are colonized by mycorrhizal fungi and accumulate carotenoids and carotenoid metabolites.[19,20] In the Norway spruce *Picea abies*, three genes encoding a DXP synthase are present in encoding genes with differential expression in stems of young saplings.[21] A type I *dxs* gene is constitutively expressed in bark tissues and is not affected by stress conditions (wounding or fungal attack): it encodes the enzyme responsible for the synthesis of isoprenoids that are essential for the cell life. The two other type II *dxs* genes showed enhanced transcript abundance in

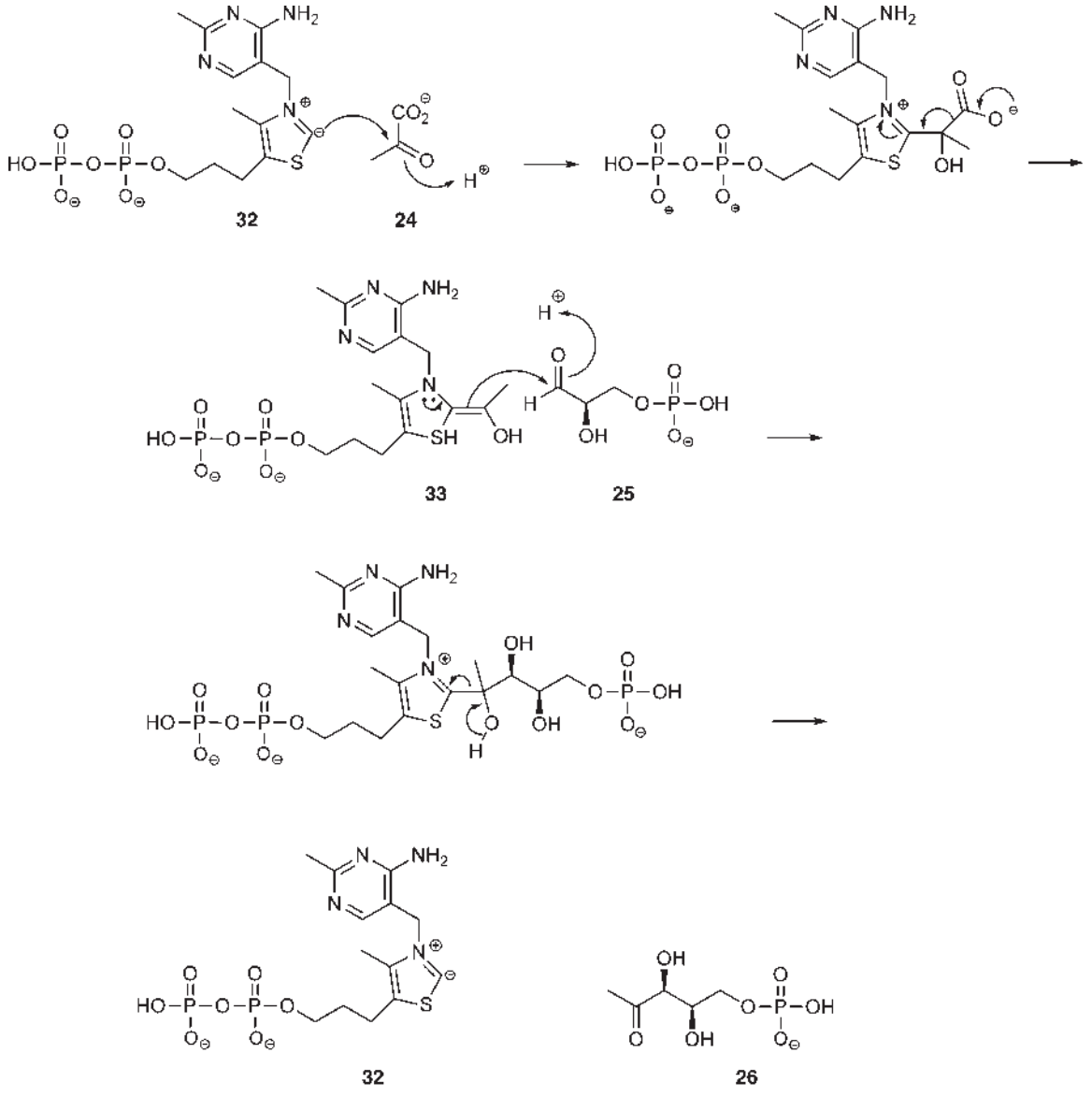


Figure 5 1-Deoxy-D-xylulose 5-phosphate synthase.

stress conditions and are probably involved in the biosynthesis of terpenoids involved in ecological interactions (mono- and diterpenes in the case of spruce, carotenoid derivatives in the case of the mycorrhizal roots).

The DXP synthase is a homodimer. Kinetic data have been obtained for the enzyme isolated from *Agrobacterium tumefaciens*,[22] *Mycobacterium tuberculosis*,[23] and *Rhodobacter capsulatus*.[24] The DXP synthase of the latter bacterium is characterized by an ordered mechanism where GAP binds before pyruvate.[24] X-ray structures of the complex with the cofactor thiamin diphosphate are available for a partially hydrolyzed DXP synthase from *E. coli* and for the intact protein from *Deinococcus radiodurans*.[25] The protein presents three domains, each showing homologies with the corresponding domain of transketolase and the E1 component of pyruvate dehydrogenase, but with different arrangements. The DXP synthase active site is located at the interface of domains I and II of a monomer, whereas it is located at the interface of the dimer in transketolase. The C-2 nucleophilic thiazolium carbon atom of thiamine diphosphate **32** (**Figure 5**) is exposed in a pocket most likely corresponding to the substrate-binding site. In addition to the previously identified

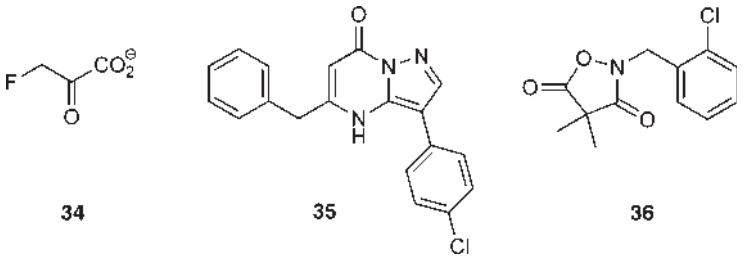


Figure 6 1-Deoxy-D-xylulose 5-phosphate synthase inhibitors.

His49 residue of the *E. coli* enzyme and participating in the proton transfer in the transketolase-catalyzed reaction,[26] several other residues putatively implied in the catalysis have been identified.

Structural variations at the level of both substrates, pyruvate **24** and GAP **25**, are accepted. According to the relative activities determined by comparison with the normal substrate pyruvate, 3-hydroxypyruvate, 2-oxobutyrate, acetaldehyde, and acetoin are poor surrogates for pyruvate.[27] GAP can be replaced by other aldose substrates: free D- and L-glyceraldehyde, glycolaldehyde, and free D-erythrose (with no identification of the reaction products) for the *E. coli* DXS,[27] D-erythrose 4-phosphate for the *M. tuberculosis* enzyme,[23] D-erythrose 4-phosphate and D-ribose 5-phosphate for the *E. coli* enzyme,[28] yielding two substrates 1-deoxy-D-fructose 6-phosphate and 1-deoxy-D-sedoheptulose 7-phosphate for the latter.

Only few inhibitors have been reported. As expected, 3-fluoropyruvate **34** (**Figure 6**) is an inhibitor of the *E. coli* ($IC_{50} = 80\ \mu mol\ l^{-1}$) and *Pseudomonas aeruginosa* ($IC_{50} = 400\ \mu mol\ l^{-1}$) enzymes.[29] It was supposed to bind covalently to the active site of the enzyme in a similar way as it does for the pyruvate dehydrogenase E1 component, but detailed kinetic analysis performed on the *R. capsulatus* enzyme showed that in the presence of GAP or D-glyceraldehyde fluoropyruvate is a competitive inhibitor with K_i values of 1.0 and 6.1 $mmol\ l^{-1}$, respectively.[24] Neither DL-α-glycerophosphate, β-glycerophosphate, 3-phospho-D-glyceric acid, 2,3-diphospho-D-glyceric acid (as possible mimics of GAP), nor phosphonoacetohydroxamate and phosphonopropionohydroxamate (as possible transition state analogues) were potent inhibitors of the *R. capsulatus* DXP synthase.[24] Structural modifications of a known transketolase inhibitor led to compounds, for example, **35** (**Figure 6**), showing modest inhibitory activity on the *M. tuberculosis* DXP synthase (with $IC_{50} = \sim 50\ \mu mol\ l^{-1}$ for the most efficient compounds), but constituting possible interesting leads for further investigations.[30]

3-Oxoclomazone **36** (**Figure 6**), a metabolite resulting from the *in vivo* oxidation of the herbicide clomazone by higher plants, was reported to have an inhibitory effect on the DXP synthase of phototrophic organisms. 3-Oxoclomazone inhibits the DXP synthase of the green alga *Chlamydomonas reinhardtii* ($IC_{50} = \sim 0.1\ mmol\ l^{-1}$)[31] and blocks isoprene **1** (**Figure 1**) emission by several plant systems and lowers significantly the phytol-containing chlorophyll and carotenoid levels during greening of etiolated barley leaves,[32] whereas clomazone itself is inactive.

1.13.2.2 2-*C*-Methyl-D-Erythritol 4-Phosphate and 2-*C*-methyl-D-Erythritol 4-Phosphate Reducto-Isomerase

1.13.2.2.1 dxr *gene discovery*

The observation that free ME is incorporated into the prenyl chain of ubiquinone **16** (**Figure 2**) and menaquinone **17** in *E. coli* opened the way for the discovery of the second gene implied in the alternative MVA-independent pathway. The search for *E. coli* mutants requiring ME for growth pointed out the damaged gene that encodes the enzyme converting DXP into MEP.[33,34] As hypothesized in the first version of the biogenetic scheme, the same enzyme, the DXP reducto-isomerase (DXR), catalyzed the rearrangement of the straight-chain DXP skeleton into the putative intermediate 2-*C*-methyl-D-erythrose 4-phosphate **37** (**Figure 7**) and the concomitant NADPH-dependent reduction of the resulting aldehyde requiring a divalent cation (e.g., Mg^{2+}, Mn^{2+}) (**Figure 7**). DXR is the most studied enzyme of the pathway. Its reaction product is MEP **27**, that is, the first intermediate showing the branched isoprene skeleton and that can be thus considered as a hemiterpene. In addition, no other role is known for MEP than that of an isoprenoid precursor, whereas DXP is also the precursor of thiamin diphosphate[35] and pyridoxal phosphate,[36] at least in *E. coli*, and free DX of

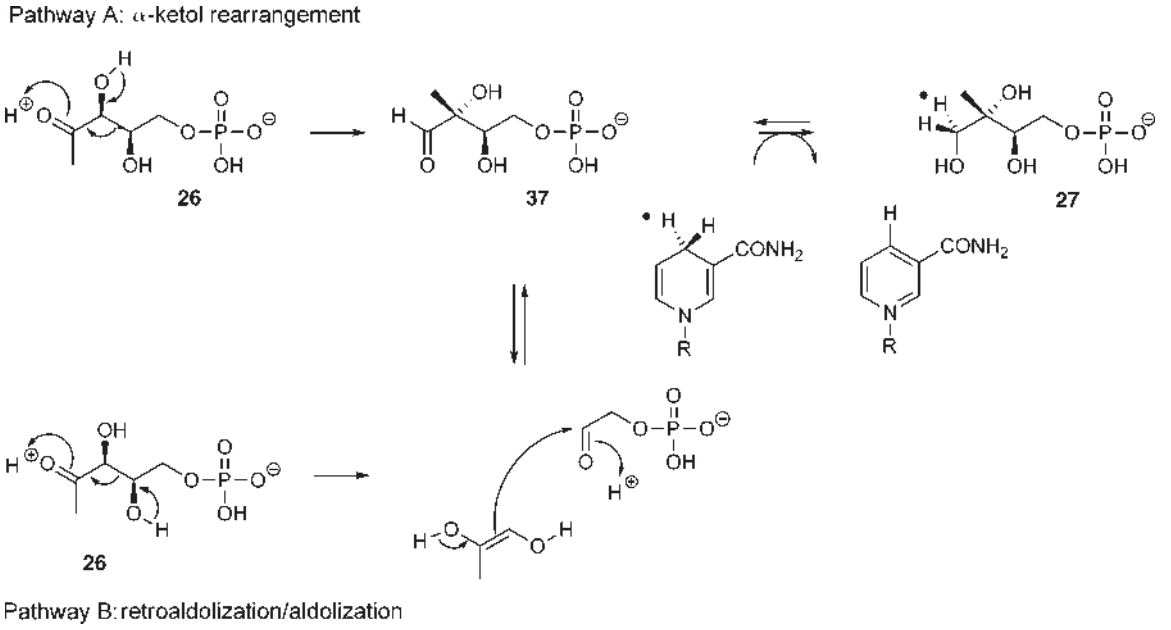


Figure 7 1-Deoxy-D-xylulose 5-phosphate reducto-isomerase.

laurencione (5-hydroxypentane-2,3-dione) in *E. coli* and *Klebsiella planticola*,[37] suggesting that DXR catalyzes the committed step of the pathway, which was accordingly named after MEP **27**, the DXR reaction product.

1.13.2.2.2 Reaction mechanism

DXR has been cloned from several organisms. The DXR proteins from *E. coli* and from *Synechocystis* sp. PCC 6803 are class B dehydrogenases, transferring stereospecifically the H_{Si} from C-4 of the cofactor in the H_{Re} position of C-1 of methylerythrose **37** (**Figure 7**).[38,39] The reaction catalyzed by the *E. coli* DXR is reversible, the equilibrium being largely displaced in favor of MEP ($K_{eq} = 45\ \mu mol\ l^{-1}$ for DXP and MEP at $150\ \mu mol\ l^{-1}$ NADPH, $k_{cat} = 116\ s^{-1}$, $K_m^{DXP} = 15\ \mu mol\ l^{-1}$, $K_m^{NADPH} = 0.5\ \mu mol\ l^{-1}$).[40] The putative intermediate aldehyde **37** (**Figure 7**) is converted into MEP **27** in the presence of NADPH and into DXP **26** in the presence of $NADP^+$.[41] Steady-state kinetic analysis of the recombinant DXR from *E. coli* was consistent with an ordered mechanism where the cofactor NADPH binds before the substrate DXP.[40] Similar kinetic data were obtained for the native or the His-tagged DXR isolated from other bacteria: *M. tuberculosis*,[42–44] *Streptomyces coelicolor*,[45] *Synechocystis* sp. PCC6803,[46] *Z. mobilis*.[47] The DXR of *E. coli* and *Synechocystis* PCC6803 have been characterized as class B dehydrogenases. They transfer stereospecifically the H_{Si} of the cofactor on the *re* face of the intermediate aldehyde **37** (**Figure 7**).[38,39,48]

The mechanism of the rearrangement, initially written as an acid-catalyzed intramolecular rearrangement of an α-ketol (**Figure 7**, pathway A), has been only recently fully disclosed. Incubation of 1-fluoroDXP **71** (**Figure 9**), a good substrate analogue, afforded first indications in favor of a retroaldol/aldol rearrangement (**Figure 7**, pathway B). A primary deuterium isotope effect was observed in single turnover incubations of 1-fluoroDXP **71** in the presence of (4*S*)-[4-^{2}H]NADPH ($k_H/k_D = 1.34 \pm 0.01$). No isotope effect was observed upon incubation of 1-fluoroDXP **71** with the other cofactor isotopomer, (4*R*)-[4-^{2}H]NADPH,[49] indicating that the rearrangement preceding the hydride transfer is more rate limiting for the conversion of DXP **26** than for the conversion of 1-fluoroDXP **71** and is enhanced by the replacement of a hydrogen by a fluorine. No burst kinetics was observed with either substrate, indicating that the release of the reaction product is not rate limiting. All these data suggest that a positive charge does not develop at C-2 of DXP and is in accordance with the retroaldol/aldol reaction. Secondary kinetic isotope effects observed upon incubation of [3-^{2}H]- and [4-^{2}H]DXP were in favor of an alternative retroaldol/aldol (**Figure 7**, pathway B) reaction and permitted to definitively reject the α-ketol rearrangement mechanism (**Figure 7**, pathway A).[50,51]

The plant DXR from *A. thaliana* has been obtained in a pseudo-mature form lacking the putative plastid-targeting sequences from a recombinant *E. coli* strain and characterized ($k_{cat} = 4.4\,s^{-1}$, $K_m^{DXP} = 132\,\mu mol\,l^{-1}$, $K_m^{NADPH} = 30\,\mu mol\,l^{-1}$).[52] Similar to the bacterial enzyme, the *Arabidopsis* DXR-catalyzed reaction is reversible, NADH can replace NADPH, but at a lower rate (14% as compared to NADPH), and the plant DXR is inhibited by fosmidomycin **38** (**Figure 8**) ($K_i = 85\,nmol\,l^{-1}$). The *dxr* gene was cloned from many plants and characterized by heterologous expression of the corresponding cDNA in *E. coli*.[53]

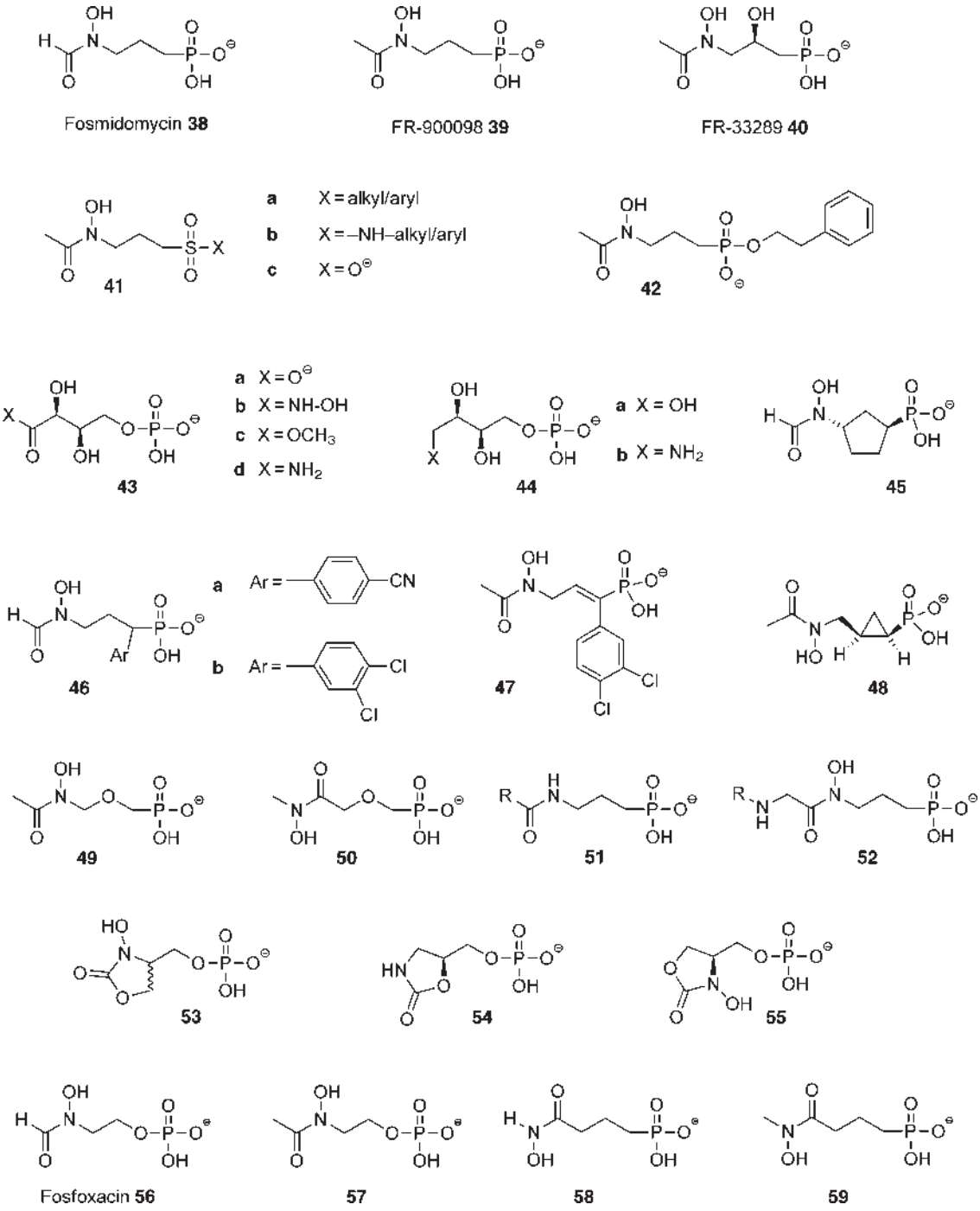


Figure 8 1-Deoxy-D-xylulose 5-phosphate reducto-isomerase inhibitors.

1.13.2.2.3 Enzyme structure

Several DXR X-ray structures of the *E. coli* enzyme have been obtained. The enzyme forms a homodimer in solution and in the crystal. A crystal structure with complexed NADPH and a sulfate ion disclosed that the enzyme is characterized by a conformational mobility. A flexible loop covers the substrate-binding site and the presence of an extra domain that is absent in other NADPH-binding enzymes.[54] The V-shaped monomer is characterized by a large N-terminal dinucleotide binding domain with an α/β topology with a parallel β-sheet and seven α-helices linked through a connective domain with five helices to a smaller C-terminal four-bundle helix domain.[54,55] The connective domain is involved in the dimerization and possesses most of the active site functional groups.[56] The crystal structure with the Mn^{2+} complex and the inhibitor fosmidomycin **38** (**Figure 8**) shows that the hydroxamate group provides two ligands to the Mn^{2+} ion in a distorted octahedral coordination sphere, whereas the phosphonate moiety is linked by several hydrogen bonds in a specific pocket. DXP, the substrate of the enzyme, can be superimposed to the inhibitor.[57] The structure of the ternary complex DXR/fosmidomycin/NADPH revealed the considerable conformational changes upon binding of the inhibitor and allowed to propose the DXR ternary complex with the cofactor and the normal substrate DXP **28** (**Figure 4**).[55] The *M. tuberculosis*[44] and *Z. mobilis*[58] enzymes showed a similar structure as those of *E. coli*. The *M. tuberculosis* DXR structure was intermediate between the open apo form and the closed holo form, pointing out significant conformational changes occurring during the catalytic process.[44]

Directed mutagenesis showed that Glu231 plays an important role in the reaction catalyzed by the *E. coli* enzyme, and that His153, His209, and His257 are involved in DXP binding.[59] In the *Synechocystis* enzyme residues Asp152, Thr153, Glu154, His155, Met206, and Glu223 were targeted.[60] Major effects on the catalysis were induced by alteration of the three acidic residues, variable effects on binding and catalysis were found for changes to Thr153 and Met206, and a His155Ala mutation had minimal effect on the kinetic parameters. A highly conserved tryptophane residue in the flexible loop was identified as potentially limiting the use of the higher DXP homologue, 1,2-dideoxy-D-hexulosephosphate **72** (**Figure 10**). This hypothesis was successfully tested by incubating the Trp204Phe mutant, which was found to use the hexose DXP higher homologue as substrate.[61]

1.13.2.2.4 Deoxyxylulose phosphate reducto-isomerase inhibitors and substrate analogues

1.13.2.2.4(i) Fosmidomycin and fosmidomycin-inspired DXR inhibitors

Fosmidomycin **38** (**Figure 8**) was described to have a potent antibacterial activity, inhibiting probably the biosynthesis of ubiquinone by an unknown mechanism.[62] It is active against *E. coli* and *Bacillus subtilis* possessing the MEP pathway, but has no effect on the growth of *Staphylococcus aureus*, which has the MVA pathway. According to this antibacterial spectrum, fosmidomycin was expected to be a specific inhibitor of a MEP pathway enzyme, and its effect on MEP pathway enzymes was tested. Fosmidomycin **38** ($K_i = 38\,\text{nmol}\,l^{-1}$), and its structural analogues FR-900098 **39** and FR-33289 **40** (**Figure 8**), proved to be potent inhibitors of the *E. coli* DXR.[63] Fosmidomycin **38** presents a slow tight-binding behavior.[64]

Fosmidomycin **38** served as lead for the conception of a whole series of possible DXR inhibitors. Attempts for the design of inhibitors were oriented toward an increased inhibitory activity, an improved bioavailability, diminishing the fast clearance, to overcome the bacterial resistance, and also to afford more information on the DXR substrate-binding site.

Replacement of the negatively charged phosphonate moiety by a neutral sulfone **41a** (**Figure 8**) or sulfonamide **41b** group results in a complete loss of activity. Some activity is retained when a negative charge is present as in the sulfonate **41c** or the phosphonate **42** monoester, and could even be improved when a hydrophobic group is present in the phosphonate monoester moiety as in the mono-phenylethyl ester of FR900098 **42** (**Figure 8**), indicating that a fairly large hydrophobic pocket is close to the phosphonate-binding site.[65] Shortening the length of the propylene spacer leads to the loss of the inhibitory activity.[66] DXP analogues where the methyl group is replaced by a hydroxy **43a**, hydroxylamino **43b**, methoxy **43c** or amino **43d** group, or the acetyl moiety by a hydroxymethyl **44a** or an aminomethyl **44b** group (**Figure 8**) have been tested for their DXR inhibitory activity.[67] The hydroxamic acid **43b** and the amino derivatives **44b** present no inhibitory activity, whereas the carboxylate

43a, the methyl ester **43c**, the amide **43d**, and the alcohol **44a** (**Figure 8**) show some limited inhibitory activity with IC_{50} values between 0.25 and 1 mmol l^{-1}. Structural variations were investigated around the fosmidomycin model. α-Aryl-substituted and conformationally restricted analogues are considered as promising antimalarial compounds ($IC_{50} = 0.20\ \mu mol\ l^{-1}$ for **45**, $0.12\ \mu mol\ l^{-1}$ for **46a**, $0.059\ \mu mol\ l^{-1}$ for **46b** on the *E. coli* DXR, **Figure 8**),[68–70] whereas α,β-unsaturated α-aryl-substituted fosmidomycin analogues such as **47** ($IC_{50} = 5.5\ \mu mol\ l^{-1}$ on the *E. coli* DXR) (**Figure 8**) exhibit a much lower affinity for DXR than fosmidomycin.[71] Some cyclopropyl analogues, for example, **48** ($IC_{50} = 0.48\ \mu mol\ l^{-1}$), are potent inhibitors for *Plasmodium falciparum*.[72] The β- and γ-oxa isosteres of fosmidomycin and FR900098 have been synthesized.[73] The β-oxa derivative of FR900098 **49** (**Figure 8**) (IC_{50} against *E. coli* and *P. falciparum* 3D7 DXR 87 and 360 nmol l^{-1}, respectively) is as active as the parent compound. Replacing the retro-hydroxamate by a hydroxamate as in **50** results in an even increased antiplasmodial activity (IC_{50} against *E. coli* and *P. falciparum* 3D7 DXR 72 and 240 nmol l^{-1}, respectively). The γ-oxa variation leads to less active compounds. A series of phosphonic acids with the same propylene spacer as fosmidomycin and FR900098 but with different substitution patterns on the amide **51** and hydroxamate **52** carbonyl group presents inhibitory activities against the *E. coli* DXR in the 1 to >30 $\mu mol\ l^{-1}$ range or *P. falciparum* DXR in the 0.4–20 $\mu mol\ l^{-1}$ range.[74] Series of fosmidomycin fragments keeping the phosphonate moiety with variations in the spacer length and replacing the retro-hydroxamate by an amino, acetamido, or 4- and 5-substituted oxazolidin-2-ones **53–55** (**Figure 8**) afford compounds with low inhibitory activity against the *E. coli* DXR in the millimolar range (e.g., $K_i = 0.75$ mmol l^{-1} for racemic **53** on the *E. coli* enzyme).[64] The slow tight binding of fosmidomycin, with probably some cooperativity, is accompanied by conformational changes of the protein and reorganization of the active site, making the rational design of DXR inhibitors not obvious.[64]

Fosfoxacin **56** (**Figure 8**), the phosphate analogue of fosmidomycin isolated from the fermentation broth of *Pseudomonas fluorescens*, and its acetyl homologue **57** are even more potent inhibitors of the *Synechocystis* sp. PCC6803 DXR.[75] Replacement of the phosphonate moiety by a carboxylate or a sulfamate or replacement of the *N*-hydroxy, *N*-formyl group by an *N*-methyl, *N*-formyl, or an *N*-methyl, *N*-acetyl group results in no significant inhibitory activity.[75] Only the replacement of the *N*-hydroxy, *N*-formyl group by the reverse hydroxamate **58** or its *N*-methyl homologue **59** (**Figure 8**) showed a strong competitive slow-binding inhibition activity on the *E. coli* DXR ($K_i = 169$ and 54 nmol l^{-1}, respectively) similar to that induced by fosmidomycin ($K_i = 40$ nmol l^{-1})[76] or the *Synechocystis* DXR ($K_i = 4\ \mu mol\ l^{-1}$).[75]

Prodrugs of fosmidomycin and FR900098 have been tested *in vivo*, essentially for their antimalarial activity. They were obtained by modification of the phosphonate moiety that was masked as phosphoester derivatives. Diaryl and acyloxyalkyl esters of FR900098 (**60**, **61**, **Figure 9**) with increased activity against the rodent

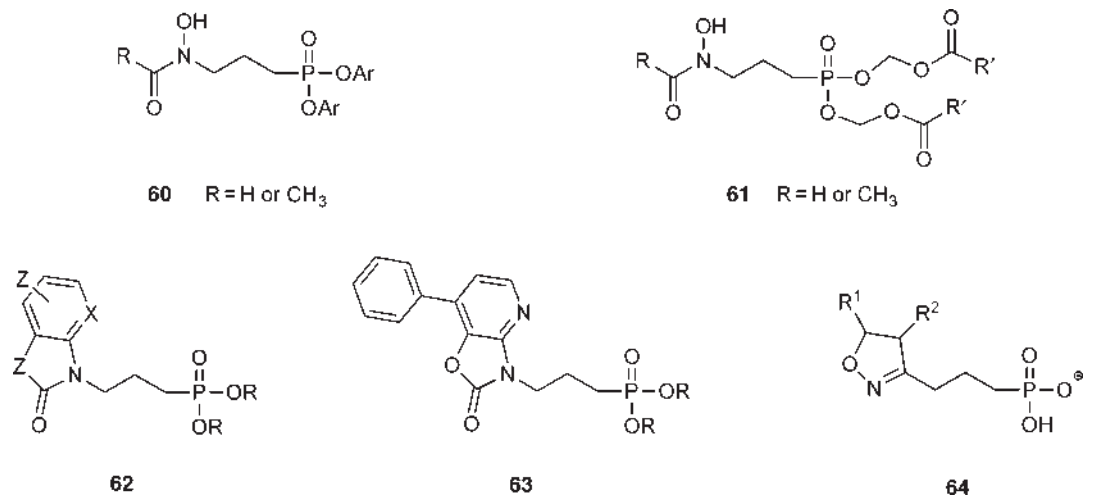


Figure 9 Prodrugs of 1-deoxy-D-xylulose 5-phosphate reducto-isomerase inhibitors.

malaria parasite *Plasmodium vinckei* were thus obtained.[77–79] Bis(pivaloyloxymethyl) esters **61** (**Figure 9**) proved particularly efficient.[80,81] The prodrug of the α-substituted 3,4-dichlorobenzyl derivative of fosmidomycin **46b** (**Figure 8**) is about twice as active against *P. falciparum* as the corresponding prodrug of fosmidomycin, but less active than the FR900098 prodrug.[80] Similarly, the same type of bis(pivaloyloxymethyl)ester prodrug of the phosphonate **50** (**Figure 8**) is also twice as active against the same parasite as the corresponding prodrug of FR900098.[73]

Fosmidomycin **38** (**Figure 8**) affects also isoprenoid biosynthesis in plants, blocking the MEP pathway. It inhibited isoprene **1** (**Figure 1**) emission in leaves from *Populus nigra*, *Platanus* × *acerifolia* and *Chelidonium majus*, lycopene, a carotenoid, biosynthesis in ripening tomatoes, chlorophyll and carotenoid biosynthesis in primary barley leaves and *L. gibba* axenic cultures.[82] In chromo- and chloroplasts prepared from several higher plants, fosmidomycin inhibited in the presence of ATP the conversion of DXP into geranylgeraniol **6** (**Figure 1**) and β-carotene **15** (**Figure 2**), but not the incorporation of MEP in accordance with an inhibition of the plant DXR.[83] Fosmidomycin is, however, much less active on the plant DXR than on the bacterial enzymes with IC_{50} significantly higher than that reported for the *E. coli* enzyme (0.28 and 0.70 $\mu mol\,l^{-1}$ for the *A. thaliana* and the barley enzymes at 0.25 and 0.50 $mmol\,l^{-1}$ DXP concentration, respectively).[31]

Fosmidomycin analogues with a benzoxazolone, benzoxazolomethine, or an oxazolopyridine moiety (e.g., **62**, **Figure 9**) did not affect the growth of *Catharanthus roseus* cell cultures at the tested concentration (125 $\mu mol\,l^{-1}$), but induced a significant decrease (45–85%) of the concentration of ajmalicin, an alkaloid containing a monoterpene geranyl diphosphate **5** (**Figure 1**) derived moiety that is synthesized through MEP pathway.[84] Substitution at C-4 on the pyridine ring even increased in some cases the inhibitory activity on ajmalicin biosynthesis that was even nearly fully blocked at a 100 $\mu mol\,l^{-1}$ concentration in the case of a phenyl substituent as in the case of **63** (**Figure 9**), which is even more effective than fosmidomycin.[85] 4,5-Dihydroxyisooxazole fosmidomycin analogues **64** (**Figure 9**) had no effect on ajmalicin biosynthesis, but the corresponding phosphonate diethyl esters and the diethyl ester of fosmidomycin unexpectedly induced an increased accumulation of the alkaloid by an unknown mechanism.[86]

1.13.2.2.4(ii) NADPH and substrate analogues The cofactor analogue dihydro-NADPH **79** (**Figure 10**) provided the *E. coli* DXR a competitive inhibition pattern against NADPH and a noncompetitive inhibition pattern against DXP, and fosmidomycin is an uncompetitive inhibitor against NADPH and showed a slow, tight-binding competitive inhibition pattern against DXP.[41]

The phosphonate isoster of DXP **65** (**Figure 10**) is a substrate of the DXR from *E. coli* ($k_{cat} = 74\,min^{-1}$, $K_m = 120\,\mu mol\,l^{-1}$), yielding the phosphonate isoster of MEP **66**,[87,88] and of the *Synechocystis* sp. PCC6803 DXR ($K_m = 690\,\mu mol\,l^{-1}$).[89] DXP analogues lacking each a hydroxyl group have been tested for investigating the reaction mechanism. The C-3 hydroxy being involved in the α-ketol rearrangement, whereas the C-4 hydroxy group is implied in the retroaldol/aldol reaction (**Figure 7**, pathway B). Neither (3*S*)-3-hydroxypentan-2-one 5-phosphate **67** (4-deoxy-DXP) nor (4*S*)-4-hydroxypentan-2-one 5-phosphate **68** (3-deoxy-DXP) is converted by the *E. coli* DXR but behaves as reversible mixed-type inhibitor with respective 120 and 800 $\mu mol\,l^{-1}$ K_i values;[40] they also behave as weak competitive inhibitors of the *Synechocystis* sp. PCC6803 DXR with 30 and 150 $\mu mol\,l^{-1}$ K_i values.[89] 1,1,1-TrifluoroDXP **69**, 1,1-difluoroDXP **70**, 1-fluoroDXP **71**, and 1,2-dideoxy-D-hexulose 6-phosphate **72**, the DXP analogue where the methyl group is replaced by an ethyl group, are all poor inhibitors, probably because of the enhanced steric bulk at C-1. 1-FluoroDXP **71** is even an acceptable substrate of the enzyme ($k_{cat} = 37\,s^{-1}$, $K_m = 227\,\mu mol\,l^{-1}$).[49] 3-Fluoro- **73** ($K_i = 444\,\mu mol\,l^{-1}$) and 4-fluoroDXP **74** ($K_i = 733\,\mu mol\,l^{-1}$) are noncompetitive inhibitors with respect to DXP.[90] The diastereomeric 2:1 mixture of 1,1,1-trifluoro-D-xylitol 5-phosphate and 1,1,1-trifluoro-D-lyxitol 5-phosphate **75** behaves like a weak reversible competitive inhibitor ($K_i = 360\,\mu mol\,l^{-1}$).[91]

5-FluoroMEP **76** (**Figure 10**) was synthesized as possible mechanism-based inactivator in the case of the retroaldol/aldol rearrangement mechanisms. No irreversible inhibition was observed. It is a weak competitive inhibitor, probably because it cannot be oxidized to the intermediate aldehyde.[92] The 1-methyl homologue **77** (**Figure 10**) of the intermediate aldehyde **37** (**Figure 7**) had no significant inhibitory effect on the *E. coli* DXR.[74] 1,2-Dideoxy-L-*threo*-3-hexulose 6-phosphate **72** (**Figure 10**) ($K_i = 630\,\mu mol\,l^{-1}$, 1-methyl DXP), 4-*epi*-DXP **78** (**Figure 10**) ($K_i = 180\,\mu mol\,l^{-1}$), and

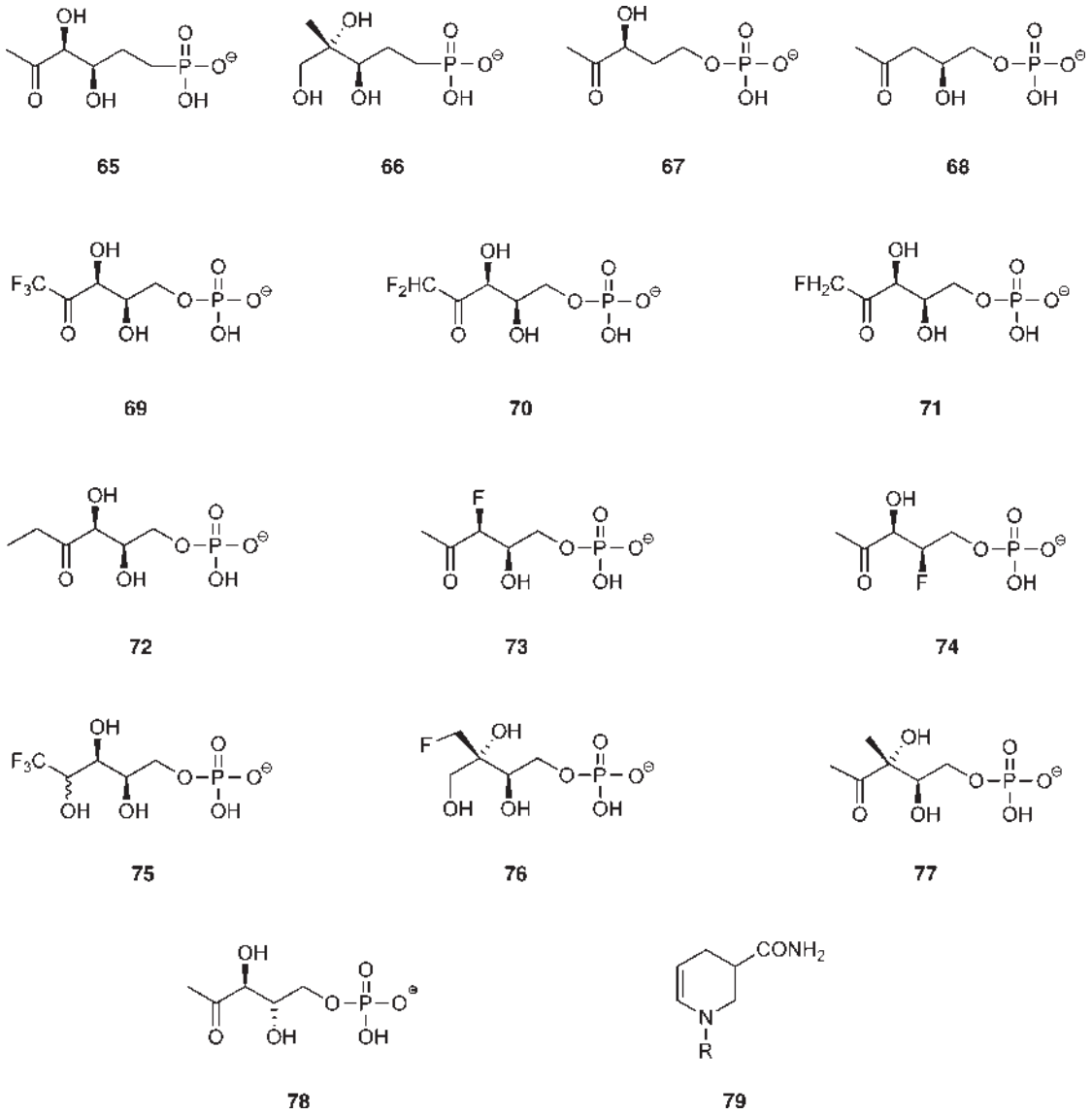


Figure 10 Substrate and product analogues of 1-deoxy-D-xylulose 5-phosphate reducto-isomerase.

2*S*,3*R*-dihydroxybutyramide 4-phosphate **43d** (**Figure 8**) ($K_i = 90\,\mu mol\,l^{-1}$) behaved as weak competitive inhibitors of the *Synechocystis* sp. PCC6803 DXR.[89]

1.13.2.3 From 2-*C*-Methyl-D-Erythritol 4-Phosphate to 2-*C*-Methyl-D-Erythritol 2,4-Cyclodiphosphate

The ^{13}C-labeling experiments allowed to determine the origin of the carbon atoms and to identify the first two intermediates, pyruvate **24** and GAP **25** (**Figure 4**). From these data, it was possible to postulate the structures of the first two C_5 intermediates, DXP **26** and MEP **27**, and to find the genes encoding the first two enzymes of the novel pathway. The next steps were identified at the level of the genes and the enzymes by a combination of labeling experiments and molecular biology methods. A unified nomenclature has been proposed for the MEP pathway enzymes, genes, and mutants in *E. coli* and *A. thaliana* and will mainly be followed in this chapter.[93]

1.13.2.3.1 4-Diphosphocytidyl-2-C-methyl-D-erythritol and 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase

Incubation of ^{3}H-labeled MEP with a crude cell-free system from *E. coli* in the presence of ATP afforded by TLC monitoring a radiolabeled metabolite, which was supposed to be an intermediate in the pathway.[94] Partial purification resulted in a severe loss of enzyme activity, suggesting that a low-molecular-weight metabolite was lost during the purification procedure. Testing a series of nucleotide 5′-triphosphates showed that cytosine triphosphate (CTP) was quite efficient for this conversion of MEP. In addition, it was found that radioactivity was incorporated into the unknown metabolites from [α-^{32}P]CTP and not from [γ-^{32}P]CTP. First details on its structure were obtained on a sample obtained with partially purified enzyme by ^{31}P NMR spectroscopy showing two signals ($\delta = 7.2$ and 7.8 ppm) with a ^{31}P/^{31}P-coupling constant ($^{2}\mathcal{J} = 20$ Hz) and by ^{13}C NMR showing four signals with ^{13}C/^{31}P-coupling constants ($\mathcal{J} = 5$–9 Hz). The complete structure was obtained by two-dimensional correlated NMR spectroscopy, revealing the spin networks of ME and cytidine. A 4-diphosphocytidyl-2-*C*-methyl-D-erythritol (CDP–ME) structure was consequently assigned. A search in the databases for a gene encoding an enzyme catalyzing the coupling of a polyol phosphate with a nucleotide diphosphate resulted in the identification of the recently discovered *Haemophilus influenzae acs1* gene,[95] encoding a bifunctional enzyme with ribulose 5-phosphate reductase/CDP–ribitol pyrophosphorylase activities. The 5′moiety of this *acs1 H. influenzae* gene is similar to the unannotated *ygbP* gene, which is also found in *E. coli* and other bacteria possessing the MEP pathway. Cloning and expression of the *ygbP E. coli* gene showed that the recombinant protein catalyzes the formation of CDP-ME **28** from CTP and MEP at high rate (**Figure 4**). The enzyme is quite substrate specific, accepting only MEP ($K_M = 3\ \mu mol\ l^{-1}$) and CTP ($K_m = 131\ \mu mol\ l^{-1}$) as substrates. GTP can replace to CTP some extent. Neither ribitol 5-phosphate, nor erythritol 4-phosphate, ATP, UTP, and ITP were substrates. The enzyme requires a divalent cation such as Mg^{2+}, Mn^{2+}, and Co^{2+}. Incorporation of radiolabeled CDP–ME **28** into carotenoids of *Capsicum annuum* chromoplasts represented the first proof that this ME derivative is an intermediate in the MEP pathway. An unannotated gene with similarities with the *E. coli ygbP* gene was also found in *A. thaliana*. The cognate protein catalyzed also the formation of CDP–ME **28** from CTP ($K_M = 114\ \mu mol\ l^{-1}$) and MEP ($K_m = 500\ \mu mol\ l^{-1}$) and was characterized by an N-terminal plastid-targeting sequence.[96] More recently, the enzymes from two *M. tuberculosis* strains were also characterized with K_m values of 59/43 and 53/92 $\mu mol\ l^{-1}$, for MEP and CTP, respectively, and an activity over a broad pH range from 6.0 to 9.0.[97,98] The enzyme is thermally instable above 50 °C, the circular dichroism spectra revealing that the tertiary structure is altered at higher temperature.[98]

Another, more general, approach led to the identification of the CDP–ME synthase.[99] Since isoprenoid biosynthesis is essential for bacteria, any loss of function of MEP pathway enzyme is lethal. For the search of MEP pathway genes, *E. coli* strains capable of utilizing MVA, which is not a normal metabolite for this bacterium, were constructed. For this the MVA kinase, the MVA phosphate kinase, and the MVA diphosphate decarboxylase genes from *Streptomyces* CL190 were introduced into *E. coli*. This transformant is thus capable of utilizing exogenous MVA added to the culture medium for the biosynthesis of IPP. Using this transformant, mutants with an absolute requirement for MVA were prepared. This indicated that they had a defect in an MEP pathway gene and resulted in the identification of the missing MEP pathway genes. Overexpression of these genes resulted in the production and the purification of the cognate enzymes in quantities that were sufficient for their characterization. Among theses gene products was the CDP–ME synthase. The reaction product, CDP–ME **28** (**Figure 4**) was directly identified by ^{1}H-, ^{13}C-, and ^{31}P-NMR spectroscopy.

Structures of the *E. coli* CDP–ME synthase were obtained with high resolution for highly ordered monoclinic form of the apo enzyme and the enzyme complexed with both CTP/Mg^{2+} and CDP–ME/Mg^{2+} [100] and with a lower resolution for the tetragonal crystal form in the absence of ligands.[101] Pulse-chase experiments established that the reaction is an ordered sequential mechanism with first binding of the CTP cofactor.[102] On the basis of the first three-dimensional structures, putative active site residues were selected for directed mutagenesis. Lys27 and Lys273 are essential for catalytic activity and proposed for the stabilization of a pentacoordinate phosphate transition state resulting from the addition of the MEP phosphate to the CTP α-phosphate. Thr140, Arg109, Asp106, and Thr165 are involved in the binding and the correct orientation of the MEP substrate.

Genome sequence comparisons showed that the *E. coli ygbP* gene is part of a small operon containing also the unannotated *ychB* and *ygbB* genes. A comparative analysis of all published complete genomes indicated that these three genes always accompany the known MEP pathway genes and are also present in other bacteria possessing

the MEP pathway.[94,103] Moreover, several microorganisms possess genes encoding putative fusion proteins with *ygbP* and *ygbB* domains, suggesting that both the YgbP and YgbB enzymes belong to the MEP pathway. To test this hypothesis, the *E. coli ychB* and *ygbB* genes were expressed in a homologous host.

1.13.2.3.2 4-D-Diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate and 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase

The recombinant YchB protein from *E. coli* catalyzes the ATP-depending phosphorylation of the tertiary C-2 hydroxy group of CDP–ME **28** yielding 2-phospho CDP–ME **29** identified by NMR spectroscopy after incubation of various ^{13}C-labeled substrate samples (**Figure 4**).[104] First proof that 2-phospho CDP–ME **29** is a MEP pathway intermediate was obtained with an already mentioned plant system: a ^{14}C-radiolabeled sample was efficiently incorporated into the carotenoids of chromoplasts from *C. annuum*.[104] Mutation performed on an *E. coli* transformant capable of utilizing MVA as a precursor for isoprene units allowed also identifying the essential *ychB* gene encoding the enzyme phosphorylating CDP–ME **28**.[103]

The putative recombinant catalytic domains (residues 81–401) of the corresponding YchB protein from tomato expressed in *E. coli* were purified to apparent homogeneity and also catalyzed the conversion of CDP–ME **28** yielding its 4-phospho derivative **29**.[105] Both the bacterial and the plant enzymes require a divalent cation, preferably Mg^{2+}.

Crystal structures of the CDP–ME kinase were obtained. The 2 Å resolution structure of the *E. coli* enzyme as a ternary complex with the substrate and a nonhydrolyzable ATP analogue (adenosine 5′-(β,γ-imino)triphosphate) display an α/β-fold that is characteristic for the galactose kinase/serine kinase/MVA kinase/phosphoMVA kinase superfamily with two domains, one for the cofactor and another for the substrate binding, and a catalytic center positioned in a deep cleft between the two domains.[106] The 1.7 Å resolution structure of the *Thermus thermophilus* enzyme showed the same characteristics as those of the GHMP kinase superfamily and pointed out the role of conserved lysine and aspartate residues, which are involved in the catalysis and are also found in the *E. coli* enzyme and in the MVA kinase.[107] The pyrimidine moiety of the substrate is involved in important interactions in the active site. Only CDP–ME **28** is phosphorylated; its uridine analogue is left intact.

The knowledge of the three-dimensional structure of the *E. coli* enzyme permitted to design inhibitors fitting into the active site. A first series of competitive inhibitors (e.g., racemic **80**, **Figure 11**, $K_i = 0.29\ \mu mol\ l^{-1}$ for the *E. coli* enzyme) with inhibition constants in the upper nanomolar range were obtained. They are drug-like, containing no phosphate or phosphonate moiety, and do not occupy the adenine-binding site.[108,109] The cocrystal structure of the *Aquifex aeolicus* enzyme/inhibitors complex shows that the inhibitor binds in a substrate-like mode and not in an adenosine-like mode and that the cyclopropyl ring is inserted in a small site, which is not occupied by the substrate. A second series of competitive inhibitors of the *E. coli* enzyme correspond to cytidine-based compounds, for example, **81** ($K_i = 85.8\ \mu mol\ l^{-1}$) and **82** ($K_i = 71.2\ \mu mol\ l^{-1}$).[110]

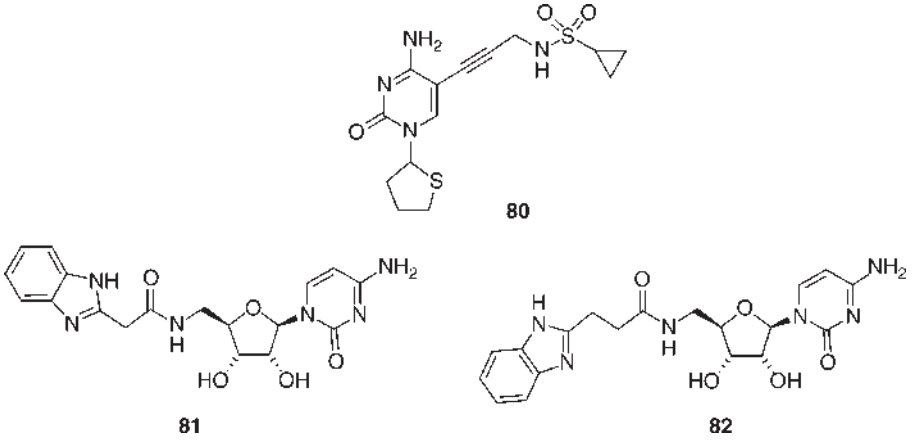


Figure 11 Inhibitors of 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase.

They were initially designed to inhibit the *E. coli* CDP–ME synthase but were finally shown to inhibit the next step catalyzed by the CDP–ME kinase. The binding mode of the inhibitors was verified in a crystal structure of the *A. aeolicus* enzyme/inhibitor crystal that complemented the *E. coli* enzyme structure.

1.13.2.3.3 2-C-m,ethyl-D-erythritol 2,4-cyclodiphosphate and 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

In *E. coli*, the *ygbB* gene accompanies the *ygbP* and the *ychB* genes in a small operon. In many bacteria, the putative orthologues of the *E. coli ygbB* gene are tightly linked or even fused to putative orthologues of *ygbP*. Homologous expression of the *E. coli ygbB* gene in a recombinant *E. coli* strain yielded a soluble protein catalyzing the conversion of 2-phospho-CDP–ME **29** into 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate **30** (MEcDP), whose structure was established by ^{13}C-NMR using ^{13}C-Labeled substrates (**Figure 4**).[111] MEcDP **30** was a known natural product, being accumulated in the fermentation broth of *Desulfovibrio desulfuricans*[112] or in millimolar concentrations in the cytoplasm of some bacteria (e.g., *Corynebacterium ammoniagenes*) upon benzylviologen treatment.[113,114] The enzyme reaction depends on a divalent cation, Mg^{2+} or Mn^{2+}. No other cofactor is required. ^{13}C-Labeled MEcDP **30** was efficiently incorporated into the carotenoids of isolated chromoplasts from *C. annuum* and *Narcissus pseudonarcissus*, establishing its role of intermediate in the MEP pathway.[115] As a side reaction, the *ygbB* protein also converted CDP–ME **28** into 2-*C*-methyl-D-erythritol 3,4-cyclophosphate **83** (**Figure 12**),[111] which is not incorporated into the chromoplast carotenoids and is most probably an *in vitro* reaction product without metabolic significance. Similar properties were found for the corresponding *P. falciparum* enzyme, converting CDP–ME **28** into 2-*C*-methyl-D-erythritol 3,4-cyclophosphate **83** and 2-phospho-CDP–ME **29** into 2-phospho-2-*C*-methyl-D-erythritol 3,4-cyclophosphate **84** (**Figure 12**).[116] The search for mutants of the *E. coli* transformant capable of utilizing exogenous MVA as isoprenoid precursors also led to the characterization of the same *ygbB* gene and gene product.[117]

Crystal structures were obtained for the *E. coli* enzyme. The protein adopts a biologically active bell-shaped homotrimeric structure around a central hydrophobic cavity and three externally facing active sites.[118,119] In a first structure, the protein contains a tetrahedral zinc-binding site with Asp8, His10, and His42 as ligands, the fourth coordination site being occupied by the β phosphate of the substrate. Soaking the crystals with CDP permitted to identify the active site.[118] Four factors are putatively involved in the substrate binding and in the formation of the active conformer of the substrate: binding of the CDP β phosphate by the Zn^{2+} ion, fixation of the conformation between the CDP α and β phosphate groups by the Mg^{2+} ion, orientation of the ME moiety in a complementary cavity and finally binding the 2-phosphate in a phosphate binding cavity.[118] Three other structures were independently described containing either a Mn^{2+} cation in a tetrahedral geometry linked to the residues complexing Zn^{2+} in the former structure, Mn^{2+}, CMP, and MEcDP, revealing the structural requirements for substrate, product, and Mn^{2+} recognition,[119] or Mn^{2+} and CDP.[120] In the latter structure, an octahedral Mn^{2+} ion is positioned between the α and β phosphate groups acting in concert with a tetrahedral Zn^{2+} ion to align and polarize the substrate. These studies were corroborated by additional structures of the *T. thermophilus* enzyme complex with CDP–ME **28**, which lacks the phosphate group involved in the nucleophilic attack required for the formation of the cyclodiphosphate group, and the normal substrate CDP–ME 2-phosphate **29**.[121] In the latter case, only cytosine monophosphate was found in the active site. The enzyme presents a hydrophobic cavity capable of trapping farnesyl diphosphate for the *Shewanella oneidensis* protein[122] or IPP **2** and/or DMAPP **3**, geranyl diphosphate **4** or farnesyl diphosphate **5** (**Figure 1**) for the *E. coli* protein,[123] suggesting a feedback regulation by metabolites derived from steps located downstream in the isoprenoid biosynthetic pathway.

From the knowledge of the three-dimensional structure, fluorescent inhibitors were designed.[124] The X-ray crystal structures of the tertiary complex protein/Zn^{2+}/inhibitor showed that the fluorescent ligands **85** and **86** (**Figure 12**) occupy both a rigid well-conserved pocket binding cytidine and a larger pocket of an adjacent monomer binding of the ME moiety of the substrate and also containing the tetrahedrally coordinated Zn^{2+} ion and pointed out π-stacking of the base against Ala131 and the backbone from Ala100 to Leu106, hydrogen bonds between the ribose C-2′and C-3′hydroxyl groups and the carboxylate of Asp56. Fluorescence-binding titration gave K_d values of 36 and 15 $\mu mol\,l^{-1}$ for compounds **7** and **9**, respectively. Based on computer modeling, weak nonphosphate inhibitors with a cytosine analogue moiety **87–89** (**Figure 12**) were obtained with IC_{50} values in the upper micromolar range.[125] Inhibition is thus obtained without interactions with the diphosphate binding site and with the Asp56 side chain, which interacts with the ribose moiety of the cytosine

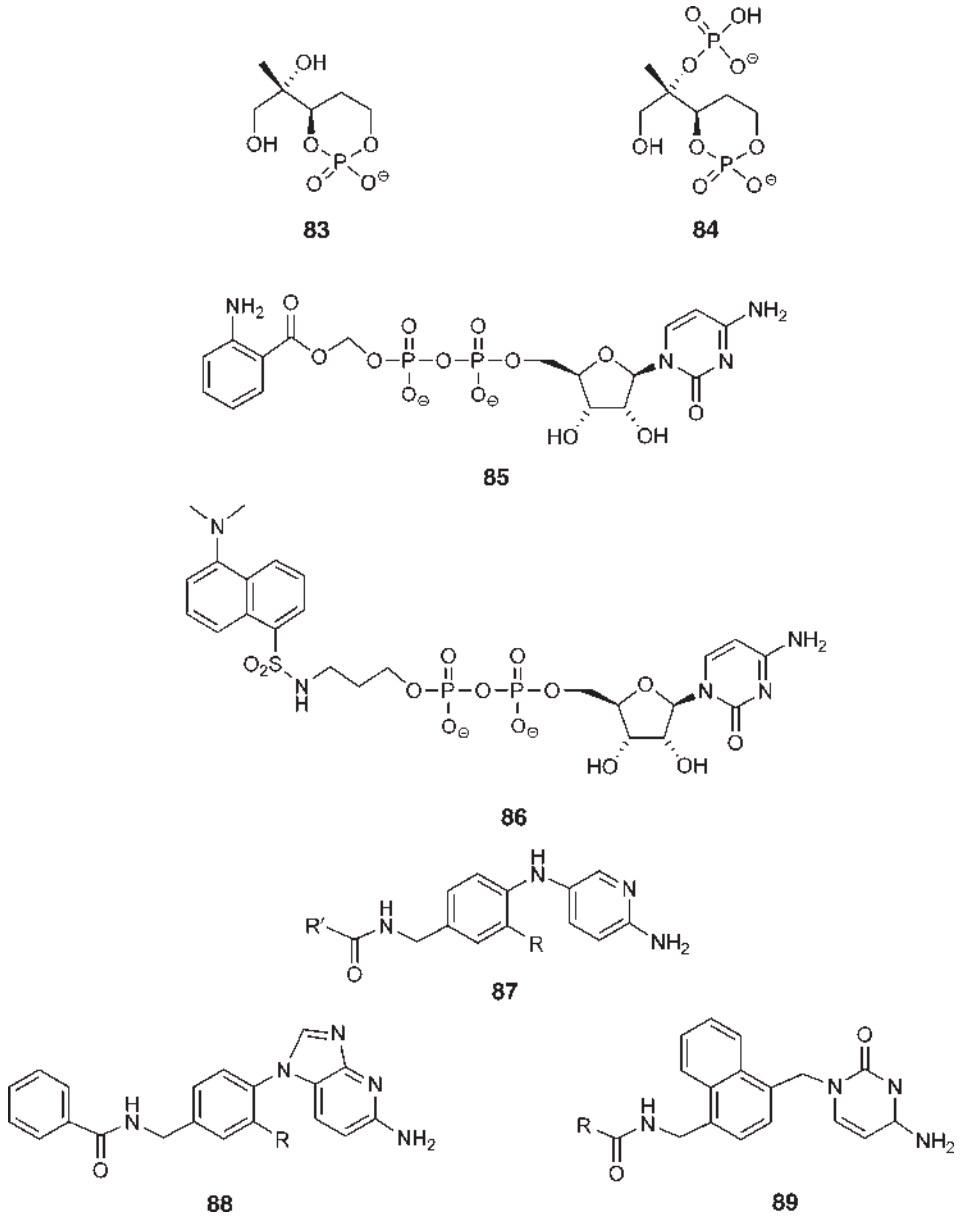


Figure 12 2-*C*-Methyl-D-erythritol 2,4-cyclodiphosphate synthase: side-reaction products (**83**, **84**) and inhibitors (**85–89**).

analogues. An additional subpocket was detected: it accommodates an aromatic spacer between the cytosine-like moiety and rings entering a flexible hydrophobic region of another pocket.

1.13.2.3.4 Bacterial bifunctional 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase/ 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

In α- and ε-proteobacteria, the genes encoding the CDP–ME synthase and the 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate synthase are fused.[126] The corresponding bifunctional protein from *Campylobacter jejuni* obtained in a recombinant *E. coli* strain catalyzes two nonconsecutive steps: the conversion of MEP **27** into CDP–ME **28** in the presence of CTP and the transformation of CDP–ME 2-phosphate **29** into MEcDP **30** at catalytic rates of 19 and 7 μmol l^{-1} mg^{-1} min^{-1} in the presence of a divalent metal ion, Zn^{2+}, Mn^{2+}, and Mg^{2+} being the best.[126] The crystal structure of the *C. jejuni* bifunctional enzyme showed an elongated hexamer with *D*3 symmetry, and

protein–protein complex formation has been found with the *A. tumefaciens* CDP–ME kinase.[127] The same bifunctional enzyme has been characterized in *Mesorhizobium loti*. Incubation with MEP **27**, CTP, and ATP afforded only CDP–ME **28**.[128] After addition of CDP–ME kinase from *E. coli*, the same system produced MEcDP **30**. Time-dependency studies pointed out the absence of substrate channeling between the active site of the bifunctional CDP–ME synthase/MEcDP synthase and the CDP–ME kinase from *A. tumefaciens*.[129] No intermediates are transferred from the CDP–ME synthase active site of the bifunctional enzyme to the CDP–ME kinase active site or from the CDP–ME kinase active site to the MEcDP synthase active site.

1.13.2.4 4-Hydroxy-2-Methylbut-2-Enyl Diphosphate and 2-*C*-Methyl-D-Erythritol 2,4-Cyclodiphosphate Reductase

The *gcpE* gene in *E. coli* is an essential gene and was shown to accompany the *dxr* gene encoding the DXP reducto-isomerase, the second enzyme of the MEP pathway. Its implication in the MEP pathway was demonstrated largely before its function was known. The *gcpE* deletion is lethal for *E. coli*, but can be rescued by inserting a synthetic operon encoding yeast MVA kinase, human phosphoMVA kinase, and yeast diphosphoMVA decarboxylase allowing the conversion of MVA into IPP through this partial sequence of the MVA pathway.[130] Growth of this *E. coli* transformant can be restored upon addition of exogenous MVA to the culture medium.[130,131]

First insight into the role of the *gcpE* gene in the MEP pathway was obtained by incubation of uniformly labeled [U-$^{13}C_5$]DX in recombinant *E. coli* cells engineered for the overexpression of genes encoding MEP pathway enzymes.[132] The crude cell-free extracts were directly analyzed for the accumulated metabolites by ^{13}C-NMR. MEcDP **30** was accumulated in large amounts in cells simultaneously expressing the D-xylulose kinase (required for the phosphorylation of the added ^{13}C-labeled DX precursor) and all known enzymes of the MEP pathway required for the DXP **26** metabolism, from the DXR to the MEcDP synthase (**Figure 4**). Additional expression of *gcpE* resulted in the accumulation of a novel metabolite identified by NMR spectroscopy as (*E*)-4-hydroxy-3-methylbut-2-enyl 1-diphosphate **31** (HMBPP, **Figure 4**). Incubation of [1-^{3}H]ME in a *gcpE*-deficient *E. coli* strain engineered for the utilization of exogenous MVA resulted in the accumulation of radiolabeled MEcDP **30**,[133] indicating that MEcDP **30** is the probable substrate of the GcpE protein. A crude cell-free system of an *E. coli* strain overexpressing *gcpE* converted [2-^{14}C]MEcDP in the presence of an *E. coli* alkaline phosphatase into (*E*)-2-methyl-but-2-ene-1,4-diol, which was identified by comparison using NMR spectroscopy with a synthetic reference compound.[134] This diol resulted most probably from the hydrolysis of a phosphorylated derivative by the added phosphatase, which left the cyclodiphosphate of the substrate untouched. A similar experiment performed with an *E. coli* strain overexpressing GcpE, but in the absence of a phosphatase afforded HMBPP **31**, which was identified by radiochemical methods using a sample obtained by synthesis, affording the first precursor to product relationship for MEcDP **30** and HMBPP **31**.[135] A conserved amino-acid motif is found in the GcpE protein family and matched with a [4Fe–4S] ferredoxin and an aconitase signature, suggesting that the MEcDP reductase contains a [4Fe–4S] prosthetic group (**Figure 13**) like the two latter enzymes.[135] Such [4Fe–4S] enzymes are usually quite oxygen sensitive, and GcpE enzyme tests were

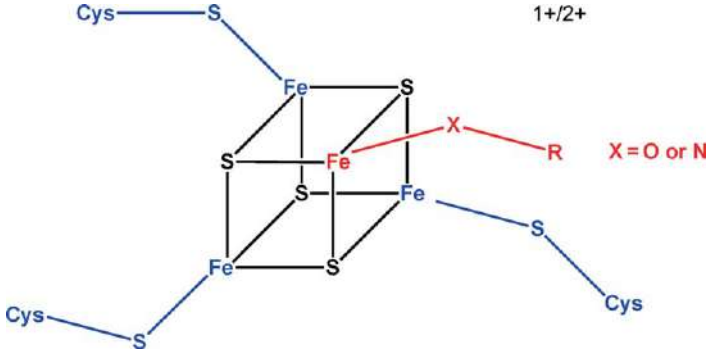


Figure 13 [4Fe–4S] cluster as prosthetic group of the 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate reductase.

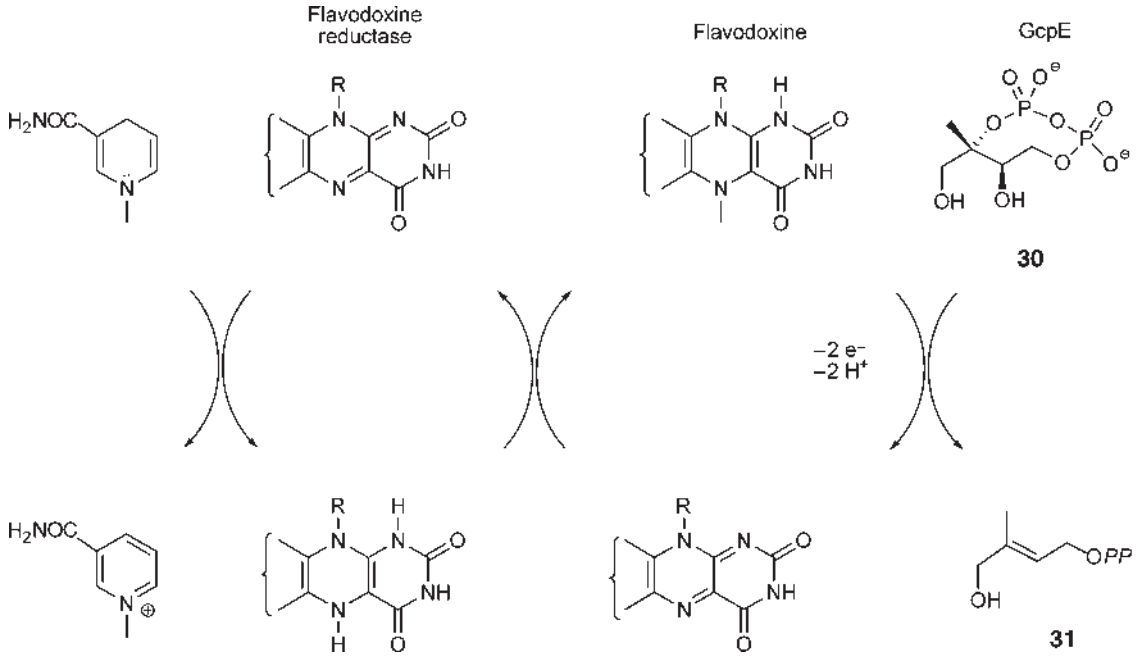


Figure 14 2-*C*-Methyl-D-erythritol 2,4-cyclodiphosphate reductase of *Escherichia coli* and associated reducing system.

subsequently best performed under an inert atmosphere and after degassing the buffers.[134,135] The MEcDP reductase catalyzes a reaction involving the cleavage of two carbon/oxygen bonds and a reduction requiring two electrons. The recognition of the *E. coli* GcpE as a [4Fe–4S] enzyme opened new perspectives for the full characterization of the prosthetic group and for the identification of the associated reducing system.[136] Oxygen-sensitive [4Fe–4S] enzymes lose their activity by destruction of the prosthetic group. Upon reconstitution in the presence of $FeCl_3$, Na_2S, and dithiothreitol under an inert atmosphere restores in the purified *E. coli* GcpE apoenzyme the intact [4Fe–4S] cubane (**Figure 13**) in the active site, which is characterized by *ca.* 4 iron and 4 inorganic sulfide ions per enzyme molecule and a characteristic visible absorption band at 410 nm. Efficient reduction of MEcDP **30** into HMBPP **31** was observed in the presence of the classical biological reducing system found in bacteria, flavodoxin/flavodoxin reductase/NADPH (**Figure 14**), or in the presence of a chemical system such as the semiquinone radical obtained by photochemical reduction of 10-methyl-5-deaza-isoalloxazine (5-deazaflavin).[136] A recombinant fusion protein including a maltose-binding protein and the *E. coli* MEcDP reductase failed to convert MEcDP **30** into HMBPP **31**, but addition of a crude cell-free system of an *E. coli* MEcDP reductase-deficient strain restored the activity, confirming that a redox system and probably auxiliary proteins are required for enzyme activity.[137] This as-isolated protein (i.e., without any reconstitution) isolated under anaerobic conditions showed activity in the presence of reducing systems such as flavodoxin/flavodoxin reductase/NADPH (specific activity above 60 nmol mg^{-1} min^{-1}) or the artificial reducing agent photoreduced 5-deazaflavin (with a lower specific activity), but no activity in the presence of dithionite.[138] This enzyme preparation showed a deficit in iron as compared to the reconstituted enzyme (2.4 iron and 4.4 sulfide ions), corresponding most likely to a mixture of enzyme with an intact [4Fe–4S] cluster, enzyme with a degraded cluster and/or apoenzyme. Disruption of the flavodoxin I *fldA* gene by transposon mutagenesis was performed in an *E. coli* strain expressing MVA pathway genes and possessing a point mutation in the *lytB* gene.[139] This construct accumulated HMBPP **31**. Additional disruption of the *fldA* gene suppressed the MVA-independent growth and drastically reduced the accumulation of HMBPP **31**. The *fldA*$^-$ mutant required MVA, indicating that this gene is essential in isoprenoid biosynthesis in aerobic growth conditions, and is most likely required as a reducing system by the iron/sulfur cluster enzymes GcpE and possibly LytB (see Section 1.13.2.5.2).

The recombinant GcpE enzyme of another bacterium, the thermophile *T. thermophilus*, has been obtained by heterologous expression in *E. coli* and purified in oxygen-free conditions.[140] With dithionite as associated reduction system, the K_m value for MEcDP was 0.42 mmol l^{-1} ($k_{cat} = 0.4\ s^{-1}$, maximal specific activity 0.6 µmol min^{-1} s^{-1} at pH 5 and 55 °C). The role of HMBPP **31** in plant isoprenoid biosynthesis was checked with *C. annuum* and *N. pseudonarcissus* chromoplasts that efficiently incorporated ^{3}H-labeled HMBPP **31** into carotenoids.[141]

Characterization of the [4Fe–4S] cluster of the bacterial *E. coli* enzyme and the *A. thaliana* plant enzyme was obtained by Mössbauer spectroscopy upon reconstitution with $^{57}FeCl_3$.[142] In the amino acid sequence of MEcDP reductase, three cysteines are highly conserved, serving to anchoring the cluster in the active site. Replacement of any of these cysteines by another amino acid resulted in a total loss of activity.[138] The Mössbauer spectra of the plant and the bacterial enzymes showed two components in a 3:1 ratio: the first component corresponds to three tetrahedral sulfur-coordinated $Fe^{2.5+}$ in a diamagnetic ground state, and the second component to a single $Fe^{2.5+}$ with one cysteine ligand being replaced by a nonsulfur ligand. This fourth iron with a ligand involving either oxygen or a nitrogen ligand was postulated to act as a Lewis acid in the elimination reaction of the C-2 hydroxy group of MEcDP **30**.[136] Kinetic studies performed with the *T. thermophilus* enzyme showed in the presence of the substrate MEcDP and dithionite as reductant that a new transient signal is detected in EPR spectroscopy, suggesting that an intermediate is bound to the cluster during the reaction.[143]

The mechanistic details of the conversion of the diol derivative (as in MEcDP **30**) into a double bond (as in HMBPP **31**) are not known. Two electrons are required for the MEcDP-catalyzed reaction. [4Fe–4S] Cluster enzymes are involved in single electron transfers. This most likely implies two consecutive single electron transfers and radical intermediates. Several hypotheses have been proposed, but are awaiting confirmation.[136,137,144]

The plant MEcDP reductase gene has been characterized in *A. thaliana*.[145] It presents two differences with the homologous *E. coli* gene: the presence of an N-terminal plastid targeting sequence and surprisingly an additional 30 kDa central domain. From the mechanistic point of view, the plant enzyme differs also from the bacterial enzyme by the associated reducing agent. Flavodoxin and the whole flavodoxin/flavodoxin reductase/NADPH reducing system are absent in plant chloroplasts. Plant isoprenoid biosynthesis in plastids is highly linked to photosynthesis (for more details, see Section 1.13.4.5). First, evidence of a protein–protein interaction between HMBPP reductase and the ferredoxin I of the cyanobacterium *Thermosynechococcus elongatus*.[146] The two proteins were shown using a bacterial two-hybrid system designed to characterize protein–protein interactions to strongly interact, and in the presence of the electron shuttle NADPH/ferredoxin oxidoreductase, ferredoxin was shown to be capable of transferring electrons on MEcDP reductase. The *A. thaliana* MEcDP reductase, in association with a purified spinach thylakoid, preparation was capable of converting MEcDP **30** into HMBPP **31** in the absence of any reducing cofactor, just upon illumination.[147] This proves that the electron flow from the photosynthesis resulting from the photo-oxidation of water can directly enter into isoprenoid biosynthesis through ferredoxin transferring these electrons onto the [4Fe–4S] cluster of MEcDP reductase (**Figure 15(a)**). In the absence of light (during the night or in roots), an electron shuttle is required: the ferredoxin–ferredoxin reductase–NADPH system can perform this task (**Figure 15(b)**).[147]

Some plants accumulate MEcDP in their chloroplasts during bright and hot days. In spinach leaves, this accumulation is stimulated by a high irradiation at supraoptimal temperatures,[148] suggesting that the conversion of MEcDP **30** into HMBPP **31** by the MEcDP reductase is rate-limiting under such oxidative stress conditions. Isoprenoid biosynthesis through the MEP pathway was, however, not blocked, suggesting that the damages induced by reactive oxygen species could be continuously repaired. In the presence of Cd^{2+}, MEcDP was accumulated, but the MEP pathway was fully blocked. Cd^{2+} does not inhibit the activity of reconstituted MEcDP reductase with an intact Fe/S cluster, suggesting that this cation interferes with the system involved in the reparation of the cluster after oxidative damage, and that the MEcDP accumulation is a signature for a HMBPP reductase-dependent bottleneck in the MEP pathway.

1.13.2.5 The Formation of Isopentenyl Diphosphate and Dimethylallyl Diphosphate and 4-Hydroxy-2-Methylbut-2-Enyl Diphosphate Reductase

1.13.2.5.1 Presence of a branching in the MEP pathway producing IPP and DMAPP

Surprisingly, the deletion of the IPP isomerase *idi* gene does not affect the growth of *E. coli*.[149] The corresponding gene product catalyzes the interconversion of IPP **2** (**Figure 3**) and DMAPP **3**, which are both required as

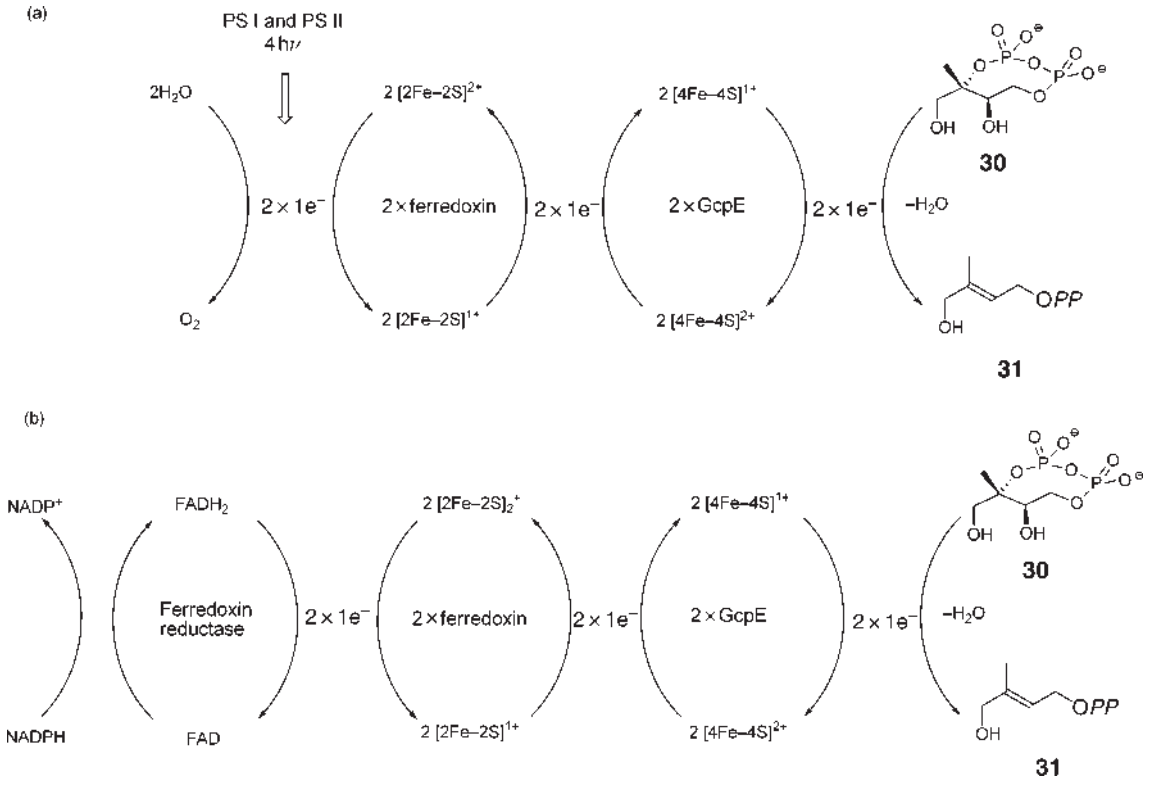


Figure 15 2-*C*-Methyl-D-erythritol 2,4-cyclodiphosphate reductase of *Arabidopsis thaliana* and associated reducing system (a) in light conditions (photosynthesis) or (b) in the dark. Reproduced from M. Seemann; B. Tse Sum Bui; M. Wolff; M. Miginiac-Maslow; M. Rohmer, *FEBS Lett.* **2006**, *580*, 1547-1552.

precursors for the isoprene units. This represented a striking feature for the isoprenoid biosynthesis: either another unknown isomerase is present in this bacterium, or the MEP pathway provides separately and independently IPP **2** and DMAPP **3** (**Figure 4**). Evidence for a branching in the MEP pathway, separately yielding IPP **2** and DMAPP **3** from a single precursor was brought by genetic methods.[150] An *E. coli* strain with a nonfunctional MEP pathway after deletion of the *dxr* gene encoding the second enzyme of the pathway was engineered for the utilization of exogenous MVA as precursor for isoprene units (for details see Section 1.13.2.3.1). This strain requires for growth the addition of either free methylerythritol or MVA to the culture medium. When it grows on MVA, the MVA pathway affords only IPP, and the bacteria are dependent on a functional IPP isomerase to produce DMAPP (**Figure 3**). When an additional deletion of the IPP isomerase *idi* gene is introduced, the novel construct does not grow anymore in the presence of MVA, indicating that there is no other way to synthesize DMAPP from IPP as the reaction catalyzed by the IPP isomerase. It grows, however, normally in the presence of ME, indicating that IPP and DMAPP are independently produced through the MEP pathway.

1.13.2.5.2 4-Hydroxy-2-methylbut-2-enyl diphosphate reductase

lytB was the candidate for the gene encoding the enzyme catalyzing the last step of the MEP pathway. It accompanied all known genes of the pathway. First evidence of its implication in isoprenoid biosynthesis was obtained with the cyanobacterium *Synechocystis* sp. PCC6803.[151] A mutant with an insertion in the promoter region of *lytB* grows slowly is depleted in photosynthetic pigments, and bleaches easily. Insertion in the coding region is lethal, but growth is surprisingly restored by supplementation of the culture medium with isopentenol and dimethylallyl alcohol, the nonphosphorylated counterparts of IPP **2** and DMAPP **3**. In an *E. coli* strain engineered for producing carotenoids, the *Synechocystis lytB* gene or the corresponding cDNA from *Adonis aestivalis*, a higher plant, enhanced production of these pigments. Definitive proof of the role of the *lytB* gene in the MEP pathway was again obtained with an *E. coli* strain engineered for the utilization of MVA as an

isoprenoid precursor.[152] Deletion of the *lytB* gene was lethal and could only be rescued when the cells were supplemented with MVA in the culture medium.

Characterization of the enzyme encoded by the *lytB* gene in *E. coli* was performed on the same way as for *gcpE*.[153] An *E. coli* construct simultaneously overexpressing all genes from the D-xylulose kinase required for the phosphorylation of free DX to the gene of the MEcDP reductase was fed with uniformly labeled [U-$^{13}C_6$]DX. ^{13}C-NMR analysis of the crude cell-free extract revealed the presence of HMBPP **31**. Additional expression of the *lytB* gene revealed the presence in the cell-free extract of IPP **2** and DMAPP **3** in an approximately 5:1 ratio (**Figure 4**), indicating that HMBPP is most likely the substrate of the LytB protein, which has two reaction products, IPP **2** and DMAPP **3**, and that LytB corresponds thus to the branching previously identified. First data on the enzyme activity of the *lytB* gene product from *E. coli* was obtained with crude cell-free extracts from a strain overexpressing the *lytB* gene,[154] which converted ^{3}H-labeled HMBPP **31** into a mixture of IPP **2** and DMAPP **3** (**Figure 16**).[154] In order to characterize IPP **2**, the major product, incubation of crude cell-free extracts of an *E. coli* strain overexpressing the *lytB* gene was performed with ^{13}C-labeled HMBPP **31** and NMR analysis of the crude reaction mixture. NADH, FAD, and divalent cations such as Co^{2+} were reported to enhance the reaction yield. These observations were not fully confirmed by further experiments performed on the purified enzyme once the right incubation conditions were found. The purified protein had no catalytic activity. This was not surprising once the nature of the [4Fe–4S] prosthetic group was recognized. Indeed, the presence of a [4Fe–4S] cluster requires manipulation of the enzyme in an inert atmosphere and/or reconstitution of the cluster. HMBPP reductase of the thermophilic bacterium *T. thermophilus* was thus isolated in a glove box under a N_2/H_2 atmosphere and converted HMBPP **31** into IPP **2** and DMAPP **3** in a 4:1 to 5:1 range.[155] Steady-state kinetic parameters could be determined: 6.6 µmol min^{-1} mg^{-1} maximal specific activity (pH 7.5, 60 °C), $k_{cat} = 3.7\ s^{-1}$, $K_m = 590\ \mu mol\ l^{-1}$ for HMBPP. The enzyme was extremely oxygen sensitive and required the presence of a reducing agent. The nonbiological dithionite was used in these first assays. As for MEcDP

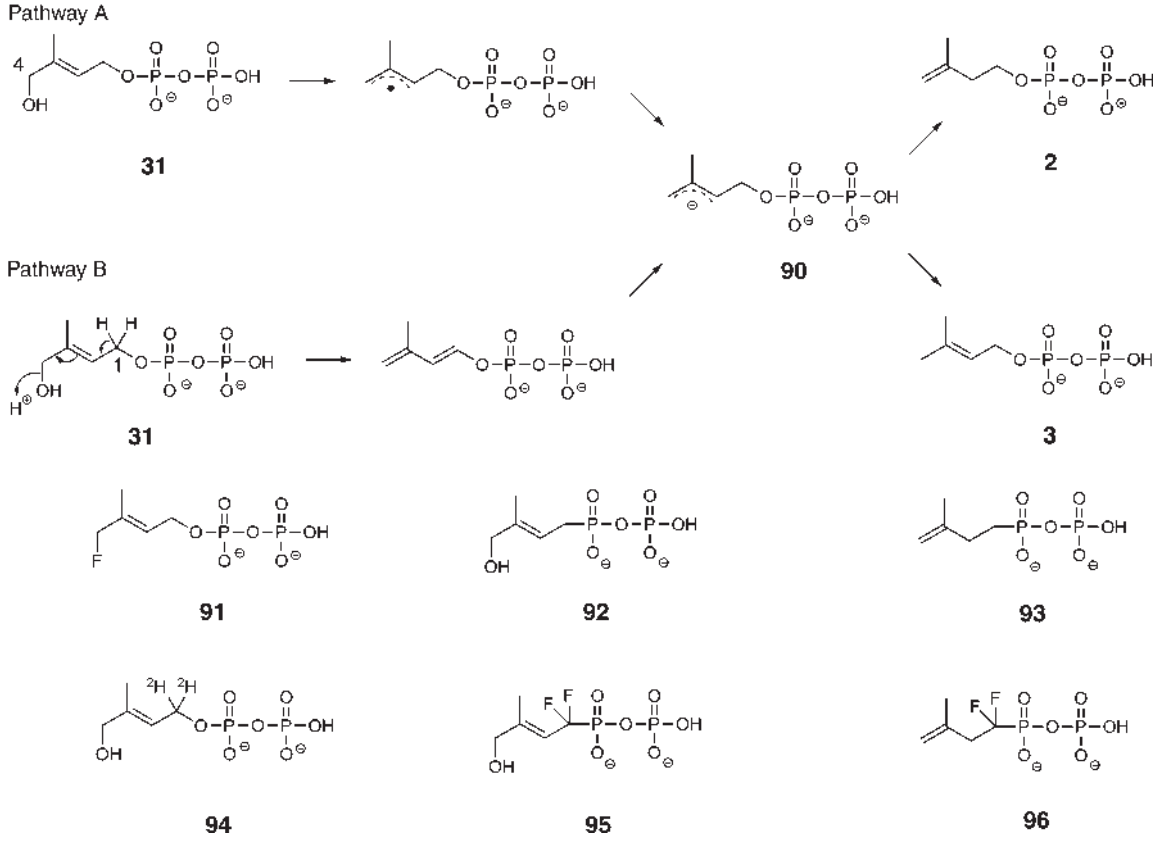


Figure 16 4-Hydroxy-3-methylbut-2-enyl diphosphate reductase: reaction mechanism and substrate analogues (**91–96**).

reductase, the 5-deazaflavin semiquinone radical obtained upon photoreduction (0.4 μmol min^{-1} mg^{-1} specific activity) or the biological system flavodoxin/flavodoxin reductase/NADPH (3 nmol min^{-1} mg^{-1} specific activity) could be associated to the purified *E. coli* HMBPP reductase and afforded IPP **2** and DMAPP **3** in a 6:1 ratio.[137] The *P. falciparum* HMBPP reductase was shown to form a stable complex with ferredoxin and to be catalytically active in the presence of the apicoplast-localized electron transfer system ferredoxin/ferredoxin reductase/NADPH, much like the plant MEcDP reductase previously described (Section 1.13.2.4).[156]

First characterization of the *E. coli* HMBPP reductase prosthetic group was obtained by EPR spectroscopy.[157] Upon reconstitution under an inert atmosphere with $FeCl_3$, Na_2S, and dithiothreitol, the reconstituted protein was characterized by its 410 nm visible absorption band ($\varepsilon = 18\,750\ mol^{-1}\,l\,cm^{-1}$) and a large increase in its iron and sulfide content (5 and 6, respectively, per protein monomer vs. 0.8 and 0.6 in the as-isolated apo-protein). Upon dithionite treatment, the reduced $[4Fe–4S]^{1+}$ form of the cluster could be characterized in the reconstituted protein by its typical EPR spectrum, whereas in the as-isolated enzyme only a [3Fe–4S] cluster, resulting from the degradation of the native [4Fe–4S] form, could be characterized.[157] In the as-isolated enzyme, the oxidized $[4Fe–4S]^{2+}$ form of the cluster cannot be characterized by EPR methods. It is certainly present to some extent as the enzyme was active in the presence of a reducing system. The as-isolated enzyme was, however, much less active than reconstituted enzyme when incubated under an inert atmosphere (e.g., with respective 11 and 30 nmol mg^{-1} min^{-1} specific activities in the presence of 170 μmol l^{-1} photoreduced 5-deazaflavin in the same experimental conditions). EPR spectroscopy experiments confirmed later the presence of the degraded $[3Fe–4S]^{1+}$ cluster in the as-isolated enzyme.[158] In the absence of any measurements on the reduced enzyme with a [4Fe–4S] cluster, which gives an EPR signal only in its reduced form, there is no proof for the presence or the absence of the intact cluster. Implementation with a gene cassette involved the overexpression of the *isc* operon involved in the assembly of the Fe/S clusters enhanced the catalytic activity of the enzyme purified under an inert atmosphere. A maximal catalytic activity was obtained with a 40 μmol l^{-1} flavodoxin and a 12 μmol l^{-1} flavodoxin reductase concentration. Replacement of any of the three cysteines (Cys12, Cys96, and Cys197) required for the binding of the cluster resulted in a nearly complete loss of the catalytic activity.[158] This feature (Cys13, Cys96, and Cys193) was found again in the first X-ray crystallographic structure of the *Aquifex aeolicus* enzyme.[159] Two histidines (His42 and His124) would allow electrostatic binding of HMBPP **31**, and Glu126 might be involved in the protonation of the anionic intermediate **90** (**Figures 16** and **17**). The crystal structure contains, however, a [3Fe–4S] cluster in place of the native [4Fe–4S] found in the active enzyme.

Similar to MEcDP reductase, the HMBPP reductase-catalyzes a reaction requiring a two-electron reduction. As a [4Fe–4S] cluster enzyme, this implies two consecutive one-electron transfers. An iron of the cluster may bind the substrate and act as Lewis acid for the elimination of the hydroxy group. The reaction involves most probably an allylic anion **90** as intermediate that can be protonated at C-2 yielding IPP **2** or at C-4 yielding DMAPP **3** (**Figure 16**). When the HMBPP reductase-catalyzed reaction is performed *in vitro* in ^2H_2O with the purified enzyme in the presence of flavodoxin/flavodoxin reductase/NADPH as reducing system, deuterium is thus found at C-2 of IPP and C-4 of DMAPP.[160] *In vivo*, the protonation mechanism may be more sophisticated, involving the transfer of another proton not directly originating from water. In an experiment designed to characterize the reduction steps of the MEP pathway, the bacterium *Z. mobilis* was grown in the presence of [1-^{2}H]glucose as a sole source of carbon and energy.[161] *Zymomonas mobilis* has no complete tricarboxylic acid cycle, utilizes only hexoses as a carbon source, and produces huge amounts of ethanol by fermentation in anaerobic growth conditions. This implies that all reducing agents derived from NADPH/NADH are synthesized through the catabolism of glucose involving only three enzymes: glucose 6-phosphate dehydrogenase, GAP dehydrogenase, and ethanol dehydrogenase. This means that *in vivo* the ^2H of [1-^{2}H]glucose is transferred by the first enzyme onto $NADP^+$, and that by the action of the next two enzymes a 1:1 pool of the two ^2H-labeled NAD(P)H isotopomers is synthesized. Any reduction step should thus leave a deuterium signature on the corresponding carbon atom. This is effectively observed on the triterpenic hopane skeleton for the carbon atoms issued from C-4 of IPP corresponding to the NADPH-dependent reduction catalyzed by the DXR. ^2H-labeling is, however, also found on the C-2 carbon atoms corresponding to the protonation site of the HMBPP reductase mentioned above. This suggests that the ^2H transferred from [1-^{2}H]glucose onto NADPH is successively transferred without significant loss onto the flavines of flavodoxin and flavodoxin reductase, and is finally released as a $^2H^+$, serving for the protonation of the allylic intermediate anion **90** (**Figure 17**).[162]

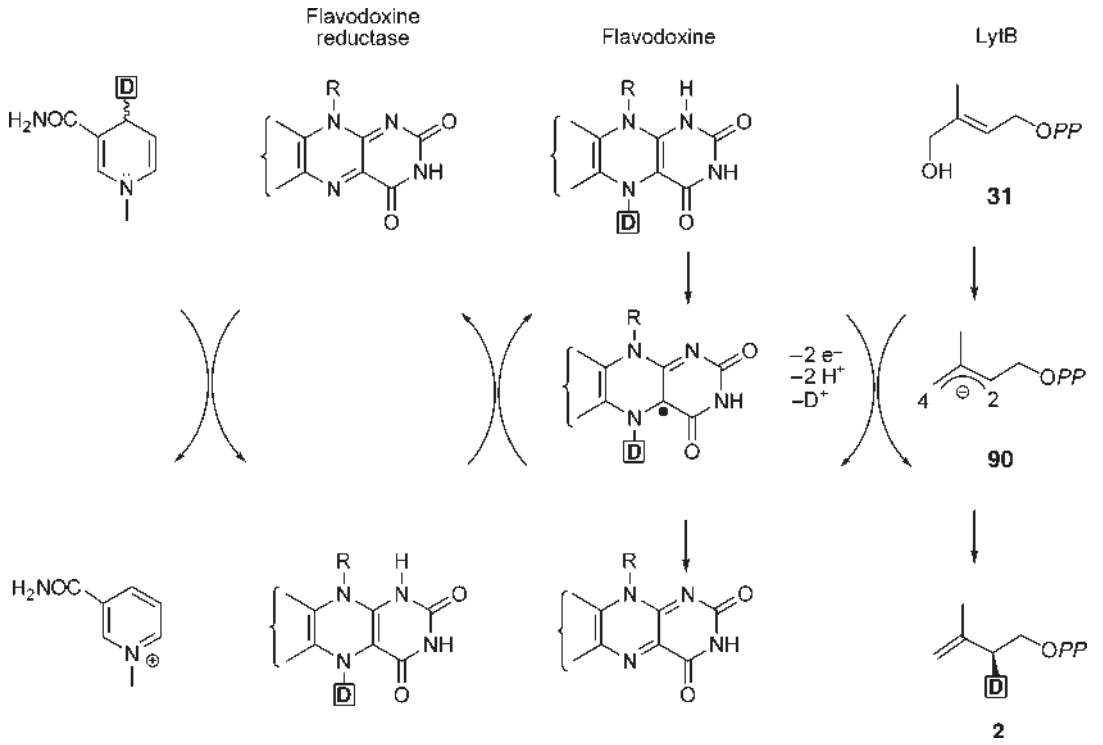


Figure 17 4-Hydroxy-3-methylbut-2-enyl diphosphate reductase from *Zymomonas mobilis*: hypothetical mechanisms of deuterium transfer from NADPH onto isopentenyl diphosphate **2**. Reproduced from M. Rohmer, *Pure Appl. Chem*. **2007**, *78*, 739–751.

In order to investigate mechanistic aspects of the HMBPP reductase-catalyzed reaction (**Figure 16**), two probes, (*E*)-3-(fluoromethyl)-but-2-enyl diphosphate **91** and (*E*)-4-hydroxy-3-methyl-2-but-2-enyl pyrophosphonate **92** (**Figure 16**), with, a C–F or a C–P bond were tested on the *E. coli* enzyme in the presence of flavodoxin/flavodoxin reductase/NADPH.[163] The reported catalytic characteristics of this preparation versus HMBPP reductase were 11.6 min^{-1} for the k_{cat} and a value lower than 15 $\mu mol\,l^{-1}$ for the K_m. Both analogues **91** ($K_m = 3.95\ mmol\,l^{-1}$, $k_{cat} = 0.55\ min^{-1}$) and **92** ($K_m < 15\ \mu mol\,l^{-1}$, $k_{cat} = 0.44\ min^{-1}$) were substrates of the enzyme: analogue **91** is converted into IPP **2** and DMAPP **3** in a 7:1 ratio, whereas the phosphonate **92** yielded only the shorter phosphonate analogue of IPP **93** (**Figure 17**). The increased K_m (200-fold compared to that of HMBPP) and the decreased k_{cat} (4.7% of that of HMBPP) recorded for the fluorinated analogue **91** are consistent with the binding of the HMBPP substrate **31** through its hydroxy group. In addition, the strong electronegativity of fluorine precludes any homolytic cleavage of the C–F bond and is thus in favor of the heterolytic cleavage of the C–O bond in HMBPP **31**. The similar K_m values for the phosphonate analogue **92** and for HMBPP **31** suggest a similar binding for the two ligands. The formation of a single product **93**, corresponding to an IPP analogue, suggested, however, that the shortening of the chain induced by the missing oxygen atom is not optimal for the protonation reaction.

In order to test whether a C–H bond cleavage occurs at C-1 (**Figure 16**, pathway B), the deuterated isotopomer [1,1-$^2H_2$]HMBPP **94** was tested on the HMBPP reductase.[164] No primary isotope effect is observed. The steady-state kinetic parameters are the same as those reported for the natural abundance of HMBPP. This observation is in favor of mechanism A (**Figure 16**, pathway A). It is compatible with mechanism B only if the elimination of the C-1 proton is not the rate-limiting step. In order to differentiate between these two hypotheses, the action of (*E*)-4-hydroxy-3-methyl-2-butenyl-1,1-difluoro diphosphonate **95** (**Figure 16**) was evaluated on the HMBPP reductase. The difluorinated phosphonate **95** is a poor substrate ($k_{cat} = 0.022\ min^{-1}$, 500-fold lower than that of the normal substrate HMBPP **31**) and is converted into the difluorinated IPP analogue **96** (**Figure 16**). Even if the difluorophosphonate **95** is a substrate with an activity 20-fold lower than that of HMBPP, this experiment is mechanistically important,

suggesting that the catalyzed reaction does not involve a C-1 protonation/deprotonation sequence (**Figure 16**, mechanism B).

1.13.3 Distribution of the MEP Pathway

1.13.3.1 Distribution of the MEP Pathway in Prokaryotes

The MEP pathway was discovered while investigating isoprenoid biosynthesis in selected bacteria.[8] First, data on the distribution were obtained from feeding experiments with ^{13}C-labeled precursors, mainly glucose.[165] Once the genes of the MEP pathway were known and more and more bacterial genomes were fully sequenced, a bioinformatic approach, searching for the MEP pathway genes, afforded a much better overview on the respective distribution of the MVA and the MEP pathways.[160,166,167] Another approach utilized the effect on carotenoid biosynthesis of pathway-specific inhibitors, either fosmidomycin for the MEP pathway, or mevinolin for the MVA pathway. Inhibition by fosmidomycin (mevinolin had no effect) suggested that the Microbacteriaceae synthesize these pigments through the MEP pathway.[168] The MEP pathway is widely distributed among Eubacteria and is the predominant pathway. It is thus easier to list the prokaryotes utilizing the MVA pathway: Myxobacteria such as *Myxococcus fulvus*[165] and *M. xanthus*,[160] *Borrelia burgdorferi*, *Legionella pneumophila*, many Gram-positive (but by far not all of them) such as *Enterococcus*, *Streptococcus*, *Staphylococcus*, and *Lactobacillus*,[160] and finally all investigated Archaea.[160,169]

Few Eubacteria possess both pathways. Some, if not most, *Streptomyces* utilize exclusively the MEP pathway for the biosynthesis of their isoprenoids. Some have both pathways and utilize the MEP pathway for the biosynthesis of essential isoprenoids, for example, the prenyl chain of the menaquinones, during the exponential growth phase, and the MVA pathway for the formation of isoprenoids issued from secondary metabolism and of the prenyl moiety of antibiotics of mixed metabolic origin.[170–172] There is thus a differentiation at the isoprenoid level between primary metabolism (MEP pathway) and secondary metabolism (MVA pathway) in these microorganisms. *Listeria monocytogenes* and *L. innocua* possess both MVA and MEP pathways. Both are functional but dispensable for viability of the cells, and disruption of the MVA pathway results in an upregulation of the MEP pathway.[173] In all other prokaryotes, the two pathways are mutually exclusive.

1.13.3.2 Presence of the MEP Pathway in Phototrophic Phyla

1.13.3.2.1 Higher plants, ferns, and mosses

Incorporation of ^{13}C-labeled glucose into the diterpenoids of *G. biloba* embryos shed light for the first time on the presence of the alternative MEP pathway in plants.[10,11] The dichotomy found in plants at the level of the biosynthesis of isoprenoids was also pointed out for the first time in this work: the cytosolic MVA pathway leading as expected to the sterols and the plastidial MEP pathway yielding the diterpenoids. It is surprising that the MEP pathway was not discovered earlier. Incorporation attempts performed with labeled acetate, the precursor of the MVA pathway, in plant systems failed to significantly label plastid isoprenoids, plant mono-, and diterpenoids. Plant systems do not easily convert acetate into pyruvate **24** and GAP **25**, the first two precursors of the MEP pathway (**Figure 4**), especially if this carbon source is in competition with carbohydrates, or simply CO_2 that is easily and rapidly incorporated into the precursors of the MEP pathway. The simple replacement of acetate by glucose permitted differentiating between the two pathways and opened the way for a whole series of labeling experiments that definitively solved old problems related to plant isoprenoid biosynthesis.[5] Are thus essentially derived from MEP pathway hemiterpenes (C_5) derived from DMAPP (isoprene in many plants,[174] the C_5 prenyl moiety of umbelliferone from *Apium graveolens*,[175] humulone,[176] and xanthohumol from hops *Humulus lupulus*,[177] the dimethylallyl moiety of glabrol in hairy root cultures of *Glycyrrhiza glabra*,[178] of the *Cinchona* 'Robusta' anthraquinone,[179] of hyperforin, a phlorogucinol derivative from *Hypericum perforatum*,[180] and the benzofurane derivatives of *Tagetes patula*,[181] 2-methyl-3-buten-2-ol from needles of *Pinus ponderosa*,[182] the hemiterpene moiety of isoprenyl phenyl ethers of the liverwort *Trichocolea tomentella*,[183] the two isoprene units of the lavandulyl group of sophoraflavone and the two dimethylallyl groups of lupalbigenin from *Sophora flavescens* cultured cells[184]), monoterpenes (C_{10}) derived from geranyl diphosphate (geraniol, menthone, pulegone, thymol from several essential oils,[185] bornyl acetate in the liverwort

Conocephalum conicum,[186] linalyl acetate in *Mentha citrata*,[187] the monoterpene moiety of isoprenyl phenyl ethers of the liverwort *T. tomentella*[183] and of cannabinoids in *Cannabis sativa*,[188] the monoterpene iridoid glucoside secologanin in a *C. roseus* cell culture,[189] loganin from *Rauwolfia serpentina*[190] and the indole alkaloid monoterpene-secologanine moiety of camptothecin in the hairy roots of *Ophiorrhiza pumila*,[191] artemisiaketone and other monoterpenes in *Tanacetum vulgare*,[192] the chrysanthemic acid moiety of pyrethrin I of *Chrysanthemum cinerariaefolium*,[193] monoterpenes from two carrot varieties[194]), diterpenoids (C_{20}) (taxol from a cell culture of *Taxus chinensis*,[195] labdane diterpenoids in *Marrubium vulgare*,[196] bicyclic and tetracyclic diterpenes of *Scoparia dulcis*,[197] a neo-*epi*-verrucosane diterpene, a sacculatadienedial and phytol in the liverwort *Fossombronia alaskana*,[198,199] steviol in *Stevia rebaudiana*,[200] abscissic acid and lipiferolide in *Liriodendron tulipifera*,[201] phytol in *Tanacetum vulgare*,[192] salvinorin A, a hallucinogenic clerodane diterpene from *Salvia divinorum*[202]) and carotenoids (C_{40}) derived from geranylgeranyl diphosphate (apocarotenoids from arbuscular mycorrhizal roots),[20] and the C_{45} all-*trans* polyprenyl chains of solanesol[203] and plastoquinone.[12] In contrast, sesquiterpenes (C_{15}) and triterpenoids (C_{30}), including sterols, derived from farnesyl diphosphate, and the prenyl chain of ubiquinone in tobacco BY-2 cell cultures are synthesized by the MVA pathway.[204] Exceptions are, however, known. The monoterpene geranyl moiety of shikonin from *Lithospermum erythrorhizon* revealed a labeling pattern consistent with the MVA pathway upon feeding with [1-^{13}C]glucose.[205] The contribution of the two pathways to the isoprenoid production is, however, not as clear-cut. Incorporations of [1-^{13}C]glucose have shown that some sesquiterpenes are derived from the MEP pathway: for example, germacrene D in *Solidago canadensis*,[206] the sesquiterpene epoxide hodgsonox from the liverwort *Lepidolaena hodgsoniae*[207] and anthecotuloid, an irregular sesquiterpene lactone from *Anthemis cotulata*.[208] In addition, cross-talk between the cytosolic and the plastidial compartment regularly occurs, leading to labeling patterns of mixed origin, suggesting new possibilities for the regulation of the production of isoprene units in plants (see Section 1.13.4.5).

Confirmation of the localization of the MEP pathway in the plant plastids was obtained by analyzing the nucleotide sequences of the genes. All MEP pathway genes are characterized in plants by the presence of a putative plastid-targeting sequence of variable length.[53] In fact, the MEP pathway of the plastids can be considered as a relictual bacterial metabolic pathway brought into a eukaryotic cell by the cyanobacterial symbiont at the origin of the chloroplasts. The MEP pathway was characterized quite early in the cyanobacterium *Synechocystis* PCC6714.[209] Using a green fluorescent protein assay, it was shown that all MEP pathway enzymes are localized in the chloroplasts of *A. thaliana*.[210]

MEP and MVA pathways can be differentiated by the labeling patterns of the isoprene units after the feeding of ^{13}C-labeled glucose. The method is sure, but requires sufficient amounts of terpenoids to run NMR spectra, and is accordingly time-consuming. They can also be differentiated by incubation with a more specific precursor such as labeled DX for the MEP pathway, or MVA for the MVA pathway. Such methods based on the feeding of an external precursor are, however, open to criticisms. They do not correspond to physiological conditions and may affect the internal balance of the intermediates and perturb metabolic fluxes, especially in plants, which normally only depend on CO_2 and photosynthesis for their carbon source. As the two pathways utilize two different sequences of enzyme reactions, involving formation and cleavage of carbon–carbon and carbon–hydrogen bond with different thermodynamic and kinetic isotope effects, it should be possible to differentiate them by their $^{12}C/^{13}C$ or $^{1}H/^{2}H$ natural isotope ratios. In the Lima bean, *Phaseolus lunatus*, 4,8-dimethylnona-1,3,7-triene (−37.4), a sesquiterpene derivative synthesized through the MVA pathway is thus ^{13}C-depleted as compared to ocimene (−29.0), a monoterpene synthesized through the MEP pathway.[211] Carbon and hydrogen isotope fractionation was also observed in the lipid biosynthesis of *Cryptomeria japonica* with differentiation of the isotope signatures of MEP and MVA pathways.[212]

1.13.3.2.2 Algae and related nonphototrophic phyla

First proof for the presence of the MEP pathway in the unicellular green alga *Scenedesmus obliquus* was obtained after feeding the cells with several ^{13}C-labeled glucose isotopomers.[213] Surprisingly, all investigated isoprenoids including sterols, phytol, carotenoids, and the prenyl chains of plastoquinone were solely derived from the MEP pathway. Similar investigations performed with [1-^{13}C]glucose confirmed the presence of the sole MEP pathway in two other unicellular green algae, *Chlorella fusca* and *C. reinhardtii*.[209] A series of labeling experiments performed later with [1-$^{2}H_1$]DX and DL-[2-^{13}C]MVA and investigation by mass spectrometry of the resulting labeling patterns of sterols and phytol were consistent with the sole presence of the MEP pathway

in representatives of four classes of the Chlorophyta (Chlorophyceae, Ulvophyceae, Trebouxiophyceae, Prasinophyceae, which are in its present acceptance not a monophyletic clade).[214] The MEP pathway is also involved in the biosynthesis of the acyclic triterpenes, botryococcenes, and tetramethylsqualene in the green alga *Botryococcus braunii* race B.[215] Using the same labeling method, a mixed origin was found for the isoprenoids of the Charophyceae lineage (*Spirogyra* sp., *Klebsophormidium flaccidum*, *Mesostigma viride*), which is phylogenetically related to higher plants.[214] The absence of the cytosolic MVA pathway has been now confirmed by the completely sequenced genome of *C. reinhardtii*, Chlorophyta (green algae). Genome library screenings suggested that the MVA pathway is a characteristic feature of all phototrophic eukaryotes and was specifically lost in a common ancestor of the Chlorophyta.[216]

The same dichotomy as that observed in higher plants was found in the red alga *Cyanidium caldarium* and in the Chrysophyte *Ochromonas danica* upon feeding of ^{13}C-labeled glucose isotopomers[209] and in the diatoms *Phaeodactylum tricornutum* and *Nitzschia ovalis* upon feeding with either [1-^{13}C]acetate, [2,2-$^{5}H_2$]DX or $^{13}CO_2$ (the latter being the best carbon source for these algae).[217] Cytoplasmic sterols were synthesized through the MVA pathway and plastid isoprenoids through the MEP pathway.

The Euglenophyte *Euglena gracilis*, grown in heterotrophic conditions in the presence of [1-^{13}C]glucose, synthesizes ergosterol and phytol through the MVA pathway.[209] In contrast, a subsequent study, utilizing either [1-^{13}C]glucose or [5,5-$^{2}H_2$]DX, suggested that the MEP pathway contributes to carotenoid biosynthesis in *E. gracilis* and not to the formation of phytol, which is obtained from the MVA pathway.[218]

Plastids are widespread in plant and algal lineages. Cryptic plastids are present in some nonphotosynthesizing unicellular eukaryotes that are phylogenetically related to phototrophic organisms. Search for MEP pathway genes in apicomplexan parasites resulted in the identification of the DXR gene in *Plasmodium* species responsible for malaria[219] and later of all other genes of the MEP pathway.[220,221] The MEP pathway is thus functionally in all intra-erythrocytic life stages of *P. falciparum*.[222] These genes were also found in other apicomplexan parasites such as *Eimeria tenella* and *Toxoplasma gondii*, responsible for toxoplasmosis, which lacked in addition to the MVA pathway genes, indicating that only the MEP pathway is operative in these organisms.[223,224]

Feeding with [1-^{13}C]glucose revealed that ergosterol and cycloartenol are synthesized through the MEP pathway in the nonphototrophic yeast-like alga *Prototheca wickerhamii*, suggesting its close relationship with the unicellular green algae.[225] *Perkinsus marinus*, a parasite involved in mass mortality in oyster farms and phylogenetically related to plastid-bearing dinoflagellates and apicomplexans, also possesses all genes of the MEP pathway.[226]

1.13.4 Openings and Further Developments

1.13.4.1 The MEP Pathway: A Target for Antimicrobial Drugs

Only the MVA pathway is present in humans and in animals for the biosynthesis of isoprene units. The alternative MEP pathway is present in most bacteria, including pathogenic bacteria and opportunistic pathogens, responsible for many nosocomial diseases. The generalized use of antibiotics during the past decades was followed by the regular appearance of resistance phenomena, which are now known for all major classes of antibiotics utilized in human therapeutics. The search for new targets against pathogenic microorganisms is thus an obvious necessity. The MEP pathway represented an unexplored field.[97,227–229]

The discovery that fosmidomycin, a known, but unexploited antibiotic, inhibited the second step of the MEP pathway and blocked the growth of many bacteria[61] and of *Plasmodium* spp.[219] validated the concept: early steps of isoprenoid biosynthesis through the MEP pathway can be considered as an interesting field. Fosmidomycin **38** and FR900098 **39** (**Figure 8**) blocked the growth of multidrug-resistant *P. falciparum* strains and cured mice infected with the rodent parasite *P. vinckei*. Application of the concept to human medicine indicated that fosmidomycin is a safe and effective treatment of uncomplicated malaria cases for treatments of 4 days or more.[230,231] Its use as single agent is, however, affected by unacceptable high rates of recrudescence. The search for combination partners with fosmidomycin pointed out synergies between fosmidomycin and the lincosamides, lincomycin, and essentially clindamycin. The fosmidomycin–clindamycin combination was found to be efficient in the *P. vinckei* mouse model.[220,232]

The numerous attempts to design inhibitors of MEP pathway enzymes, using either the fosmidomycin structure as a lead for DXR inhibitors, or docking studies as soon as three-dimensional X-ray structures were available for the other enzymes, have been described in earlier sections. Despite large screening programs[233] and many attempts to set up fast enzymatic tests designed for high-throughput screening,[234] no MEP pathway inhibitor is presently in clinical use. It is certainly premature to draw conclusions, but the main problems encountered essentially with fosmidomycin and fosmidomycin derivatives are representative of the major trends explored in the past few years and are discussed below.

Even if fosmidomycin is well tolerated by humans, its efficiency as an antimicrobial agent is not as interesting as expected. The antibiotic is characterized by a high clearance and is rapidly eliminated in the urine.[235] Moreover, resistance against this antibiotic appears rapidly in bacteria by a general mechanism. As a small anionic molecule, it is easily pumped out of the cells. A gene conferring the resistance to fosmidomycin was cloned from wild-type *E. coli*. It is involved in the efflux of the antibiotic.[236] Adenylate cyclase *E. coli* mutants were also found to be resistant to fosmidomycin, and to fosfomycin, another phosphonate antibiotic, indicating that both antibiotics are transported into the cells through the GAP transporter.[237]

The lack of penetration into the target cells is described as another mechanism for fosmidomycin resistance. Although the *M. tuberculosis* DXR is inhibited by fosmidomycin ($IC_{50} = 310\,nmol\,l^{-1}$), this bacterium is not sensitive to this antibiotic.[43] The resistance is apparently due to a lack of uptake of the antibiotic by the bacteria. Both *M. tuberculosis* and *M. smegmatis* are resistant to high concentrations of the antibiotic and do not accumulate intracellularly fosmidomycin, which remains in the culture medium.[238]

Only one report on the screening of natural products libraries is precisely described. Screening of several Mediterranean plant extracts on the *E. coli* DXR activity pointed out a strong antibacterial activity for the *Cercis siliquastrum* extract.[239]

1.13.4.2 Overproduction of MEP Pathway-Derived Isoprenoids in Bacteria and Plants

In most bacteria, isoprenoids are synthesized through the MEP pathway. Overproduction of isoprenoid of economical interest has been accordingly investigated by manipulating the expression of MEP pathway genes, especially those that are believed to encode rate-limiting enzymes such as the DXP synthase. In a transformed *E. coli* engineered for β-carotene and zeaxanthin production, overexpression of the *dxs*, *dxr*, and *idi* genes, respectively, encoding DXP synthase, DXP reducto-isomerase, and the IPP isomerase results in an up to 3.5-fold stimulation of carotenoidogenesis. Cotransformation with *idi* overexpression with either *dxs* or *dxr* has an additive effect yielding carotenoid concentrations up to $1.6\,mg\,g^{-1}$ (cell dry weight).[240] In other *E. coli* strains, engineered for carotenoid production, overexpression of the DXP synthase gene from *Bacillus subtilis* or *Synechocystis* sp. PCC 6803 alone induced increased lycopene and ubiquinone-8 levels as compared to controls,[241] and simultaneous overexpression of the *E. coli* DXP synthase and the *Archaeoglobus fulgidus* geranylgeranyl diphosphate synthase genes afforded large amounts of lycopene ($41\,mg\,g^{-1}$, dry weight).[242] A recombinant *E. coli* strain transformed with the DXP synthase gene from *Ps. aeruginosa* and the decaprenyl diphosphate synthase gene from *Gluconobacter suboxydans* was characterized by a twofold increase in ubiquinone-10 concentration, reaching $46\,mg\,l^{-1}$.[243]

Enhanced isoprenoid production was also attempted in plants. Upregulation of the DXP synthase or of the DXR activity, for example, in *Arabidopsis* plants, induces an increased accumulation of MEP pathway-derived isoprenoids.[244] Metabolic engineering of tomato plants has been performed by transformation with the bacterial *E. coli dxs* gene targeted to the plastids by the tomato *dxs* transit sequence. The resulting transgenic lines are characterized by a 1.6-fold increase of the carotenoid content with no changes in the cytosolic isoprenoid content.[245] The essential oil of the spike lavender *Lavandula latifolia* is essentially composed of monoterpenes synthesized through the MEP pathway. An *A. thaliana* cDNA of the DXP synthase was constitutively expressed in spike lavender. Upregulation of this enzyme results in a spectacular enhancement of the essential oil content: up to 360 and 74% in flowers and leaves, respectively, as compared to controls.[246] The constitutive expression of the Arabidopsis HMBPP reductase, the last enzyme of the MEP pathway, in tomato does not increase the carotenoid concentration in etioplasts, but light-grown transgenic plants have higher carotenoid levels.[247] Double transgenic *A. thaliana* lines overproducing the taxadiene synthase from *Taxus baccata* (cyclizing

geranylgeranyl diphosphate) and the DXP synthase or the HMBPP reductase show a twofold stronger effect on taxadiene production for HMBPP reductase overexpression as compared to DXP synthase overexpression.[247]

1.13.4.3 Isoprenoid Biosynthesis in Plants: MVA Versus MEP Pathway: Cross-Talk between the Cytosolic and the Plastidial Compartments

The allocation of the origin of isoprene units, MVA pathway for the cytosolic isoprenoids and MEP pathway for the plastidial isoprenoids, is rarely as strict as described. Evidence for exchanges of prenyl alcohols diphosphates between the two cellular compartments has been repeatedly reported. Already in the first study on ginkgo embryos,[11] the C_{20} diterpene skeleton of ginkgolides, derived from the acyclic geranylgeranyl diphosphate, was shown to be of mixed origin. If the bulk (~95%) of the C_{20} skeleton was effectively derived from MEP pathway, a significant part was assembled from a C_{15} moiety (corresponding to farnesyl diphosphate) with isoprene units derived from MVA pathway completed by a fourth isoprene unit derived from the MEP pathway. A similar dual origin was observed for the formation of the diterpene skeletons of phytol with an involvement of the MVA pathway in the synthesis of the farnesyl diphosphate-derived moiety of phytol, which is completed by a fourth unit essentially derived from the MEP pathway, in several liverworts, *Heteroscyphus planus*,[248,249] *Lophocolea heterophylla*,[249] and *Ptychanthus striatus*[250] and in the hornwort *Acanthoceros punctatus*[251] or for the FPP moiety of hinokiol, an abietane diterpene, in *Torreya nucifera*.[252] Isoprenoid skeletons with isoprene units of dual origin have been later reported in many series.

In *C. roseus* cell cultures, upon feeding of [1-^{13}C]glucose and careful analysis of the terpenoid NMR spectra, a significant contribution of the MEP pathway (~6%) was found for the formation of the sitosterol skeleton normally derived from the MVA pathway.[253,254] The MEP pathway contribution to the phytosterol biosynthesis was even of the same magnitude as that of the MEP pathway in *Croton sublyratus*.[255] The isotope distribution upon feeding of ^{13}C-labeled glucose revealed a mixed origin of the isoprene units in other series. Upon feeding with [U-$^{13}C_6$]glucose, the phytol of *C. roseus* cell culture was shown to be derived from both MEP (~60%) and MVA pathways (~40%).[254] The C_{15} sesquiterpene skeleton of bisaboloxide A and chamazulene from *Matricaria recutita* is formed from two isoprene units predominantly derived from the MEP pathway and a third unit of mixed origin, being derived from both MEP and MVA pathways.[256] The sesquiterpene β-caryophyllene in carrot roots,[194] the sesquiterpene blend from vine leaves,[257] and the nerolidyl moiety of 4-nerolidylcatechol from *Potomorphe umbellata*[258] have also a dual origin as shown from the labeling patterns obtained upon feeding with ^{2}H- or ^{13}C-labeled MVA, DX, or glucose. In the monoterpene series, a dual origin for isoprene units has been found for geraniol and (*S*)-linalool from grape berries[259] and (*S*)-linalool in the leaves of the strawberry *Fragaria* × *ananassa*.[260] In *Piper gaudichaudianum*, the hemiterpene dimethylallyl moiety of gaudichaudianic acid, a *p*-hydroxybenzoate derivative, is derived from MEP and from MVA pathway as shown by the incorporation of [1-^{13}C]glucose whereas the geranyl diphosphate-derived monoterpene chain results from the condensation of DMAPP derived from both MVA and DMAPP pathways with an IPP unit essentially made through the MEP pathway.[261] Finally, the dolichols, derived from all-*cis* polyprenols, shown in a hairy root culture from *Coluria geoides* were also shown to be of mixed origin after incorporation of ^{13}C-labeled glucose as a general precursor or ^{2}H-labeled MVA or DXP as pathway-specific precursors: statistical analysis of the ^{13}C distribution by mass spectrometry and NMR spectroscopy suggests that in the prenyl chains up to 13 isoprene units are synthesized in the plastids from IPP derived either from MVA and from MEP pathways and are completed in the cytoplasm with units solely derived from MVA pathway.[262]

In tobacco Bright Tellow-2 cell cultures, the two pathways complement one another.[263] Treatment in the low millimolar range with mevinolin, a potent inhibitor of the key enzyme of the MVA pathway, hydroxymethylglutaryl coenzyme A reductase, results in growth reduction that is alleviated by the addition of free DX. Incorporation studies with [1,1,1,4-$^{2}H_2$]DX show that not only plastoquinone as expected, but also sterols, normally derived from MVA pathway, are both synthesized through the MEP pathway in the presence of mevinolin. Growth inhibition by fosmidomycin, inhibiting DXR, the second enzyme of the MEP pathway, is restored by the addition of MVA to the culture medium; complementation with [2-^{13}C]MVA proves that in the presence of fosmidomycin even the prenyl chain of plastoquinone, a representative of the plastid isoprenoids, is derived from the MVA pathway. Best incorporation yields of the labeled precursors are obtained by

simultaneous treatment with both inhibitors, showing that at least in this plant system the two pathways can replace each other. Treatment of *A. thaliana* seedlings with the same two pathway-specific inhibitors pointed out the absence of correlation between gene-expression patterns and the accumulation of the end products of isoprenoid metabolic pathways, suggesting that posttranscriptional processes may be involved in the regulation of isoprenoid metabolism.[264] All these labeling experiments are consistent with the transport between the cytosol and plastids of C_5, C_{10}, and C_{15} allylic alcohol diphosphates. This hypothesis is supported by the fact that isolated chloroplasts from three plants (spinach, Indian mustard, and kale), envelope membrane vesicles, and proteoliposomes from solubilized proteins of spinach envelope membranes are capable of efficiently transporting IPP and geranyl diphosphate, and to a lower extent DMAPP and farnesyl diphosphate.[265] The weak incorporation of ^{14}C-labeled MVA that was reported in earlier studies on mono- and diterpene biosynthesis can now be attributed to metabolite exchanges in the frame of the cross-talk between the cytosolic and the plastidial plant cell compartments.[5]

1.13.4.4 IPP and DMAPP Production in Bacteria and Plants: HMBPP Reductase Versus IPP/DMAPP Isomerase

A most intriguing feature of the MEP pathway is the different labeling patterns obtained in isoprene units upon feeding DX with ^{2}H-labeling at C-4 **97** or ME with ^{2}H-labeling at C-3 **98** (**Figure 18**). Incorporation of [4-^{2}H]DX in an *E. coli* wild-type strain[266] or of [3,5,5,5-$^2H_4$]ME in an *E. coli* mutant with a deleted DXR gene[267] resulted in a different labeling in the isoprene units of the prenyl chains of ubiquinone and menaquinone: those derived from DMAPP, serving as starter substrate for the prenyl transferases, retain quantitatively the deuterium of the precursor, whereas those derived from IPP, serving for the elongation of the prenyl chains, do not show any deuterium content. A similar observation was made on the monoterpene skeleton of cineol upon feeding *Eucalyptus globulus* twigs with [4-^{2}H]DX.[268] A different labeling pattern was found in the isoprene units of carotenoids and phytol of a *C. roseus* cell culture grown in the presence of doubly labeled [2-^{13}C, 4-^{2}H]DX: presence of ^{13}C in all isoprene units shows that the precursor is effectively incorporated in all isoprene units that are all devoid of any ^{2}H content.[269] In contrast, feeding a tobacco Bright Yellow-2 cell

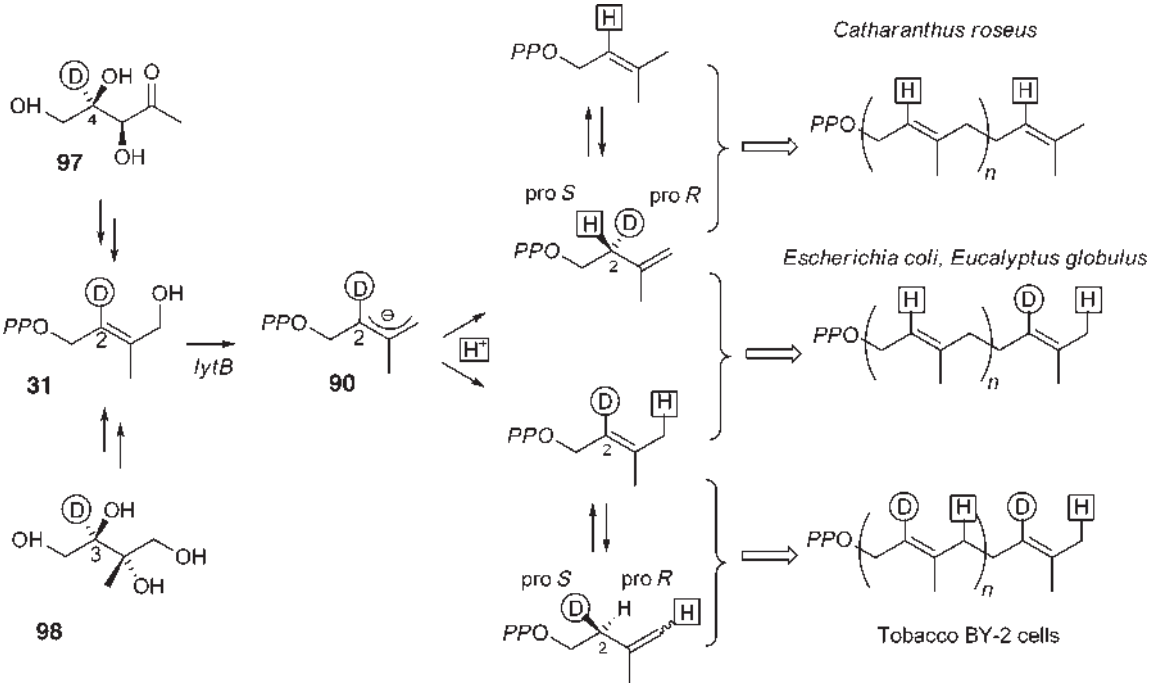


Figure 18 Plant isoprenoid biosynthesis: origin of isoprene units from isopentenyl diphosphate **2** or dimethylallyl diphosphate **3**. Reproduced from M. Rohmer, *Pure Appl. Chem*. **2007**, *78*, 739–751.

culture with [4-^{2}H]DX results in the same labeling of all isoprene units, whichever they are derived from IPP or from DMAPP. They are all deuterium labeled with the same isotope abundance (15%) corresponding to the IPP–DMAPP ratio produced by the HMBPP reductase.[270] These three, apparently contradictory, labeling patterns of isoprene units can, however, be explained in the frame of a single hypothetical biogenetic scheme (**Figure 18**). Taking into account that the pro-*S* C-2 hydrogen is eliminated on IPP by the IPP isomerase and by the *trans* prenyl transferase (**Figures 1** and **18**), the deuterium signature found in the labeling experiments described above corresponding to isoprene units issued from the DMAPP branch produced by the HMBPP reductase, and the absence of deuterium corresponding to isoprene units issued from the IPP branch. Such variations found in the origin of the isoprene units may reflect a way for modulating the origin of isoprene units that deserves more detailed investigations.

1.13.4.5 Light Influence on the MEP Pathway in Plants

Since the discovery of the MEP pathway, the regulation of the biosynthesis of plastid isoprenoids has been regularly investigated, but is by far not-yet elucidated. Significant roles in the regulation have been thus pointed out for the DXP synthase,[53] the DXP reducto-isomerase,[244] and the HMBPP reductase.[247]

Circadian rhythms have been discovered for isoprenoid biosynthesis through the MEP pathway. The emission of isoprene or 2-methyl-3-buten-2-ol, that is, both produced through the MEP pathway, is light dependent in all tested plant species, correlated with the diurnal variations of the DMAPP content of the cells and partly due to controls over DMAPP production.[271] The diurnal fluctuations observed in the emission of isoprene in Eastern cottonwood *Populus deltoides* and red oak *Quercus rubra*[272] and in popular *Populus* × *canescens* leaves[273] likely results from a rhythmicity of the MEP pathway. In the flowers of snapdragon *Antirrhinum majus*, the MEP pathway provides IPP precursors for both monoterpene (myrcene, ocimene, linalool) biosynthesies in the plastids and sesquiterpene (nerolidol) biosynthesis in the cytosol.[274] These biosynthetic processes operate in a rhythmic manner controlled by a circadian clock and a rhythmic gene expression with emission peaks of volatile terpenoids corresponding to maximal illumination. The DXP synthase transcripts, but not of the DXP reducto-isomerase, showed diurnal oscillations and were highly expressed during daytime. Cross-talk occurs in snapdragon petal epidermis unidirectionally from the plastids to the cytosol.

Activation of the MEP pathway is required in many plant defense mechanisms. Elicitor-induced production of diterpenic phytoalexins in rice (momilactones and phytocassanes) corresponds to an enhanced expression of the first three steps of the pathway catalyzed by the DXP synthase, the DXP reducto-isomerase, and 4-diphosphocytidyl ME synthase.[275] Treatment with 3-oxoclomazone, a DXS inhibitor, or fosmidomycin, a DXR inhibitor, represses the accumulation of the phytoalexins, suggesting that the MEP pathway is responsible for the sufficient production of these plant defense metabolites. According to the feeding of ^{2}H-labeled DX and the administration of pathway-specific inhibitors (fosmidomycin or lovastatin), volatile sesquiterpenes related to herbivore damages in *M. truncatula* are derived from plastidial MEP pathway and from cytosolic MVA pathway.[276] In the Lima bean, *P. lunatus*, photosynthesis is the major carbon source for the MEP pathway.[277] Leaves damaged by feeding of *Spodoptera littoralis* release in the dark phase (*Z*)-3-hexenyl acetate, whereas the monoterpene β-ocimene, synthesized through the MEP pathway,[278] is only emitted within minutes after the onset of light.

Acknowledgments

The work was partly supported by the 'Institut Universitaire de France' and by the 'Agence Nationale de la Recherche' (grants Nb ANR-05-BLAN-0217 and Nb ANR-06-BLAN-0291).

Abbreviations

CDP-ME	4-diphosphocytidyl-2-*C*-methyl-D-erythritol
DMAPP	dimethylallyl diphosphate
DX	1-deoxy-D-xylulose

DXP	1-deoxy-D-xylulose 5-phosphate
DXR	1-deoxy-D-xylulose 5-phosphate reducto-isomerase
GAP	D-glyceraldehyde 3-phosphate
HMBPP	4-hydroxy-3-methylbut-2-enyl 1-phosphate
IPP	isopentenyl diphosphate
ME	2-*C*-methyl-D-erythritol
MEcDP	2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate
MEP	2-*C*-methyl-D-erythritol 4-phosphate
MVA	mevalonate

References

1. J. D. Connolly; R. A. Hill, *Dictionary of Terpenoids*; Chapman & Hall: New York, 1992.
2. J. C. Sacchetini; C. D. Poulter, *Science* **1997**, *277*, 1788–1789.
3. K. Bloch, *Steroids* **1992**, *57*, 378–383, and references cited therein.
4. D. A. Bochar; J. A. Friesen; C. V. Stauffacher; V. W. Rodwell, Biosynthesis of Mevalonic Acid from Acetyl-CoA. In *Comprehensive Natural Products Chemistry*; D. Cane, Ed.; Pergamon Press: Oxford, 1999; Vol. 2, pp 15–44, and references cited therein.
5. M. Rohmer, A Mevalonate-Independent Route to Isopentenyl Diphosphate. In *Comprehensive Natural Products Chemistry*; D. Cane, Ed.; Pergamon Press: Oxford, 1999; Vol. 2, pp 45–67, and references cited therein.
6. D. E. Cane; T. Rossi; A. M. Tillman; J. P. Pachlatko, *J. Am. Chem. Soc.* **1981**, *103*, 1838–1843.
7. G. Flesch; M. Rohmer, *Eur. J. Biochem.* **1988**, *175*, 405–411.
8. M. Rohmer; M. Knani; P. Simonin; B. Sutter; H. Sahm, *Biochem. J.* **1993**, *295*, 517–524, and references cited therein.
9. M. Rohmer; M. Seemann; S. Horbach; H. Sahm, *J. Am. Chem. Soc.* **1996**, *118*, 2564–2566.
10. M. Schwarz, Terpen-Biosynthese in *Ginkgo Biloba*: Eine Überraschende Geschichte. Ph.D. Thesis, Nb ETH 10951, Eidgenössische Technische Hochschule, Zurich, Switzerland, 1994.
11. M. Schwarz; D. Arigoni, Gingkolide biosynthesis. In *Comprehensive Natural Products Chemistry*; D. Cane, Ed.; Pergamon Press: Oxford, 1999; Vol. 2, pp 367–400, and references cited therein.
12. H. K. Lichtenthaler; J. Schwender; A. Disch; M. Rohmer, *FEBS Lett.* **1997**, *400*, 271–274.
13. J. Wungsintaweekul; S. Herz; S. Hecht; W. Eisenreich; R. Feicht; F. Rohdich; A. Bacher; M. H. Zenk, *Eur. J. Biochem.* **2001**, *268*, 310–316.
14. A. Hemmerlin; D. Tritsch; M. Hartmann; K. Pacaud; J. F. Hoeffler; A. van Dorsselaer; M. Rohmer; T. J. Bach, *Plant Physiol.* **2006**, *142*, 441–457.
15. C. A. Testa; R. M. Cornish; C. D. Poulter, *J. Bacteriol.* **2004**, *186*, 473–480.
16. G. A. Sprenger; U. Schörken; T. Wiegert; S. Grolle; A. A. De Graaf; S. V. Taylor; T. D. Begley; S. Bringer-Meyer; H. Sahm, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12857–12862.
17. L. M. Lois; N. Campos; S. Rosa-Putra; K. Danielsen; M. Rohmer; A. Boronat, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2105–2110.
18. B. M. Lange; M. R. Wildung; D. McCaskill; R. Croteau, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2100–2104.
19. M. H. Walter; J. Hans; D. Strack, *Plant J.* **2002**, *31*, 243–254.
20. M. H. Walter; D. S. Floss; J. Hans; T. Fester; D. Strack, *Phytochemistry* **2007**, *68*, 130–138.
21. M. A. Phillips; M. H. Walter; S. G. Ralph; P. Dabrowska; K. Luck; E. M. Urós; W. Boland; D. Strack; M. Rodríguez-Concepción; J. Bohlmann; J. Gershenzon, *Plant Mol. Biol.* **2007**, *65*, 243–257.
22. J. K. Lee; D. K. Oh; S. Y. Kim, *J. Biotechnol.* **2007**, *128*, 555–566.
23. A. M. Bailey; S. Mahapatra; P. J. Brennan; D. C. Crick, *Glycobiology* **2002**, *12*, 813–820.
24. L. M. Eubanks; C. D. Poulter, *Biochemistry* **2003**, *42*, 1140–1149.
25. S. Xiang; G. Usunow; G. Lange; M. Busch; L. Tong, *J. Biol. Chem.* **2007**, *282*, 2676–2682.
26. J. Querol; M. Rodriguez-Concepcion; A. Boronat; S. Imperial, *Biochem. Biophys. Res. Commun.* **2001**, *289*, 155–160.
27. M. Schürmann; M. Schürmann; G. A. Sprenger, *J. Mol. Catal. B Enzym.* **2002**, *19–20*, 247–252.
28. J. Querol; C. Grosdemange-Billiard; M. Rohmer; A. Boronat; S. Imperial, *Tetrahedron Lett.* **2002**, *43*, 8265–8268.
29. B. Altincicek; M. Hintz; S. Sanderbrand; J. Wiesner; E. Beck; H. Jomaa, *FEMS Microbiol. Lett.* **2000**, *190*, 329–333.
30. J. Mao; H. Eoh; R. He; Y. Wang; B. Wan; S. G. Franzblau; D. C. Crick; A. P. Kozikowski, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5320–5323.
31. C. Mueller; J. Schwender; J. Zeidler; H. K. Lichtenthaler, *Biochem. Soc. Trans.* **2000**, *28*, 792–793.
32. J. Zeidler; J. Schwender; C. Mueller; H. K. Lichtenthaler, *Biochem. Soc. Trans.* **2000**, *28*, 796–798.
33. T. Kuzuyama; S. Takahashi; H. Watanabe; H. Seto, *Tetrahedron Lett.* **1998**, *39*, 4509–4512.
34. S. Takahashi; T. Kuzuyama; H. Watanabe; H. Seto, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9879–9884.
35. J. H. Park; P. C. Dorrestein; H. Zhai; C. Kinsland; F. W. McLafferty; T. P. Begley, *Biochemistry* **2003**, *42*, 12430–12438.
36. D. E. Cane; S. Du; J. K. Robinson; Y. Hsiung; I. D. Spenser, *J. Am. Chem. Soc.* **1999**, *121*, 7722–7723.
37. S. Rosa-Putra; L. Charon; K. Danielsen; C. Pale-Grosdemange; L. M. Lois; N. Campos; A. Boronat; M. Rohmer, *Tetrahedron Lett.* **1998**, *39*, 6185–6188.
38. T. Radykewicz; F. Rohdich; J. Wungsintaweekul; S. Herz; K. Kis; W. Eisenreich; A. Bacher; M. H. Zenk; D. Arigoni, *FEBS Lett.* **2000**, *465*, 156–160.
39. P. J. Proteau; Y. H. Woo; T. Williamson; C. Phaosiri, *Org. Lett.* **1999**, *1*, 921–923.

40. J. F. Hoeffler; D. Tritsch; C. Grosdemange-Billiard; M. Rohmer, *Eur. J. Biochem.* **2002**, *269*, 4446–4457.
41. A. T. Koppisch; D. T. Fox; B. S. J. Blagg; C. D. Poulter, *Biochemistry* **2002**, *41*, 236–243.
42. A. Argyrou; J. S. Blanchard, *Biochemistry* **2004**, *43*, 4375–4384.
43. R. K. Dhiman; M. L. Schaeffer; A. M. Bailey; C. A. Testa; H. Scherman; D. C. Crick, *J. Bacteriol.* **2005**, *187*, 8395–8402.
44. L. M. Henrikson; C. Björkelid; S. L. Mowbray; T. Unge, *Acta Crystallogr. D Biol. Crystallogr.* **2006**, *D62*, 807–813.
45. D. E. Cane; C. Chow; A. Lillo; I. Kang, *Bioorg. Med. Chem. Lett.* **2001**, *9*, 1467–1477.
46. X. Yin; P. J. Proteau, *Biochim. Biophys. Acta* **2003**, *1652*, 75–81.
47. S. Grolle; S. Bringer-Meyer; H. Sahm, *FEMS Microbiol. Lett.* **2000**, *191*, 131–137.
48. D. Arigoni; J. L. Giner; S. Sagner; J. Wungsintaweekul; M. H. Zenk; K. KIs; A. Bacher; W. Eisenreich, *Chem. Commun.* **1999**, 1127–1128.
49. D. T. Fox; C. D. Poulter, *Biochemistry* **2005**, *44*, 8360–8368.
50. U. Wong; R. J. Cox, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 4926–4929.
51. J. W. Munos; X. Pu; S. O. Mansoorabadi; H. J. Kim; H. W. Liu, *J. Am. Chem. Soc.* **2009**, *131*, 2048–2049.
52. F. Rohdich; S. Lauw; J. Kaiser; R. Feicht; P. Köhler; A. Bacher; W. Eisenreich, *FEBS J.* **2006**, *273*, 4446–4458.
53. M. Rodríguez-Concepción; A. Boronat, *Plant Physiol.* **2002**, *130*, 1079–1089, and references cited therein.
54. S. Yajima; T. Nonaka; T. Kuzuyama; H. Seto; K. Ohsawa, *J. Biochem.* **2002**, *131*, 313–317.
55. A. Mac Sweeney; R. Lange; R. P. M. Fernandez; H. Schulz; G. E. Dale; A. Douanganath; P. J. Proteau; C. Oefner, *J. Mol. Biol.* **2005**, *345*, 115–127.
56. K. Reuter; S. Sanderbrand; H. Jomaa; J. Wiesner; I. Steinbrecher; E. Beck; M. Hintz; G. Klebe; M. Stubbs, *J. Biol. Chem.* **2002**, *277*, 5378–5384.
57. S. Steinbacher; J. Kaiser; W. Eisenreich; R. Hubert; A. Bacher; F. Rohdich, *J. Biol. Chem.* **2003**, *278*, 18401–18407.
58. S. Ricagno; S. Grolle; S. Bringer-Meyer; H. Sahm; Y. Lindqvist; G. Schneider, *Biochim. Biophys. Acta* **2004**, *1698*, 37–44.
59. T. Kuzuyama; S. Takahashi; M. Takagi; H. Seto, *J. Biol. Chem.* **2000**, *275*, 19928–19932.
60. R. P. M. Fernandes; P. J. Proteau, *Biochim. Biophys. Acta* **2006**, *1764*, 223–229.
61. R. P. M. Fernandes; C. Phaosiri; P. J. Proteau, *Arch. Biochem.* **2005**, *444*, 159–164.
62. Y. Shigi, *J. Antimicrob. Chemother.* **1989**, *24*, 131–145.
63. T. Kuzuyama; T. Shimizu; S. Takahashi; H. Seto, *Tetrahedron Lett.* **1998**, *39*, 7913–7916.
64. L. Mercklé; A. de Andrés-Gómez; B. Dick; R. J. Cox; C. R. A. Godfrey, *ChemBioChem* **2005**, *6*, 1866–1874.
65. J. Perruchon; R. Ortmann; M. Altenkämper; K. Silber; J. Silber; J. Wiesner; H. Jomaa; G. Klebe; M. Schlitzer, *ChemMedChem* **2008**, *3*, 1232–1241.
66. K. Hemmi; H. Takeno; M. Hashimoto; T. Kamiya, *Chem. Pharm. Bull.* **1982**, *30*, 111–118.
67. J. R. Walker; C. D. Poulter, *J. Org. Chem.* **2005**, *70*, 9955–9959.
68. T. Haemers; J. Wiesner; S. Van Poecke; J. Goeman; D. Henschker; E. Beck; H. Jomaa; S. Van Calenbergh, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1888–1891.
69. T. Haemers; J. Wiesner; R. Busson; H. Jomaa; S. Van Calenbergh, *Eur. J. Org. Chem.* **2006**, *17*, 3856–3863.
70. V. Devreux; J. Wiesner; H. Jomaa; J. Rozenski; J. Van der Eycken; S. Van Calenbergh, *J. Org. Chem.* **2007**, *72*, 3783–3789.
71. V. Devreux; J. Wiesner; H. Jomaa; J. Van der Eycken; S. Van Calenbergh, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 4920–4923.
72. V. Devreux; J. Wiesner; J. L. Goeman; J. Van der Eycken; H. Jomaa; S. Van Calenbergh, *J. Med. Chem.* **2006**, *49*, 2656–2660.
73. T. Haemers; J. Wiesner; D. Giessman; T. Verbruegghen; U. Hillaert; R. Ortmann; H. Jomaa; A. Link; M. Schlitzer; S. Van Calenbergh, *Bioorg. Med. Chem.* **2008**, *16*, 3361–3371.
74. D. Giessmann; P. Heidler; T. Haemers; S. Van Calenbergh; A. Reichenberg; H. Jomaa; C. Weidemeyer; S. Sanderbrand; J. Wiesner; A. Link, *Chem. Biodivers.* **2008**, *5*, 643–655.
75. Y. W. Woo; R. P. M. Fernandes; P. J. Proteau, *Bioorg. Med. Chem.* **2006**, *14*, 2375–2385.
76. L. Kuntz; D. Tritsch; C. Grosdemange-Billiard; A. Hemmerlin; A. Willem; T. J. Bach; M. Rohmer, *Biochem. J.* **2005**, *366*, 127–135.
77. A. Reichenberg; J. Wiesner; C. Weidemeyer; E. Dreiseidler; S. Sanderbrand; B. Altincicek; E. Beck; M. Schlitzer; H. Jomaa, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 833–835.
78. R. Ortmann; J. Wiesner; A. Reichenberg; D. Henschker; E. Beck; H. Jomaa; M. Schlitzer, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2163–2166.
79. R. Ortmann; J. Wiesner; A. Reichenberg; D. Henschker; E. Beck; H. Jomaa; M. Schlitzer, *Arch. Pharm. Chem.* **2005**, *338*, 305–314.
80. K. Schlüter; R. D. Walter; B. Bergmann; T. Kurz, *Eur. J. Med. Chem.* **2006**, *41*, 1385–1397.
81. T. Kurz; K. Schlüter; U. Kaula; B. Bergmann; R. D. Walter; D. Geffken, *Bioorg. Med. Chem.* **2006**, *14*, 5121–5135.
82. J. Zeidler; J. Schwender; C. Müller; J. Wiesner; C. Weidemeyer; E. Beck; H. Jomaa; H. K. Lichtenthaler, *Z. Naturforsch.* **1998**, *53c*, 980–986.
83. M. Fellermeier; K. Kis; S. Sagner; U. Maier; A. Bacher; M. H. Zenk, *Tetrahedron Lett.* **1999**, *40*, 2743–2746.
84. M. Courtois; Z. Mincheva; F. Andreu; M. Rideau; M. C. Viaud-Massuard, *J. Enzyme Inhib. Med. Chem.* **2004**, *19*, 559–565.
85. Z. Mincheva; M. Courtois; F. Andreu; M. Rideau; M. C. Viaud-Massuard, *Phytochemistry* **2005**, *66*, 1797–1803.
86. Z. Mincheva; M. Courtois; J. Crèche; M. Rideau; M. C. Viaud-Massuard, *Bioorg. Med. Chem.* **2004**, *12*, 191–197.
87. O. Meyer; C. Grosdemange-Billiard; D. Tritsch; M. Rohmer, *Org. Biomol. Chem.* **2003**, *1*, 4367–4372.
88. G. Hirsch; C. Grosdemange-Billiard; D. Tritsch; M. Rohmer, *Tetrahedron Lett.* **2004**, *45*, 519–521.
89. C. Phaosiri; P. J. Proteau, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5309–5312.
90. A. Wong; J. W. Munos; V. Devasthali; K. A. Johnson; H. W. Liu, *Org. Lett.* **2004**, *6*, 3625–3628.
91. O. Meyer; C. Grosdemange-Billiard; D. Tritsch; M. Rohmer, *Tetrahedron Lett.* **2007**, *48*, 711–714.
92. J. Munos; X. Pu; H. W. Liu, *Bioorg. Med. Chem. Lett.* **2007**, *18*, 3090–3094.
93. M. A. Phillips; P. León; A. Boronat; M. Rodríguez-Concepción, *Trends Plant Sci.* **2008**, *12*, 619–623.
94. F. Rohdich; J. Wungsintaweekul; M. Fellermeier; S. Sagner; S. Herz; K. Kis; W. Eisenreich; A. Bacher; M. H. Zenk, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11758–11763.
95. A. Follens; M. Veiga-Da Cunha; R. Merckx; E. Van Schaftingen; J. Van Eldere, *J. Bacteriol.* **1999**, *181*, 2001–2007.

96. F. Rohdich; J. Wungsintaweekul; W. Eisenreich; G. Richter; C. A. Schuhr; S. Hecht; M. H. Zenk; A. Bacher, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6451–6456.
97. H. Eoh; A. C. Brown; L. Buetow; W. N. Hunter; T. Parish; D. Kaur; P. J. Brennan; D. C. Crick, *J. Bacteriol.* **2007**, *189*, 8922–8927.
98. W. Shi; J. Feng; M. Zhang; X. Lai; S. Xu; X. Zhang; H. Wang, *J. Biochem. Mol. Biol.* **2007**, *40*, 911–920.
99. T. Kuzuyama; M. Takagi; K. Kaneda; T. Dairi; H. Seto, *Tetrahedron Lett.* **2000**, *41*, 703–706.
100. S. B. Richard; M. E. Bowman; W. Kwiatkowski; I. Kang; C. Chow; A. M. Lillo; D. E. Cane; P. Noel, *Nat. Struct. Biol.* **2001**, *8*, 641–648.
101. L. E. Kemp; C. S. Bond; W. N. Hunter, *Acta Crystallogr. D Biol. Crystallogr.* **2003**, *D59*, 607–610.
102. S. B. Richard; A. M. Lillo; C. N. Tetzlaff; M. E. Bowman; J. P. Noel; D. E. Cane, *Biochemistry* **2004**, *43*, 12189–12197.
104. H. Lüttgen; F. Rohdich; S. Herz; J. Wungsintaweekul; S. Hecht; C. A. Schuhr; M. Fellermeier; S. Sagner; M. H. Zenk; A. Bacher; W. Eisenreich, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1062–1067.
103. T. Kuzuyama; M. Takagi; K. Kaneda; H. Watanabe; T. Dairi; H. Seto, *Tetrahedron Lett.* **2000**, *41*, 2925–2928.
105. F. Rohdich; J. Wungsintaweekul; H. Lüttgen; M. Fischer; W. Eisenreich; C. A. Schuhr; M. Fellermeier; N. Schramek; M. H. Zenk; A. Bacher, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8251–8256.
106. L. Miallau; M. S. Alphey; L. E. Kemp; G. A. Leonard; S. M. McSweeney; S. Hecht; A. Bacher; W. Eisenreich; F. Rohdich; W. N. Hunter, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9173–9178.
107. T. Wada; T. Kuzuyama; S. Satoh; S. Kuramitsu; S. Yokoyama; S. Unzai; J. R. H. Tame; S. Y. Park, *J. Biol. Chem.* **2003**, *278*, 30022–30027.
108. A. K. H. Hirsch; S. Lauw; P. Gersbach; W. B. Schweizer; F. Rohdich; W. Eisenreich; A. Bacher; F. Diederich, *ChemMedChem* **2007**, *2*, 806–810.
109. A. K. H. Hirsch; M. S. Alphey; S. Lauw; M. Seet; L. Barandun; W. Eisenreich; F. Rohdich; W. N. Hunter; A. Bacher; F. Diederich, *Org. Biomol. Chem.* **2008**, *6*, 2719–2730.
110. C. M. Crane; A. K. H. Hirsch; M. S. Alphey; T. Sgraja; S. Lauw; V. Ilarionova; F. Rohdich; W. Eisenreich; W. N. Hunter; A. Bacher; F. Diederich, *ChemMedChem* **2008**, *3*, 91–101.
111. S. Herz; J. Wungsintaweekul; C. A. Schuhr; S. Hecht; H. Lüttgen; S. Sagner; M. Fellermeier; W. Eisenreich; M. H. Zenk; A. Bacher; F. Rohdich, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2486–2490.
112. D. L. Turner; H. Santos; P. Fareleira; I. Pacheco; J. Legall; A. V. Xavier, *Biochem. J.* **1992**, *285*, 387–390.
113. D. Ostrovsky; I. Shipanova; L. Sibeldina; A. Shashkov; E. Kharatian; I. Malyarova; G. Tantsyrev, *FEBS Lett.* **1992**, *298*, 159–161.
114. D. Ostrovsky; A. Shashkov; A. Sviridov, *Biochem. J.* **1993**, *295*, 901–902.
115. M. Fellermeier; M. Raschke; S. Sagner; J. Wungsintaweekul; C. A. Schuhr; S. Hecht; K. Kis; T. Radykewicz; P. Adam; F. Rohdich; W. Eisenreich; A. Bacher; D. Arigoni; M. Zenk, *Eur. J. Biochem.* **2001**, *268*, 6302–6310.
116. F. Rohdich; W. Eisenreich; J. Wungsintaweekul; S. Hecht; C. A. Schuhr; A. Bacher, *Eur. J. Biochem.* **2001**, *268*, 3190–3197.
117. M. Takagi; T. Kuzuyama; K. Kaneda; H. Watanabe; T. Dairi; H. Seto, *Tetrahedron Lett.* **2000**, *41*, 3395–3398.
118. S. Steinbacher; J. Kaiser; J. Wungsintaweekul; S. Hecht; W. Eisenreich; S. Gerhardt; A. Bacher; F. Rohdich, *J. Mol. Biol.* **2002**, *316*, 79–88.
119. S. B. Richard; J. L. Ferrer; M. E. Bowmann; A. M. Lillo; C. N. Tezlaff; D. E. Cane; J. P. Noel, *J. Biol. Chem.* **2002**, *277*, 8667–8672.
120. L. E. Kemp; C. S. Bond; W. N. Hunter, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6591–6596.
121. H. Kishida; T. Wada; S. Unzai; T. Kuzuyama; M. Takagi; T. Terada; M. Shirouzu; S. Yokoyama; J. R. H. Tame; S. Y. Park, *Acta Crystallogr. D Biol. Crystallogr.* **2003**, *D59*, 23–31.
122. S. Ni; H. Robinson; G. C. Marsing; D. E. Bussiere; M. A. Kennedy, *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *D60*, 1949–1957.
123. L. E. Kemp; M. S. Alphey; C. S. Bond; M. A. J. Ferguson; S. Hecht; A. Bacher; W. Eisenreich; F. Rohdich; W. N. Hunter, *Acta Crystallogr. D Biol. Crystallogr.* **2005**, *D61*, 45–52.
124. C. M. Crane; J. Kaiser; N. L. Ramsden; S. Lauw; F. Rohdich; A. Bacher; F. Diederich, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 1069–1074.
125. C. Baumgartner; C. Eberle; F. Diederich; S. Lauw; F. Rohdich; W. Eisenreich; A. Bacher, *Helv. Chim. Acta* **2007**, *90*, 1043–1068.
126. M. Gabrielsen; F. Rohdich; W. Eisenreich; T. Gräwert; S. Hecht; A. Bacher; W. N. Hunter, *Eur. J. Biochem.* **2004**, *271*, 3028–3035.
127. M. Gabrielsen; C. S. Bond; I. Hallyburton; S. Hecht; A. Bacher; W. Eisenreich; F. Rohdich; W. N. Hunter, *J. Biol. Chem.* **2004**, *279*, 52753–52761.
128. C. A. Testa; C. Lherbet; F. Pojer; J. P. Noel; C. D. Poulter, *Biochim. Biophys. Acta* **2005**, *1764*, 85–96.
129. C. Lherbet; F. Pojer; S. B. Richard; J. P. Noel; C. D. Poulter, *Biochemistry* **2006**, *45*, 3548–3553.
130. N. Campos; M. Rodríguez-Concepción; M. Seemann; M. Rohmer; A. Boronat, *FEBS Lett.* **2001**, *488*, 170–173.
131. B. Altincicek; A. K. Kollas; S. Sanderbrand; J. Wiesner; M. Hintz; E. Beck; H. Jomaa, *J. Bacteriol.* **2001**, *183*, 2411–2416.
132. S. Hecht; W. Eisenreich; P. Adam; S. Amslinger; K. Kis; A. Bacher; D. Arigoni; F. Rohdich, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14837–14842.
133. M. Seemann; N. Campos; M. Rodríguez-Concepción; J. F. Hoeffler; C. Grosdemange-Billiard; A. Boronat; M. Rohmer, *Tetrahedron Lett.* **2002**, *43*, 775–778.
134. M. Seemann; N. Campos; M. Rodríguez-Concepción; E. Ibañez; T. Duvold; A. Boronat; M. Rohmer, *Tetrahedron Lett.* **2002**, *43*, 1413–1415.
135. M. Wolff; M. Seemann; C. Grosdemange-Billiard; D. Tritsch; N. Campos; M. Rodríguez-Concepción; A. Boronat; M. Rohmer, *Tetrahedron Lett.* **2002**, *43*, 2555–2559.
136. M. Seemann; B. Tse Sum Bui; M. Wolff; D. Tritsch; N. Campos; A. Boronat; A. Marquet; M. Rohmer, *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 4337–4339.
137. F. Rohdich; F. Zepeck; P. Adam; S. Hecht; J. Kaiser; R. Laupitz; T. Gräwert; S. Amslinger; W. Eisenreich; A. Bacher; D. Arigoni, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1586–1591.
138. F. Zepeck; T. Gräwert; J. Kaiser; N. Schramek; W. Eisenreich; A. Bacher; F. Rohdich, *J. Org. Chem.* **2005**, *70*, 9168–9174.
139. K. J. Puan; H. Wang; T. Dairi; T. Kuzuyama; C. T. Morita, *FEBS Lett.* **2005**, *579*, 3802–3806.
140. A. K. Kollas; E. C. Duin; M. Eberl; B. Altincicek; M. Hintz; A. Reichenberg; D. Henschker; A. Henne; I. Steinbrecher; D. N. Ostrovsky; R. Hedderich; E. Beck; H. Jomaa; J. Wiesner, *FEBS Lett.* **2002**, *532*, 432–436.

141. W. Gao; R. Loeser; M. Raschke; M. A. Dessoy; M. Fulhorst; H. Alpermann; L. A. Wessjohann; M. H. Zenk, *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 2604–2607.
142. M. Seemann; P. Wegner; V. Schünemann; B. Tse Sum Bui; M. Wolff; A. Marquet; A. X. Trautwein; M. Rohmer, *J. Biol. Inorg. Chem.* **2005**, *10*, 131–137.
143. D. Adedej; H. Hernandez; J. Wiesner; U. Köhler; H. Jomaa; E. C. Duin, *FEBS Lett.* **2007**, *581*, 279–283.
144. W. Brandt; M. A. Dessoy; M. Fulhorst; W. Gao; M. H. Zenk; L. A. Wessjohann, *ChemBioChem* **2004**, *5*, 311–323.
145. J. Querol; N. Campos; S. Imperial; A. Boronat; M. Rodríguez-Concepción, *FEBS Lett.* **2002**, *514*, 343–346.
146. K. Okada; T. Hase, *J. Biol. Chem.* **2005**, *280*, 20672–20679.
147. M. Seemann; B. Tse Sum Bui; M. Wolff; M. Miginiac-Maslow; M. Rohmer, *FEBS Lett.* **2006**, *580*, 1547–1552.
148. C. Rivasseau; M. Seemann; A. M. Boisson; P. Streb; E. Gout; R. Douce; M. Rohmer; R. Bligny, *Plant Cell Environ.* **2009**, *32*, 82–92.
149. F. M. Hahn; A. P. Hurlburt; C. D. Poulter, *J. Bacteriol.* **1999**, *181*, 4499–4504.
150. M. Rodríguez-Concepción; N. Campos; L. M. Lois; C. Maldonaldo; J. F. Hoeffler; C. Grosdemange-Billiard; M. Rohmer; A. Boronat, *FEBS Lett.* **2000**, *473*, 328–332.
151. F. X. Cunningham; T. P. Lafond; E. Gantt, *J. Bacteriol.* **2000**, *182*, 5841–5848.
152. B. Altincicek; A. K. Kollas; M. Eberl; J. Wiesner; S. Sanderbrand; M. Hintz; E. Beck; H. Jomaa, *FEBS Lett.* **2001**, *499*, 37–40.
153. F. Rohdich; S. Hecht; K. Gärtner; P. Adam; C. Krieger; S. Amslinger; D. Arigoni; A. Bacher; W. Eisenreich, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1158–1163.
154. P. Adam; S. Hecht; W. Eisenreich; J. Kaiser; D. Arigoni; A. Bacher; F. Rohdich, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12108–12113.
155. B. Altincicek; E. C. Duin; A. Reichenberg; R. Hedderich; A. K. Kollas; M. Hintz; S. Wagner; J. Wiesner; E. Beck; H. Jomaa, *FEBS Lett.* **2002**, *532*, 437–440.
156. R. C. Röhrich; N. Englert; K. Troschke; A. Reichenberg; M. Hintz; F. Seeber; E. Balconi; A. Aliverti; G. Zanetti; U. Köhler; M. Pfeiffer; E. Beck; H. Jomaa; J. Wiesner, *FEBS Lett.* **2005**, *579*, 6433–6438.
157. M. Wolff; M. Seemann; B. Tse Sum Bui; Y. Frapart; D. Tritsch; A. G. Estrabot; M. Rodríguez-Concepción; A. Boronat; A. Marquet; M. Rohmer, *FEBS Lett.* **2003**, *541*, 115–120.
158. T. Gräwert; J. Kaiser; F. Zepeck; R. Laupitz; S. Hecht; S. Amslinger; N. Schramek; E. Schleicher; S. Weber; M. Haslbeck; J. Buchner; C. Rieder; D. Arigoni; A. Bacher; W. Eisenreich; F. Rohdich, *J. Am. Chem. Soc.* **2004**, *126*, 12847–12855.
159. I. Rekittke; J. Wiesner; R. Röhrich; U. Demmer; E. Warkentin; W. Xu; K. Troschke; M. Hintz; J. H. No; E. C. Duin; E. Oldfield; H. Jomaa; U. Ermler, *J. Am. Chem. Soc.* **2008**, *130*, 17206–17207.
160. R. Laupitz; T. Gräwert; C. Rieder; F. Zepeck; A. Bacher; D. Arigoni; F. Rohdich; W. Eisenreich, *Chem. Biodivers.* **2004**, *1*, 1367–1376.
161. L. Charon; C. Pale-Grosdemange; M. Rohmer, *Tetrahedron Lett.* **1999**, *40*, 7231–7234.
162. M. Rohmer, *Pure Appl. Chem.* **2007**, *78*, 739–751.
163. Y. Xiao; Z. K. Zhao; P. Liu, *J. Am. Chem. Soc.* **2008**, *130*, 2164–2165.
164. Y. Xiao; P. Liu, *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 9722–9725.
165. S. Rosa-Putra; A. Disch; J. M. Bravo; M. Rohmer, *FEMS Microbiol. Lett.* **1998**, *164*, 169–175.
166. Y. Boucher; W. F. Doolittle, *Mol. Microbiol.* **2000**, *37*, 703–716.
167. B. M. Lange; T. Rujan; W. Martin; R. Croteau, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13172–13177.
168. S. M. Trutko; L. V. Dorofeeva; L. I. Evtushenko; D. N. Ostrovskii; M. Hintz; J. Wiesner; H. Jomaa; B. P. Baskunov; V. K. Akimenko, *Microbiology* **2005**, *74*, 335–341.
169. A. Smit; A. Mushegian, *Genome Res.* **2000**, *10*, 1468–1484.
170. H. Seto; H. Watanabe; K. Furihata, *Tetrahedron Lett.* **1996**, *37*, 7979–7982.
171. H. Seto; N. Orihara; K. Furihata, *Tetrahedron Lett.* **1998**, *39*, 9497–9500.
172. Y. Hamano; T. Dairi; M. Yamamoto; T. Kuzuyama; N. Itoh; H. Seto, *Biosci. Biotechnol. Biochem.* **2002**, *66*, 808–819.
173. M. Begley; C. G. M. Gahan; A. K. Kollas; M. Hintz; C. Hill; H. Jomaa; M. Eberl, *FEBS Lett.* **2004**, *561*, 99–104.
174. J. G. Zeidler; H. K. Lichtenthaler; H. U. May; F. W. Lichtenthaler, *Z. Naturforsch.* **1997**, *52c*, 15–23.
175. V. Stanjek; J. Piel; W. Boland, *Phytochemistry* **1999**, *50*, 1141–1145.
176. M. Goese; K. Kammhuber; A. Bacher; M. H. Zenk; W. Eisenreich, *Eur. J. Biochem.* **1999**, *263*, 447–454.
177. S. Berwanger; J. Zapp; H. Becker, *Planta Med.* **2005**, *71*, 530–534.
178. Y. Asada; W. Li; T. Yoshikawa, *Phytochemistry* **2000**, *55*, 323–326.
179. Y. S. Han; R. van der Hejden; A. W. M. Lefeber; C. Erkelen; R. Verpoorte, *Phytochemistry* **2002**, *59*, 45–55.
180. P. Adam; D; Arigoni; A. Bacher; W. Eisenreich, *J. Med. Chem.* **2002**, *45*, 4786–4793.
181. L. Margl; C. Ettenhuber; I. Gyurjan; M. H. Zenk; A. Bacher; W. Eisenreich, *Phytochemistry* **2005**, *66*, 887–899.
182. J. Zeidler; H. K. Lichtenthaler, *Planta* **2001**, *213*, 323–326.
183. A. J. Barlow; H. Becker; K. P. Adam, *Phytochemistry* **2001**, *57*, 7–14.
184. H. Yamamoto; P. Zhao; K. Inoue, *Phytochemistry* **2002**, *60*, 263–267.
185. W. Eisenreich; S. Sagner; M. H. Zenk; A. Bacher, *Tetrahedron Lett.* **1997**, *38*, 3889–3892.
186. R. Thiel; K. P. Adam; J. Zapp; H. Becker, *Pharm. Pharmacol. Lett.* **1997**, *7*, 103–105.
187. D. J. Fowler; J. T. Hamilton; A. J. Humphrey; D. O'Hagan, *Tetrahedron Lett.* **1999**, *40*, 3803–3806.
188. M. Fellermeier; W. Eisenreich; A. Bacher; M. H. Zenk, *Eur. J. Biochem.* **2001b**, *268*, 1596–1604.
189. A. Contin; R. van der Heijden; A. W. M. Lefeber; R. Verpoorte, *FEBS Lett.* **1998**, *434*, 413–416.
190. D. Eichinger; A. Bacher; M. H. Zenk; W. Eisenreich, *Phytochemistry* **1999**, *51*, 223–236.
191. Y. Yamasaki; M. Kitajima; M. Arita; H. Takayama; H. Sudo; M. Yamazaki; N. Aimi; K. Saito, *Plant Physiol.* **2004**, *134*, 161–170.
192. D. Umlauf; J. Zapp; H. Becker; K. P. Adam, *Phytochemistry* **2004**, *65*, 2463–2470.
193. K. Matsuda; Y. Kikuta; A. Haba; K. Nakayama; Y. Katsuda; A. Hatanaka; K. Komai, *Phytochemistry* **2005**, *66*, 1529–1535.
194. D. Hampel; A. Mosandl; M. Wüst, *Phytochemistry* **2005**, *66*, 305–311.
195. W. Eisenreich; B. Menhard; P. J. Hylands; M. H. Zenk; A. Bacher, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6431–6436.
196. W. Knöss; B. Reuter; J. Zapp, *Biochem. J.* **1996**, *326*, 449–454.

197. T. Hayashi; T. Asai; U. Sankawa, *Tetrahedron Lett.* **1999**, *40*, 8239–8243.
198. W. Eisenreich; C. Rieder; C. Grammes; G. Hessler; K. P. Adam; H. Becker; D. Arigoni; A. Bacher, *J. Biol. Chem.* **1999**, *274*, 36312–36320.
199. U. Hertewich; J. Zapp; H. Becker; K. P. Adam, *Phytochemistry* **2001**, *58*, 1049–1054.
200. N. Totté; L. Charon; M. Rohmer; F. Compernolle; I. Baboeuf; J. M. C. Geuns, *Tetrahedron Lett.* **2000**, *41*, 6407–6410.
201. N. Hirai; R. Yoshida; Y. Todoroki; H. Ohigashi, *Biosci. Biotechnol. Biochem.* **2000**, *64*, 1448–1458.
202. L. Kutrzeba; F. E. Dayan; J. L. Howell; F. Feng; J. L. Giner; J. K. Zjawiony, *Phytochemistry* **2007**, *68*, 1872–1881.
203. E. Fukusaki; S. Takeno; T. Bamba; H. Okumoto; H. Katto; S. Kajiyama; A. Kobayashi, *Biosci. Biotechnol. Biochem.* **2004**, *68*, 1988–1990.
204. A. Disch; A. Hemmerlin; T. J. Bach; M. Rohmer, *Biochem. J.* **1998**, *331*, 615–621.
205. S. M. Li; S. Hennig; L. Heide, *Tetrahedron Lett.* **1998**, *39*, 2721–2724.
206. P. Steliopoulos; M. Wüst; K. P. Adam; A. Mosandl, *Phytochemistry* **2002**, *60*, 12–20.
207. A. J. Barlow; S. D. Lorimer; E. R. Morgan; R. T. Weavers, *Phytochemistry* **2003**, *63*, 25–29.
208. J. van Klink; H. Becker; S. Andersson; W. Boland, *Org. Biomol. Chem.* **2003**, *1*, 1503–1508.
209. A. Disch; J. Schwender; C. Müller; H. K. Lichtenthaler; M. Rohmer, *Biochem. J.* **1998**, *333*, 381–388.
210. M. H. Hsieh; C. Y. Chang; S. J. Hsu; J. J. Chen, *Plant Mol. Biol.* **2008**, *66*, 663–673.
211. A. Jux; G. Gleixner; W. Boland, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 2091–2093.
212. Y. Chikaraishi; H. Naraoka; S. R. Poulson, *Phytochemistry* **2004**, *65*, 323–330.
213. J. Schwender; M. Seemann; H. K. Lichtenthaler; M. Rohmer, *Biochem. J.* **1996**, *316*, 73–80.
214. J. Schwender; C. Gemünden; H. K. Lichtenthaler, *Planta* **2001**, *212*, 416–423.
215. Y. Sato; Y. Ito; S. Okada; M. Murakami; H. Abe, *Tetrahedron Lett.* **2003**, *44*, 7035–7037.
216. C. Grauvogel; J. Petersen, *Gene* **2007**, *396*, 125–133.
217. J. H. Cvejic; M. Rohmer, *Phytochemistry* **2000**, *53*, 21–28.
218. D. Kim; M. R. Fritz; P. J. Proteau, *J. Nat. Prod.* **2004**, *67*, 1067–1069.
219. H. Jomaa; J. Wiesner; S. Sanderbrand; B. Altincicek; C. Weidemeyer; M. Hintz; I. Türbachova; M. Eberl; J. Zeidler; H. K. Lichtenthaler; D. Soldati; E. Beck, *Science* **1999**, *285*, 1573–1576.
220. J. Wiesner; D. Henschker; D. B. Hutchinson; E. Beck; H. Jomaa, *Antimicrob. Agents Chemother.* **2002**, *46*, 2889–2894.
221. J. Wiesner; H. Jomaa, *Curr. Drug Targets* **2007**, *8*, 3–13.
222. M. B. Cassera; F. C. Gozzo; F. L. D'Alexandri; E. F. Merino; H. A. del Portillo; V. J. Peres; I. C. Almeida; M. N. Eberlin; G. Wunderlich; J. Wiesner; H. Jomaa; E. A. Kimura; A. M. Katzin, *J. Biol. Chem.* **2004**, *279*, 51749–51759.
223. M. Clastre; A. Goubard; A. Prel; Z. Mincheva; M. C. Viaud-Massuart; D. Bout; M. Rideau; F. Velge-Roussel; F. Laurent, *Exp. Parasitol.* **2007**, *116*, 375–384.
224. S. N. J. Moreno; Z. H. Li, *Expert Opin. Ther. Targets* **2008**, *12*, 253–263.
225. W. Zhou; W. D. Nes, *Tetrahedron Lett.* **2000**, *41*, 2791–2795.
226. M. Matsuzaki; H. Kuroiwa; T. Kuroiwa; K. Kita; H. Nosaki, *Mol. Biol. Evol.* **2008**, *25*, 1167–1179.
227. C. A. Testa; M. J. Brown, *Curr. Pharm. Biotechnol.* **2003**, *4*, 248–259.
228. F. Rohdich; A. Bacher; W. Eisenreich, *Bioorg. Chem.* **2004**, *32*, 292–308.
229. H. Eoh; P. J. Brennan; D. C. Crick, *Tuberculosis* **2009**, *89*, 1–11.
230. M. A. Missinou; S. Borrmann; A. Schindler; S. Issifou; A. A. Adegnika; P. B. Matsiegui; R. Binder; B. Lell; J. Wiesner; T. Baranek; H. Jomaa; P. G. Kremsner, *Lancet* **2002**, *360*, 1941–1942.
231. B. Lell; R. Ruangweerayut; J. Wiesner; M. A. Missinou; A. Schindler; T. Baranek; M. Hintz; D. Hutchinson; H. Jomaa; P. G. Kremsner, *Antimicrob. Agents Chemother.* **2003**, *47*, 735–738.
232. S. Oyakhirome; S. Issifou; P. Pongratz; F. Barondi; M. Ramharter; J. Kun; M. Missinou; B. Lell; P. Kremsner, *Antimicrob. Agents Chemother.* **2007**, *51*, 1869–1871.
233. E. B. Gottlin; R. E. Benson; S. C. Connary; B. Antonio; K. Duke; E. S. Payne; S. S. Ashraf; D. J. Christensen, *J. Biomol. Screen.* **2003**, *8*, 332–339.
234. V. Illarionova; J. Kaiser; E. Ostrozhenkova; A. Bacher; M. Fischer; W. Eisenreich; F. Rohdich, *J. Org. Chem.* **2006**, *71*, 8824–8834.
235. H. P. Kuemmerle; T. Murakawa; H. Sakamoto; N. Sato; T. Konishi; F. De Santis, *Int. J. Clin. Pharmacol. Ther. Toxicol.* **1985**, *23*, 521–528.
236. S. Fujisaki; S. I. Ohnuma; T. Horiuchi; I. Takahashi; S. Tsukui; Y. Nishimura; T. Nishino; M. Kitabatake; H. Inokuchi, *Gene* **1996**, *175*, 83–87.
237. Y. Sakamoto; S. Furukawa; H. Ogihara; M. Yamasaki, *Biosci. Biotechnol. Biochem.* **2003**, *67*, 2030–2033.
238. A. C. Brown; T. Parish, *BMC Microbiol.* **2008**, *8*, 78–86.
239. J. Kaiser; M. Yassin; S. Prakash; N. Safi; M. Agami; S. Lauw; E. Ostroshenkova; A. Bacher; F. Rohdich; W. Eisenreich; J. Safi; A. Golan-Goldirsch, *Phytomedicine* **2007**, *14*, 242–249.
240. M. Albrecht; N. Misawa; G. Sandmann, *Biotechnol. Lett.* **1999**, *21*, 791–795.
241. M. Harker; P. M. Bramley, *FEBS Lett.* **1999**, *448*, 115–119.
242. C. W. Wang; M. K. Oh; J. C. Liao, *Biotechnol. Prog.* **2000**, *16*, 922–926.
243. S. J. Kim; M. D. Kim; J. H. Choi; S. Y. Kim; Y. W. Ryu; J. H. Seo, *Appl. Microbiol. Biotechnol.* **2006**, *72*, 982–985.
244. L. Carretero-Paulet; A. Cairó; P. Botella-Pavía; O. Besumbes; N. Campos; A. Boronat; M. Rodríguez-Concepción, *Plant Mol. Biol.* **2006**, *62*, 683–695.
245. E. M. A. Enfissi; P. D. Fraser; L. M. Lois; A. Boronat; W. Schuch; P. M. Bramley, *Plant Biotechnol. J.* **2005**, *3*, 17–27.
246. J. Muñoz-Bertomeu; I. Arrillaga; R. Ros; J. Segura, *Plant Physiol.* **2006**, *142*, 890–900.
247. P. Botella-Pavía; O. Besumbes; M. A. Phillips; L. Carretero-Paulet; A. Boronat; M. Rodríguez-Concepción, *Plant J.* **2004**, *40*, 188–199.
248. K. Nabeta; T. Ishikawa; H. Okuyama, *J. Chem. Soc. Perkin Trans. I* **1995**, 3111–3115.
249. K. Nabeta; T. Kawae; T. Saitoh; T. Kikuchi, *J. Chem. Soc. Perkin Trans. I* **1997**, 261–267.
250. R. P. Karunagoda; D. Itoh; K. Katoh; K. Nabeta, *Biosci. Biotechnol. Biochem.* **2001**, *65*, 1076–1081.

251. D. Itoh; R. P. Karunagoda; T. Fishie; K. Katoh; K. Nabeta, *J. Nat. Prod.* **2000**, *63*, 1090–1093.
252. J. W. Yang; Y. Orihara, *Tetrahedron* **2002**, *58*, 1265–1270.
253. D. Arigoni; S. Sagner; C. Latzel; W. Eisenreich; A. Bacher; M. H. Zenk, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10600–10605.
254. C. A. Schuhr; T. Radykewicz; S. Sagner; C. Latzel; M. H. Zenk; D. Arigoni; A. Bacher; F. Rohdich; W. Eisenreich, *Phytochem. Rev.* **2003**, *2*, 3–16.
255. W. De-Eknamkul; B. Potduang, *Phytochemistry* **2003**, *62*, 389–398.
256. K. P. Adam; J. Zapp, *Phytochemistry* **1998**, *49*, 953–959.
257. D. Hampel; A. Mosandl; M. Wüst, *J. Agric. Food Chem.* **2005**, *53*, 2652–2657.
258. D. C. B. Bergamo; M. J. Kato; V. da Silva Bolzani; M. Furlan, *J. Braz. Chem. Soc.* **2005**, *16*, 1406–1409.
259. F. Luang; M. Wüst, *Phytochemistry* **2002**, *60*, 451–459.
260. D. Hampel; A. Mosandl; M. Wüst, *J. Agric. Food. Chem.* **2006**, *54*, 1473–1478.
261. A. A. Lopes; D. C. Baldoqui; S. N. López; M. J. Kato; V. da; S. Bolzani; M. Furlan, *Phytochemistry* **2007**, *68*, 2053–2058.
262. K. Skorupinska-Tudek; J. Poznanski; J. Wojcik; T. Bienkowski; I. Szostkiewicz; M. Zelma-Femiak; A. Bajda; T. Chojnacki; O. Olszowska; J. Grunler; O. Meyer; M. Rohmer; W. Danikiewicz; E. Swiezewska, *J. Biol. Chem.* **2008**, *283*, 21024–21035.
263. A. Hemmerlin; J. F. Hoeffler; O. Meyer; D. Tritsch; I. A. Kagan; C. Grosdemange-Billiard; M. Rohmer; T. J. Bach, *J. Biol. Chem.* **2003**, *278*, 26666–26676.
264. O. Laule; A. Fürholz; H. S. Chang; T. Zhu; X. Wang; P. B. Heifetz; W. Gruissem; B. M. Lange, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6866–6871.
265. J. A. Bick; B. M. Lange, *Arch. Biochem. Biophys.* **2003**, *415*, 146–154.
266. J. L. Giner; B. Jaun; D. Arigoni, *Chem. Commun.* **1998**, 1857–1858.
267. L. Charon; J. F. Hoeffler; C. Pale-Grosdemange; L. M. Lois; N. Campos; A. Boronat; M. Rohmer, *Biochem. J.* **2000**, *346*, 737–742.
268. C. Rieder; B. Jaun; D. Arigoni, *Helv. Chim. Acta* **2000**, *83*, 2504–2513.
269. D. Arigoni; W. Eisenreich; C. Latzel; S. Sagner; T. Radykewicz; M. H. Zenk; A. Bacher, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1309–1314.
270. J. F. Hoeffler; A. Hemmerlin; C. Grosdemange-Billiard; T. J. Bach; M. Rohmer, *Biochem. J.* **2002**, *366*, 573–582.
271. T. N. Rosenstiel; A. J. Fischer; R. Fall; R. K. Monson, *Plant Physiol.* **2002**, *129*, 1276–1284.
272. J. L. Funk; C. G. Jones; C. J. Baker; H. M. Fuller; C. P. Giardina; M. T. Lerdau, *Ecol. Appl.* **2003**, *13*, 269–278.
273. M. Loivamäki; S. Louis; G. Cinege; I. Zimmer; R. J. Fischbach; J. P. Schnitzler, *Plant Physiol.* **2007**, *143*, 540–551.
274. N. Dudareva; S. Andersson; I. Orlova; N. Gatto; M. Reichelt; D. Rhodes; W. Boland; J. Gershenzon, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 933–938.
275. A. Okada; T. Shimizu; K. Okada; T. Kuzuyama; J. Koga; N. Shibuya; H. Nojiri; H. Yamane, *Plant Mol. Biol.* **2007**, *65*, 177–187.
276. G. Arimura; D. P. W. Huber; J. Bohlmann, *Plant J.* **2004**, *37*, 603–616.
277. G. Arimura; S. Köpke; M. Kunert; V. Volpe; A. David; P. Brand; P. Dabrowska; M. E. Maffei; W. Boland, *Plant Physiol.* **2008**, *146*, 965–973.
278. J. Piel; J. Donath; K. Bandemer; W. Boland, *Angew. Chem. Int. Ed. Engl.* **1998**, *37*, 2478–2481.

Biographical Sketch

Michel Rohmer, Professor of Chemistry, Université de Strasbourg/CNRS, Institut de Chimie, Strasbourg, France. As a chemical engineer of the Ecole Nationale Supérieure de Chimie de Strasbourg (1970), Michel Rohmer completed his Ph.D. thesis under the supervision of Professsor Guy Ourisson at the Université Louis Pasteur of Strasbourg (1975), working on the chemistry and biochemistry of prokaryotic triterpenoids. Meanwhile, he was 'assistant' and later 'maître-assistant' in Pharmacognosy at the Faculty of Pharmacy of the Université Louis Pasteur (1974–79). After postdoctoral work with Professor Carl Djerassi at Stanford University (1978–79) on sterols from marine organisms, he was promoted as Professor of organic and bio-organic chemistry first at the Université de Haute Alsace

(Ecole Nationale Supérieure de Chimie) in Mulhouse (1979–94) and later in 1994 at the Faculty of Chemistry of the University of Strasbourg.

He is a member of the 'Institut Universitaire de France' (1997), the 'Deutsche Akademie der Naturforscher Leopoldina' (2000) and the French 'Académie des Sciences' (2003).

Michel Rohmer is a specialist in the chemistry and biochemistry of isoprenoids from microorganisms and higher plants. The main discoveries include the biohopanoids, a series of pentacyclic bacterial triterpenoids, precursors of a ubiquitous family of molecular fossils, and the MEP pathway, a novel metabolic route toward the isoprene units in bacteria and plant plastids.

1.14 Prenyltransferase

Hirofumi Kurokawa and Tanetoshi Koyama, Tohoku University, Sendai, Japan

© 2010 Elsevier Ltd. All rights reserved.

1.14.1	**Introduction**	557
1.14.1.1	Mevalonate Pathway and Nonmevalonate Pathway	558
1.14.1.2	Biosynthetic Pathway of Isoprenoids Includes Various Prenyltransferases	559
1.14.2	**(*E*)-Prenyltransferases**	561
1.14.2.1	Classification of (*E*)-Prenyltransferases	561
1.14.2.1.1	Short-chain (*E*)-prenyltransferases	561
1.14.2.1.2	Medium-chain (*E*)-prenyltransferases	562
1.14.2.1.3	Long-chain (*E*)-prenyltransferases	563
1.14.2.2	Stereochemistry of (*E*)-Prenyltransferases	563
1.14.2.3	Three-Dimensional Structure of (*E*)-Prenyltransferases	564
1.14.2.3.1	FPP synthase	564
1.14.2.3.2	Geranylgeranyl diphosphate synthase	565
1.14.2.4	Catalytic Mechanism	565
1.14.2.4.1	Roles of conserved amino acid residues	565
1.14.2.4.2	Random chemical mutagenesis	568
1.14.2.5	Product Chain-Length Regulation Mechanism	568
1.14.3	**(*Z*)-Prenyl Diphosphate Synthases**	569
1.14.3.1	Classification of (*Z*)-Prenyltransferases	571
1.14.3.1.1	Short- and long-chain (*Z*)-prenyltransferases	571
1.14.3.1.2	Long-chain (*Z*)-prenyltransferases	572
1.14.3.2	Stereochemistry of (*Z*)-Prenyltransferase	572
1.14.3.3	Three-Dimensional Structure of (*Z*)-Prenyltransferase	572
1.14.3.4	Catalytic Mechanism	574
1.14.3.5	Product Chain-Length Regulation Mechanism	577
1.14.4	**Conclusions**	579
References		580

1.14.1 Introduction

Isoprenoids constitute the most diverse group of natural products. Prenyltransferases catalyze the fundamental isoprenoid chain-elongation process to produce prenyl diphosphate with various chain lengths and stereochemistries. All of the compounds are derived from linear isoprenoid diphosphates synthesized from the isomeric 5-carbon unit intermediates isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP).[1–3] The reactions are regulated to proceed continuously and terminate precisely at definite chain lengths according to the specificities of individual prenyltransferases. Isoprenoid compounds play crucial metabolic and structural roles in cells. Sterols, known as structural components of biological membranes controlling membrane fluidity, play important roles as precursors of steroid hormones in animals and plants. Protein prenylation, a posttranslational modification mediated by covalent attachment of a farnesyl or geranylgeranyl moiety onto specific proteins, such as Ras, trimeric G-proteins, and yeast-mating pheromone α-factor, is required for membrane association of these regulatory proteins that function in signal transduction cascade of higher organisms. Prenyl quinines are used as electron carriers required for the mitochondrial respiratory chain and as cofactors in electron transport chains in plastids of higher plants. Three phytohormones, abscisic acid, gibberellins, and cytokinins, are also included in isoprenoids. Dolichol plays an important role as a glycosyl carrier lipid in the biosynthesis of N-linked glycoproteins, O-linked mannose to tryptophan,

and glycosylphosphatidylinositol (GPI)-anchored protein. Other classes of isoprenoids are carotenoids (visual pigments and photoprotective agents), monoterpenes (insect sex pheromones and fragrances), and sesquiterpenes (plant defensive agents).

1.14.1.1 Mevalonate Pathway and Nonmevalonate Pathway

The biosynthetic pathways of isoprenoids can be divided into three phases: (1) formation of IPP, (2) condensation of IPP to synthesize linear isoprenoids, and (3) condensation, cyclization, and modification of linear isoprenoids (**Figure 1**). The first phase of the isoprenoid biosynthetic pathway is the formation of the C_5 building unit of isoprenoid IPP, which includes the formation of DMAPP or isomerization of IPP into DMAPP. The acetate/mevalonate pathway, in which three molecules of acetyl-CoA condense successively to form 3-hydroxyl-3-methylglutaryl-CoA (HMG-CoA) and followed to a key intermediate molecule, mevalonate, had commonly been considered as the sole metabolic pathway for the formation of IPP and DMAPP in all organisms.

The biosynthesis of IPP has long been assumed to proceed exclusively through the acetate/mevalonate pathway.[4] However, several lines of evidence indicated that the metabolic pathway of isoprenoids in certain eubacteria, plants, and green algae could not be solely explained by the mevalonate pathway.[5] Rhomer *et al.*[6,7] have discovered a novel pathway to IPP in which mevalonate is not involved. On the basis of labeling patterns of bacterial hopanoids and ubiquinones derived from ^{13}C-labeled metabolites of glycolysis as well as [^{13}C]acetate, they have proved that glyceraldehyde 3-phosphate occupies the branch point of glycolysis and nonmevalonate pathway to IPP. They have proposed another metabolic pathway of IPP from eubacteria initiates with a condensation of glyceraldehyde 3-phosphate and a C_2 unit resulted from thiamine diphosphate-dependent decarboxylation of pyruvate to form 1-deoxy-D-xylulose 5-phosphate (DXP).[6,7] Because DXP is known to be used in other metabolic pathways, such as the biosynthesis of pyridoxal phosphate and thiamine diphosphate,[8–10] methyl-D-erythritol 4-phosphate (MEP), which is converted from DXP by DXP

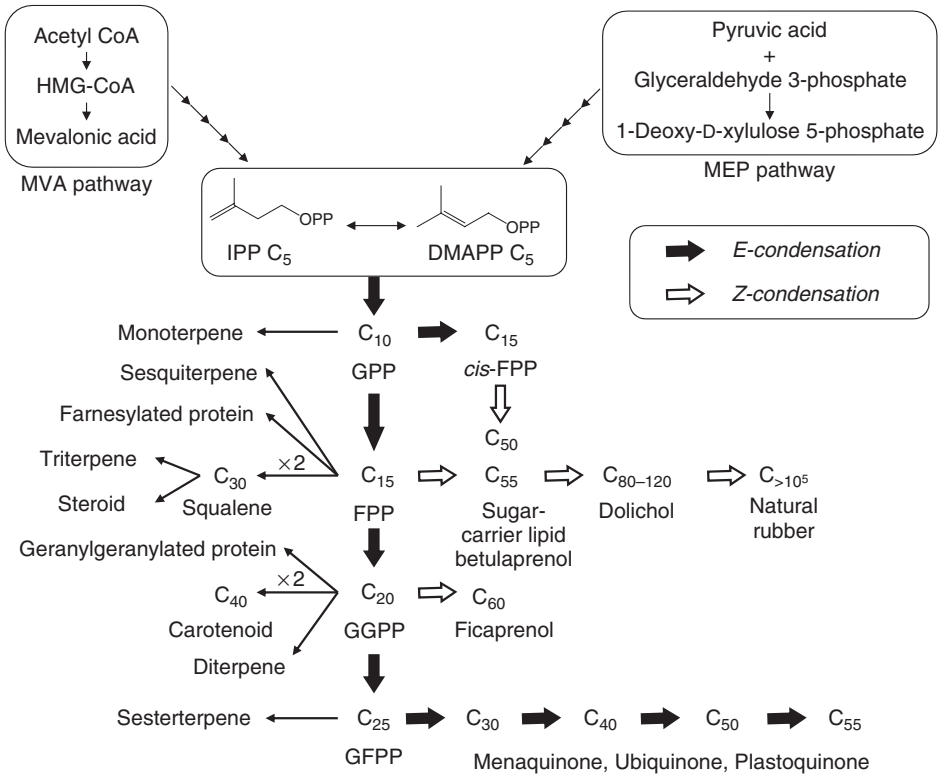


Figure 1 Biosynthetic pathways of isoprenoids.

reductoisomerase, is considered to be the first committed precursor of the mevalonate-independent pathway.[11] Therefore, this pathway is commonly termed as the nonmevalonate pathway or the MEP pathway. In the past decade, enzymes catalyzing in the MEP pathway were identified from various organisms, and an overall metabolic pathway was elucidated.[12,13] It has also been shown that the nonmevalonate pathway is operative in the biosynthesis of not only particular isoprenoids in bacteria but also of some terpenoids in plants[14] and algae.[15]

1.14.1.2 Biosynthetic Pathway of Isoprenoids Includes Various Prenyltransferases

In the biosynthesis of isoprenoids, all carbon skeletons are synthesized from linear isoprenoids at the second phase in isoprenoid biosynthetic pathways (**Figure 1**). The carbon chain length of naturally occurring linear isoprenoids is widely distributed from geranyl diphosphate (GPP, C_{10}) to natural rubber ($C_{>100\,000}$). Every linear isoprenoid is formed by sequential condensation of IPP and allylic diphosphates with actions of prenyl chain-elongating enzymes, commonly called prenyltransferases. In a broad sense, prenyltransferases include all enzymes that catalyze the transfer of allylic prenyl groups to acceptor molecules, such as IPP, aromatic intermediates of quinones, or specific proteins. In this review, for convenience, prenyltransferase refers to prenyl diphosphate synthase that catalyzes the sequential condensation of IPP to allylic diphosphates. The reactions catalyzed by prenyltransferases start by the formation of allylic cations after the elimination of pyrophosphate ions to form allylic prenyl diphosphate, followed by the addition of an IPP with stereospecific removal of a proton at the 2-position.

Comparison of many kinds of prenyltransferases revealed that prenyltransferases can be classified into two major groups according to the protein structure and the stereochemistry; *trans*- or (*E*)-prenyltransferases and *cis*- or (*Z*)-prenyltransferases (**Figure 2**).[13] The biosynthesis of linear isoprenoids can be divided into two phases (**Figures 1** and **3**). In the first phase, short-chain allylic diphosphates, GPP (C_{10}), farnesyl diphosphate (FPP, C_{15}), and geranylgeranyl diphosphate (GGPP, C_{20}) are formed by the action of *trans*- or (*E*)-prenyltransferases. These (all-*E*)-short-chain prenyl diphosphates are then used as allylic primer substrates for additional IPP condensation with (*E*)- or (*Z*)-configuration by (*E*)- or (*Z*)-prenyltransferases, respectively. The prenyltransferases responsible for each linear isoprenoid with a specific number of isoprene units strictly recognize the prenyl chain lengths of the allylic substrates and ultimate products. From the viewpoints of enzymology and organic chemistry, this property of prenyltransferases is one of the most interesting research topics regarding the catalytic mechanisms of prenyl chain-elongating enzymes. The absolute stereochemistry of reactions catalyzed by each type of prenyltransferase with respect to the face of the double bond of IPP

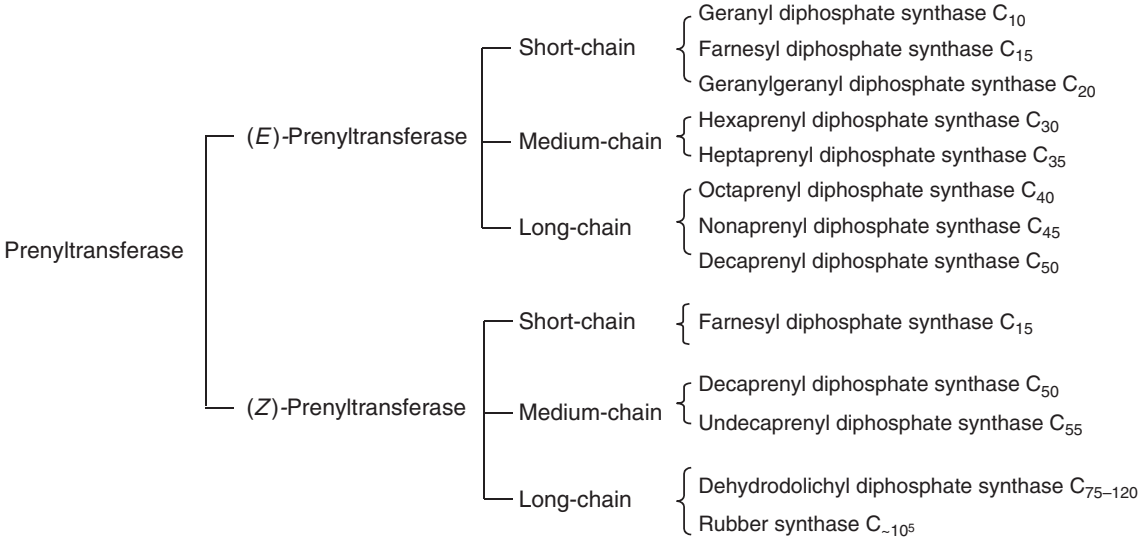


Figure 2 Classification of the enzymes that catalyze prenyl chain elongation.

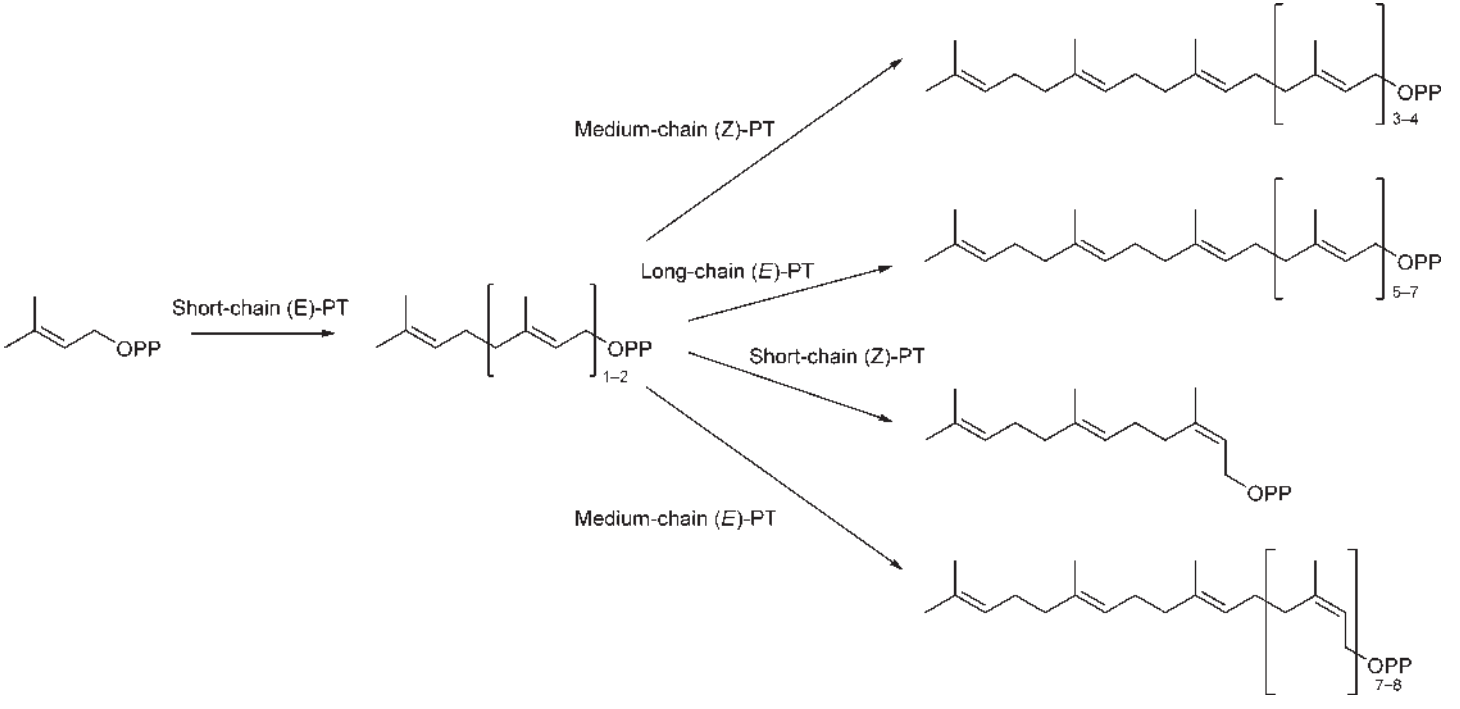


Figure 3 Biosynthesis of linear isoprenoids catalyzed by prenyltransferases (PT).

during the carbon–carbon bond formation was revealed using (all-*E*)-heptaprenyl diphosphate (HepPP) synthase ((*E*)-prenyltransferase) and UPP synthase ((*Z*)-prenyltransferase) from *Bacillus subtilis*.[16,17] The only difference between the reaction catalyzed by *trans*- and *cis*-prenyltransferases is the prochirality of the proton that is eliminated from the 2-position of IPP, that is, pro-*R* for (*E*)-prenyltransferase and pro-*S* for (*Z*)-prenyltransferase, suggesting a similar structural property in the catalytic centers of (*E*)- and (*Z*)-prenyltransferases. However, recent molecular analyses and crystal structure determinations of *Micrococcus luteus* B-P 26 UPP (C_{55}) synthase,[18] which catalyzes sequential *Z*-condensation of IPP, showed that not only the primary but also the three-dimensional (3D) structure of (*Z*)-prenyltransferases was totally different from that of (*E*)-prenyltransferases.[18–20]

1.14.2 (*E*)-Prenyltransferases

(*E*)-Prenyltransferases catalyze the synthesis of all-*trans*-prenyl diphosphates with various chain lengths from C_{10} to C_{50} (**Figure 2**). Bacterial enzymes that catalyze the synthesis of short-chain-length prenyl diphosphates such as C_{15} and C_{20}, and long-chain prenyl diphosphates with chain lengths from C_{40} to C_{50}, show a homodimeric structure (**Table 1**). Although some enzymes, for example, GPP synthase, show high-order oligomeric states, a homodimer appears to be a minimum functional unit for catalysis. The enzymes synthesize the prenyl diphosphate with medium chain lengths of C_{30} or C_{35} showing unique heterodimeric structure (**Table 1**). Since each component alone does not show any enzymatic activity, the heterodimer is essential for their enzymatic activity. FFP (C_{15}) synthase (*trans*-FPS) plays an important role for isoprenoid biosynthesis. The human *trans*-FPS is the molecular target of nitrogen-containing bisphosphonates, currently used as clinical inhibitors of bone resorption diseases, for example, osteoporosis.

In this section we will mainly consider the molecular mechanisms of the catalytic functions of (*E*)-prenyltransferases based on the recently accumulating reports accomplished by mutational and crystallographic structural studies.

1.14.2.1 Classification of (*E*)-Prenyltransferases

1.14.2.1.1 Short-chain (E)-prenyltransferases

Short-chain prenyl diphosphate synthases include FPP synthases (geranyltranstransferase (EC 2.5.1.10)) and GGPP synthases (farnesyltranstransferase (EC 2.5.1.30)). The enzymes require no cofactor except divalent metal ions such as Mg^{2+} or Mn^{2+}, which are essential cofactors for all prenyltransferases.

Table 1 Properties of typical *E*-prenyltransferases

Species	*Enzyme involved*	*Subunit structure*	*Elongation range*	*Reference(s)*
Lithospermum erythrorhizon	GPP synthase	n.d.	$C_5 \rightarrow C_{10}$	21
Saccharomyces cerevisiae	FPP synthase	Homodimer	$C_5 \rightarrow C_{15}$	22, 23
Rat liver	FPP synthase	Homodimer	$C_5 \rightarrow C_{15}$	24
Bacillus stearothermophilus	FPP synthase	Homodimer	$C_5 \rightarrow C_{15}$	25, 26
Methanobacterium thermoautorophicum	GGPP synthase	Homodimer	$C_5 \rightarrow C_{15}$,C_{20}	27
Bovine brain	GGPP synthase	Homooligomer	$C_{15} \rightarrow C_{20}$	28
Sulfolobus acidocaldarius	GGPP synthase	Homodimer	$C_5 \rightarrow C_{20}$	29
Natronobacterium pharaonis	FGPP synthase	n.d.	$C_{20} \rightarrow C_{30}$	30
Micrococcus luteus B-P 26	HexPP synthase	Heterodimer	$C_{15} \rightarrow C_{30}$	31, 32, 13
B. stearothermophilus	HepPP synthase	Heterodimer	$C_{15} \rightarrow C_{35}$	34, 35
Bacillus subtilis	HepPP synthase	Heterodimer	$C_{15} \rightarrow C_{35}$	36, 37
Escherichia coli	OctPP synthase	Homodimer	$C_{15} \rightarrow C_{40}$	38
M. luteus	SPP synthase	Homodimer	$C_{15} \rightarrow C_{45}$	39
Paracoccus denitrificans	DecPP synthase	Homodimer	$C_{15} \rightarrow C_{50}$	40, 41

n.d., not determined.

Organisms constitutively contain at least one of these short-chain prenyl diphosphate synthases for the production of prenyl diphosphates that act as the priming substrates for the other groups of prenyltransferases. The short-chain prenyl diphosphates are also biosynthetic precursors of various isoprenoids including steroids, carotenoids, and prenylated proteins. It is worth noting that GGPP synthase occurs not only in plants and bacteria but also in mammals. The GGPP synthase from fission yeast, which is essential for sporulation, is a heterodimer of FPP synthase, Fps1, and an FPS-like protein, Spo9.[42]

The prenyltransferases in this group have been studied most extensively. Each enzyme of this group has a homodimeric structure in which the subunits are tightly bound to each other.

GPP synthase (dimethylallyltranstransferase (EC 2.5.1.1)), which catalyzes the condensation of DMAPP and IPP to GPP, the key precursor of monoterpene biosynthesis, exists in higher plants producing monoterpenes. This synthase has been purified from *Lithospermum erythrorhizon* cell culture,[21] *Salvia officinalis* leaves,[43] *Arabidopsis thaliana*,[44] *Abies grandis*,[45] and several higher dicotyledon plant leaves.[46] GPP synthase purified from the isolated oil glands of spearmint appeared to be a heterodimer consisting of 28 and 37 kDa FPP synthase-like subunits.[47] Interestingly, some bark beetle genera (Coleoptera: Scolytidae) produce GPP as a precursor of monoterpenes that function as aggregation and dispersion pheromones.[48] Farnesylgeranyl diphosphate (FGPP) synthase has also been purified from a haloalkaliphilic archaeon *Natronobacterium pharaonis*.[30]

1.14.2.1.2 Medium-chain (E)-prenyltransferases

Bacteria are good sources for prenyltransferases especially for the synthesis of long-chain prenyl diphosphates, because they produce menaquinones or ubiquinones having polyprenyl side chains whose chain lengths vary in a species-specific manner.[49] Thus, it was of interest to see whether there exist enzymes corresponding to the variety of the chain length of quinone side chains.

During the search for prenyltransferases in bacteria, we found that the C_{30} chain elongation is catalyzed by hexaprenyl diphosphate (HexPP) synthase (*trans*-pentaprenyltranstransferase (EC 2.5.1.33)), which has a novel subunit system. It catalyzes the synthesis of (all-*E*)-HexPP by adding three molecules of IPP to FPP, but it cannot catalyze the steps of synthesis of GPP or FPP from DMAPP and IPP ($C_5 \rightarrow C_{10} \rightarrow C_{15}$).[31] This enzyme occurs in *M. luteus* B-P 26 that produces menaquinone-C_{30}. Similarly, HepPP synthase ((*E*)-hexaprenyltranstransferase (EC 2.5.1.30)) has been found in *B. subtilis*, which produces exclusively menaquinone-C_{35}.[36] It catalyzes the synthesis of (all-*E*)-HepPP by adding four molecules of IPP to FPP, but it is not able to catalyze the $C_5 \rightarrow C_{10} \rightarrow C_{15}$ reactions either. FPP is supplied by FPP synthase, which occurs in all bacteria.

The purification of the HexPP synthase involved unexpected difficulties, not only due to the substrate specificities, but also due to the fact that the synthesis of HexPP requires two separable proteins, neither of which is catalytically active alone.[31] Therefore, the HexPP synthase had been elusive before this fact was disclosed. Similarly, two different proteins constituting HepPP synthase were separated and characterized from *B. subtilis*.[50] The components of HexPP synthase and HepPP synthase are so specific that neither of the two components of one enzyme is exchangeable with that of the other enzyme. Therefore, they appear to be novel heterodimeric enzymes with subunits easily dissociable under physiological conditions.

To obtain insight into the dynamic mechanism of the cooperative interaction between the components A and B of the HexPP synthase, we studied the properties of the heteromeric HexPP synthase using Superose 12 gel filtration.[51] Components A and B, when mixed without substrates, gave two elution peaks at 24 and 27 kDa positions, respectively. These eluates showed no enzyme activity. However, when a mixture of components A and B, which had been incubated in the presence of Mg^{2+} and FPP, was filtered on Superose 12, a protein fraction that showed HexPP synthase activity was eluted at a position of approximately 50 kDa, suggesting that this fraction corresponds to a complex, A–B substrate, which probably represents a catalytically active state of the enzyme. Formation of a complex of components A and B was also observed in the presence of a relatively high concentration of PP_i or one of the substrates, IPP and FPP.[52] These results have led us to propose a mechanism of this unusual enzyme system as shown in **Figure 4**. In the presence of Mg^{2+} and FPP, components A and B exist as a ternary complex of the two components and FPP + Mg^{2+}, which is catalytically active as HexPP synthase. After three molecules of IPP are condensed to produce HexPP, which is water insoluble, the complex dissociates into components A and B with a

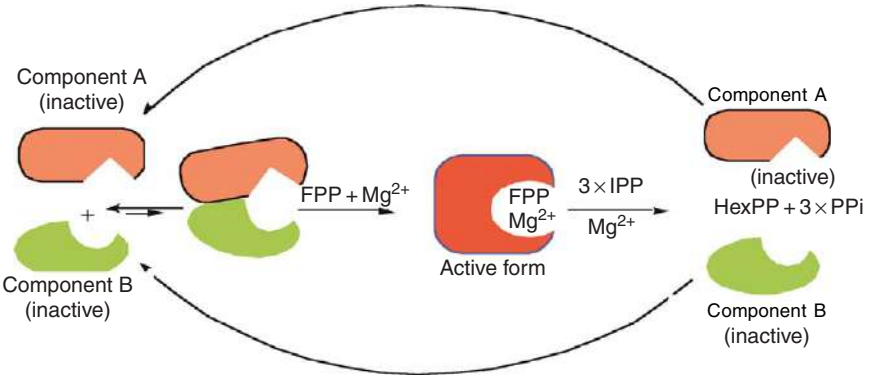


Figure 4 Mechanism of the dynamic formation and dissociation of the ternary complex of the heteromeric components of HexPP synthase.

concomitant release of HexPP and PP_i (**Figure 4**). This might account for the ability of these soluble proteins to exert an efficient turnover in the synthesis of amphipathic molecules from the water-soluble substrate without association with membrane components. By inactivation studies using Cys- and Arg-specific reagents, we have shown that the catalytic site of the HexPP synthase is formed by cooperative interaction between the two components and that Cys and Arg residues on component B play important roles in the catalytic activity.[32]

1.14.2.1.3 Long-chain (E)-prenyltransferases

By analogy with HexPP and HepPP synthases, it was predicted that the enzymes for the synthesis of polyprenyl diphosphate with chain lengths longer than C_{35} might also be composed of two-component systems. However, this is not the case. (*E*)-Nonaprenyl (solanesyl) diphosphate (SPP) synthase (*trans*-octaprenyltranstransferase (EC 2.5.1.11)) was isolated from *M. luteus*[53] as an enzyme catalyzing the chain elongation up to C_{45}-PP from GPP (C_{10}), and it was found that this enzyme is a homodimeric protein that is functionally active by itself.[39] In this respect, it resembles the enzymes in the group of prenyltransferase I, but it differs in that it requires a protein factor to maintain efficient catalytic turnover. This protein factor was separated as a high-molecular component of a soluble fraction of the same bacterium. It stimulates dramatically the enzymatic synthesis of SPP in a time-dependent manner, but it does not affect the stability of the enzyme.[39] Bovine serum albumin or a detergent such as Tween 80 can substitute for this protein, showing a similar mode of stimulation. These results suggest that the protein facilitates catalytic turnover by removing from the active site hydrophobic products, which otherwise would inhibit the reaction.

Bacterial (all-*E*)-octaprenyl diphosphate (C_{40}) synthase and (all-*E*)-decaprenyl diphosphate (C_{50}) synthase have been found and partially purified from *Escherichia coli*[54] and *Paracoccus denitrificans*,[40] respectively. These polyprenyl diphosphate synthases have properties similar to those of the SPP synthase described above. The protein from *M. luteus*, which stimulates the SPP synthase, acts with these synthases as well.[39] However, the same protein shows no stimulation on *E. coli* undecaprenyl diphosphate (UPP) synthase, which catalyzes the synthesis of (*Z*)-prenyl chains. Therefore, this protein is effective specifically on (all-*E*)-polyprenyl diphosphates with chain lengths of C_{40}, C_{45}, and C_{50}.

1.14.2.2 Stereochemistry of (*E*)-Prenyltransferases

From the viewpoint of organic reactions constructing carbon–carbon bonds, the mode of the reaction catalyzed by prenyltransferases is unique and interesting. The 1′–4 condensation reaction between an allylic diphosphate, such as DMAPP, and IPP proceeds with migration of the double bond at C-1 of IPP by a unique electrophilic condensation involving carbonium ions as reactive intermediates. Iteration of condensation reaction is possible in such a way.

In the 1′–4 condensation catalyzed by FPP synthase:

1. The condensation occurs with inversion of configuration at the C-1 methylene of the allylic diphosphate, DMAPP, or GPP.[55]
2. The 1′–4 bond is constructed by the addition of the allylic moiety to the *si* face of IPP.[56]
3. The *pro-R* hydrogen at C-2 of IPP is lost when the new double bond is formed.[55]

The addition of an allylic prenyl moiety occurs on the same side where the C-2 hydrogen of IPP is eliminated during the 1′–4 condensation. Since the same stereochemical relationships as described above were found in both mammalian liver and yeast, the stereochemistry of 1′–4 condensation is thought to be a highly conserved property of prenyltransferases.

It is interesting that the proton abstraction at C-2 and the addition of a prenyl group to C-4 of IPP during the transferase reaction occur at the same face of IPP as depicted in **Figure 5**, whereas the proton abstraction at C-2 and the addition to C-4 and the addition to C-4 of IPP during the IPP isomerase reaction occur at the opposite faces of the IPP molecule as shown in **Figure 5**. A ternary FPP–IPP–thioanalogue of DMAPP complex structure revealed the catalytic mechanism in terms of 3D structure.[57]

1.14.2.3 Three-Dimensional Structure of (*E*)-Prenyltransferases

1.14.2.3.1 FPP synthase

FPP synthase is one of the key enzymes responsible for cholesterol biosynthesis and exhibits changes in the level of its activity in response to the level of cholesterol in mammalian cells. Taking advantage of this fact, Clarke *et al.*[58] succeeded in isolation of a cDNA for rat liver FPP synthase. It is worth noting that the 30 amino acid 'active-site peptide' of chicken liver FPP synthase, which Rilling and coworkers had identified by photoaffinity labeling with *o*-azidophenylethyl diphosphate, was helpful as the convincing clue for the identification of the gene. Since then, cDNAs or genomic clones encoding FPP synthase have been isolated from various organisms, including human,[59,60] yeast,[61] *E. coli*,[62] *Bacillus stearothermophilus*,[25] white lupin,[63] *A. thaliana*,[64] and rubber tree.[65]

The first prenyltransferase that was purified to homogeneity was the FPP synthase from *Saccharomyces cerevisiae*.[22] Crystallization of FPP synthase was first achieved for avian liver enzyme by Reed and Rilling in 1975.[66] After 20 years, *B. stearothermophilus* FPP synthase was crystallized and a preliminary X-ray diffraction analysis was carried out to about 3 Å resolution.[67] Meanwhile, Sachettini and coworkers[68] made a great contribution to this field of study by succeeding in the crystal structure determination of avian liver FPP synthase to 2.6 Å resolution. This (*E*)-prenyltransferase enzyme is a homodimer and possesses a characteristic structure that is composed of many antiparallel α-helices forming a large cavity in each subunit (**Figure 6**). X-ray crystallographic studies have revealed that many enzymes related to isoprenoid biosynthesis, including pentalenene synthase,[70] squalene cyclase,[71] 5-*epi*-aristolochene synthase,[72] protein farnesyltransferase,[73] and FPP synthase,[68] have a common structure called isoprenoid synthase fold[74] (or terpenoid synthase fold), which is composed of 11–12 α-helices forming a large central cavity. Since then, it has been believed that the common isoprenoid synthase fold is included in all enzymes related to isoprenoid biosynthesis.

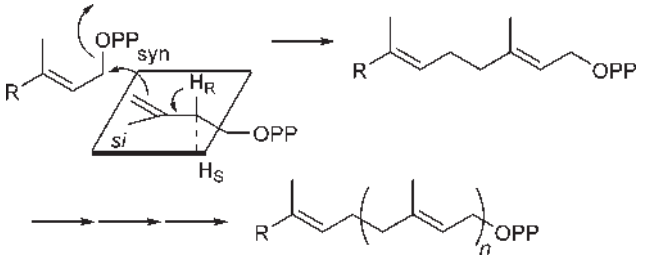


Figure 5 Stereochemistry of the 1′–4 condensation by FPP synthase.

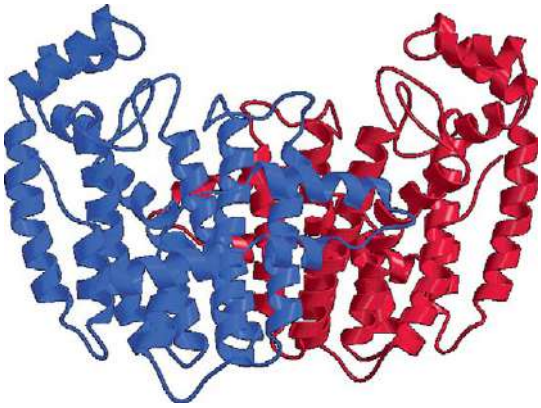


Figure 6 Crystal structure of avian liver FPP synthase (Protein Data Bank ID, 1FPS).[68] Figures were drawn by PyMol.[69]

1.14.2.3.2 Geranylgeranyl diphosphate synthase

The first enzyme whose crystal structure was determined among the GGPP synthases was from *S. cerevisiae*.[75] The structure is composed of 15 α-helices joined by connecting loops and is arranged with α-helices around a large central cavity. The N-terminal 17 amino acids of this GGPP synthase protrude from the helix core into the other subunit and contribute to the tight dimer formation. The crystal structure of human GGPP synthase reveals a hexameric arrangement.[76] Three dimers join together to form a propeller-bladed hexameric molecule. Comparison of primary sequence suggested that this quaternary structure is restricted to mammalian and insect orthologues. GGPP is tightly bound in a cavity distinct from the chain-elongation site for FPP synthase.

1.14.2.4 Catalytic Mechanism

1.14.2.4.1 Roles of conserved amino acid residues

To evaluate the roles of amino acids in the conserved regions in FPP synthases, a number of site-directed mutagenesis experiments have been carried out. Joly and Edwards[23] showed that the mutagenesis of Asp104, Asp107, Arg112, Arg113, and Asp243 of rat liver FPP synthase, which correspond to D2, D3, R1, R2, and D4 in regions II or VI, respectively (see **Figure 7**), resulted in a decreased V_{max} of approximately 1000-fold compared to the wild-type enzyme. However, no significant changes in the K_m values for either IPP or GPP were observed in the conservative mutation in which Asp and Arg residues were replaced with Glu and Lys, respectively. These results suggest that these amino acid residues are involved in catalytic efficiency rather than binding affinity. On the other hand, Asp103 (D1) and Asp247 (D6) could be replaced with Glu without marked changes in kinetic properties. Using a recombinant form of yeast enzyme extended by a C-terminal -Glu-Glu-Phe α-tubulin epitope, Song and Poulter[80] conducted mutagenesis studies in which the conserved Asp or Arg residue was replaced with Ala. They showed that the mutations of the Asp residues and nearby Arg residues in region II (D1, D2, R1, and R2) and the Asp residues in region VI (D4 and D5) drastically lowered the catalytic activity of the fused enzyme. However, the third Asp in region VI (D6) and Lys254, which is conserved downstream in region VI, were found to be replaced with Ala without marked changes in kinetic properties. We also showed that the mutagenesis of Asp224 and Asp225 (D4 and D5) of *B. stearothermophilus* FPP synthase resulted in an approximately 104–105-fold decrease in the k_{cat} values compared to the wild type.[81] On the other hand, replacement of Asp228 (D6) with Ala resulted in an approximately 10-fold decrease in the k_{cat} value and a 10-fold increase in the K_m value for IPP.

Taken altogether, these results indicate the importance of the six Asp residues in the consensus D1D2xxD3xxxxR1R2G (region II) and GxxFQxxD4D5xxD6 (region VI) motifs as follows: D2, D3, and D4 are so critical that none of them can be replaced by even Glu without significant decrease in catalytic activity, whereas D1 can be replaced by Glu but not by Ala. However, D6 is not so critical that even Ala can substitute for it.

Replacement of the first Asp residue (D4, Asp243) in the GxxFQxxD4DxxD motif in region VI of rat FPP synthase with Glu resulted in a 26-fold increase in the K_m value for IPP.[82] Replacement of the conserved Lys

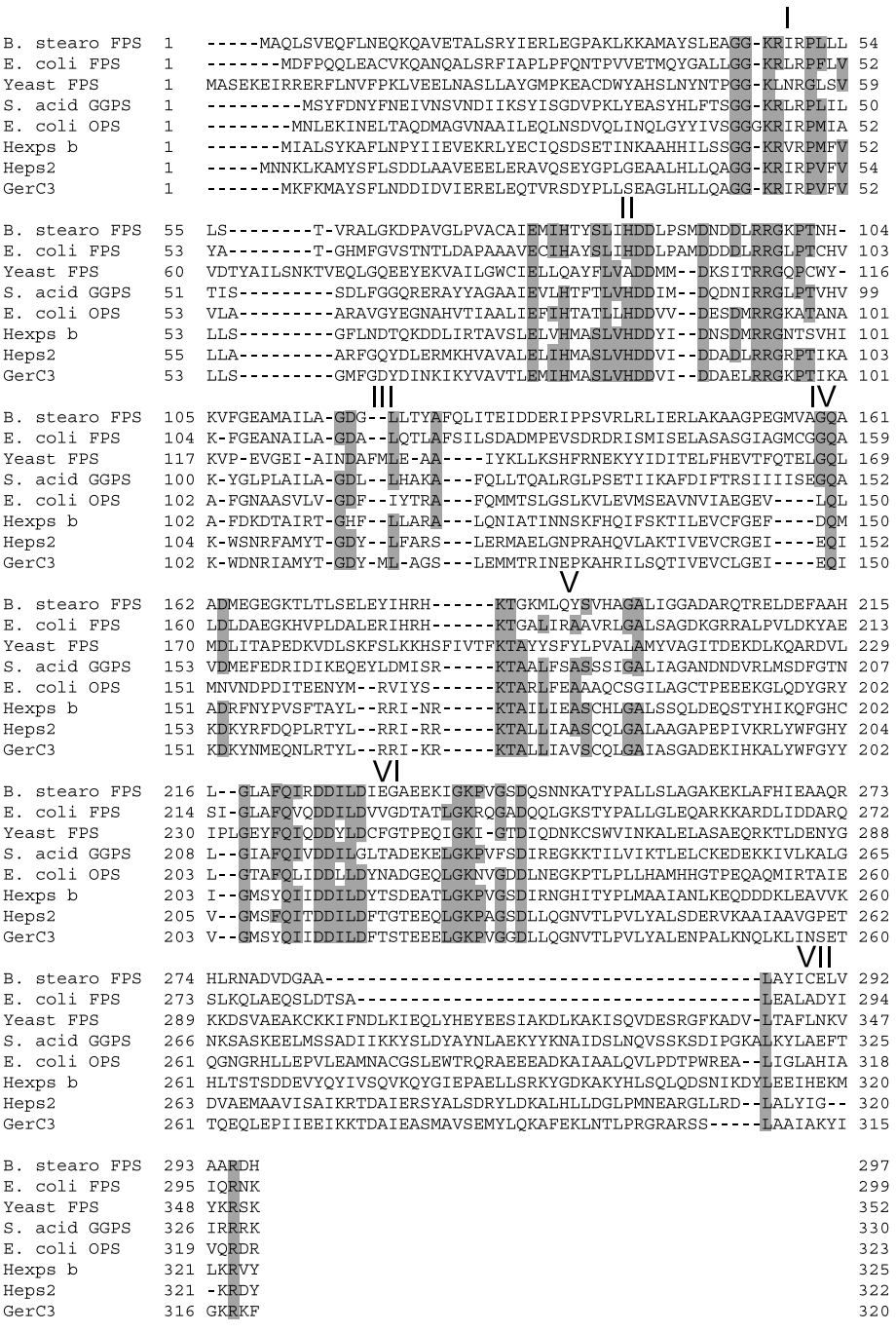


Figure 7 Amino acid sequence alignment of various prenyltransferases: B. stearo FPS, FPP synthase of *Bacillus stearothermophilus*;[25] E. coli FPS, FPP synthase of *Escherichia coli*;[62] Yeast FPS, FPP synthase of *Saccharomyces cerevisiae*;[61] S. acid GGPS, GGPP synthase of *Sulfolobus acidocaldarius*;[26] E. coli OPS, OctPP synthase of *E. coli*;[24] Hexps b, component B of HexPP synthase of *Micrococcus luteus* B-P 26;[77] Heps2, component II of HepPP synthase of *B. stearothermophilus*;[78] GerC3, component II of HepPP synthase of *Bacillus subtilis*.[79] The conserved amino acid residues in the seven conserved regions I to VII are shaded.

residues (Lys47 and Lys183) in regions I and V of *B. stearothermophilus* FPP synthase with aliphatic amino acids also resulted in marked decrease of the enzyme affinity for IPP, thereby causing a change in the distribution of products.[81] Both mutants K47I and K183A have k_{cat} values approximately 70-fold lower than that of the wild-type enzyme. Neither of them show significantly changed K_m values for GPP, but they both show markedly increased K_m values for IPP. These facts indicate that both Lys residues conserved in regions I and V contribute to the binding affinity for IPP as well as D6 (or D4 in case of rat enzyme) in region VI.

Blanchard and Karst[29] isolated a mutant gene from a yeast strain that had the unusual property of excreting prenyl alcohols such as geraniol. By comparing the nucleotide sequence of the FPP synthase of the mutant, they have shown that the unusual property is attributed to a one-point mutation resulting in a substitution of Glu for Lys197 of *S. cerevisiae* FPP synthase, which is the conserved Lys in region V and which corresponds to Lys183 in the *B. stearothermophilus* enzyme. This is interesting in the light of the above-mentioned observation[81] that both K47I and K183A showed markedly increased K_m values for IPP, since a decrease of the affinity for IPP might also cause an accumulation of GPP. We examined this possibility by analyzing the relative accumulation of GPP in the FPP synthase reaction by the mutant enzymes, and obtained similar results[81] to those observed by Blanchard and Karst[29] with the yeast mutant extracts. It is worth noting that K47I and K183A of *B. stearothermophilus* enzyme resemble each other with respect to both kinetic parameters and product distributions. These results raise the possibility that the release of GPP in the reaction catalyzed by the yeast mutant enzyme might also be due to depression of the binding affinity for IPP rather than for GPP.

The significance of the Phe-Gln motif (FQ) in region VI was examined by us.[83] The highly conserved Phe220 and Gln221 residues were replaced with Ala and Glu, respectively. These mutations resulted in 10^5- and 10^3-fold decreases in catalytic activity of the FPP synthase from *B. stearothermophilus*, respectively, indicating that the Phe residue has a crucial role in catalysis. The Michaelis constants of the Q221E mutant for DMAPP and GPP increased by approximately 25- and 2-fold, respectively, compared to the wild type, whereas those for IPP were not altered much. These results suggest that the FQ motif is involved not only in substrate binding but also in catalysis. These facts suggest a hypothetical model for the binding of an allylic diphosphate to region VI. The dramatic effect on k_{cat} by the replacement of the Phe group with Ala suggests that Phe220 might contribute to the acceleration of catalysis by stabilizing the prenyl cation formed at the beginning of the catalytic reaction through a cation–π interaction.[84]

Mammalian FPP synthase shares a C-terminal region (region VII) characterized as being so crowded with basic amino acids as shown in **Figure 7**. However, only the Arg residue at the third position from the C-terminus is completely conservative if bacterial FPP synthases are taken into account. To examine the significance of this residue, Arg295 in the *B. stearothermophilus* enzyme was mutated to Val.[85] This mutation, however, did not result in significant change in catalytic activity, indicating that the conserved Arg residue in region VII is not essential for catalysis. The R295V mutant showed a threefold increased K_m value for IPP but showed an unchanged K_m value for GPP as compared with the wild type. Song and Poulter[80] also examined the significance of the Arg350 near the C-terminal of the yeast enzyme having an -EEF epitope. The R350A mutant showed only a slight additional increase in the K_m value for IPP. The addition of the -EEF epitope to the C-terminus of the wild-type enzyme resulted in a 14-fold increase of the K_m value for IPP and a 12-fold decrease of k_{cat}.

Bacillus stearothermophilus FPP synthase is unique in that it possesses only two Cys residues in contrast to FPP synthases from other sources that have more than four Cys residues. To explore the significance of these Cys residues, we examined the effect of replacement of the Cys residues at 73 and 289 of this thermostable enzyme with Ser.[86] As a result, all of the mutant enzymes were active as FPP synthase, showing specific activities comparable to that of the wild-type enzyme. These results indicate that neither of the Cys residues is essential for the catalytic function.

Hosfield *et al.*[57] reported the crystal structure of *E. coli* FPP synthase in complex with IPP and a thioanalogue of DMAPP. The substrate analogue, whose bridging oxygen atom between P-1 and C-1 in DMAPP substituted with sulfur, showed 10^6-fold decrease in k_{cat} and thus effectively inhibits the FPP synthase against DMAPP.[87] The complex structure reveals how enzyme conformational changes organize conserved active site residues to utilize metal-induced ionization and substrate positioning for catalysis.[57]

1.14.2.4.2 Random chemical mutagenesis

Random chemical mutagenesis often provides a powerful method for obtaining a desired mutation product if an effective procedure for screening positive clones is available. To clone the GGPP synthase gene of *Sulfolobus acidocaldarius*, Ohnuma *et al.*[26] have developed an *in vivo* method for detecting the enzymatic activity by utilizing carotenoid biosynthesis genes of *Erwinia uredovora*[88] to produce a red-colored clone expressing GGPP synthase activity.

The FPP synthase gene of *B. stearothermophilus* was subjected to random mutagenesis by $NaNO_2$ treatment to construct libraries of mutated FPP synthase genes. From the libraries, mutants showing GGPP synthase activities were selected by the red–white screening method, and 11 red positive clones were obtained from 24 300 mutants.[89] Each mutant was found to contain a few amino acid substitutions in the FPP synthase, which resulted in acquisition of the catalytic activity of synthesizing GGPP as well as FPP. Mutants that had replacement of Tyr81 with His showed the most efficient production of GGPP. From the analysis of the mutations, it was possible to define three amino acids that determine the final chain length of the products. They were Leu34, Tyr81, and Val157. In particular, the mutated enzyme that has a substitution of His for Tyr81, which is situated at the fifth amino acid upstream to the DDxxD motif (FARM, first aspartate-rich motif) in region II, produced GGPP most effectively.

To investigate the role of Tyr81 of *B. stearothermophilus* FPP synthase, Ohnuma *et al.*[90] constructed 19 mutant enzymes each of which has a different amino acid at position 81. All enzymes except Y81P showed prenyltransferase activities to catalyze condensation of IPP with an allylic diphosphate. When assayed with FPP as the allylic substrate, considerable activities were observed in almost all mutated FPP synthases. These results indicate that the mutated enzymes can catalyze the chain elongation beyond FPP. When DMAPP or GPP was used as the allylic primer, the mutated enzymes, Y81A, Y81G, and Y81S, produced HexPP as the longest product. These mutants produced geranylfarnesyl diphosphate (GFPP, C_{25}) as the main product. Y81A and Y81G gave larger amounts of HexPP and GFPP than did Y81S. These mutant enzymes, in which Tyr81 was replaced with Cys, His, Ile, Leu, Gln, Thr, or Val, all produced GFPP as the longest product. These observations strongly indicate that the amino acid Tyr81 directly contacts the ω-terminal of the final (longest) product during the catalytic isoprenoid chain elongation. This interaction in the catalytic site must be the critical step in determining the chain length of the product of prenyltransferase.

Assuming that the essential structure of the binding cavity of the *B. stearothermophilus* FPP synthase is similar to that of the avian FPP synthase in view of the sequence similarity between them, it is reasonable to predict, on the basis of the crystal structure of the avian enzyme,[68] that the Tyr81 is situated on a point 11–12 Å apart from the first Asp-rich motif in the large central cavity of the FPP synthase. The distance is similar to the length of the hydrocarbon moiety of an allylic substrate bound by the diphosphate moiety at the Asp motif through a magnesium ion. As the distance is similar to the length of the hydrocarbon moiety of FPP, it is suggested that the aromatic ring of Tyr81 might prevent further chain elongations longer than C_{15}.

1.14.2.5 Product Chain-Length Regulation Mechanism

One of the most interesting research topics on prenyl chain-elongating enzymes is to understand the mechanisms by which individual prenyltransferases recognize the prenyl chain lengths of allylic substrates and products. Tarshis *et al.*[91] have reported strong evidence on the mechanism of regulation of product chain length in the avian FPP synthase reaction by X-ray analyses of some mutated enzymes that acquired the catalytic capability of producing longer chain prenyl diphosphates. An analysis of the X-ray structure[68] of the avian FPP synthase, coupled with the information about conserved amino acids of many prenyltransferases so far cloned, led them to the idea that the phenyl groups of Phe112 and Phe113 in the avian enzyme, which are situated at the fifth and fourth amino acid upstream to the DDxxD motif in region II, are important for determining the ultimate length of the hydrocarbon chains. Then they carried out several site-directed mutations in the avian enzyme with respect to these Phe residues. As a result, enzymes capable of producing GGPP (F112A), GFPP (F113S), and longer-chain prenyl diphosphates (F112A/F113S) were obtained. X-ray analyses of the structure of the F112A/F113S mutant in the absence or presence of allylic substrates indicated an alteration of the size of the binding pocket for the growing isoprenoid chain in the active site of the enzyme. The proposed binding pocket in the mutant structure was increased in depth by 5.8 Å as compared with that for

the wild-type enzyme. Allylic diphosphates were observed in the holo structures, bound through Mg^{2+} to the first two Asp residues in the DDxxD motif (Asp117, Asp121), with the hydrocarbon tails of all the ligands growing down the hydrophobic pocket toward the mutation site.

From X-ray analyses of the wild-type avian FPP synthase, Tarshis *et al.*[91] have suggested that the ultimate length of the polyisoprenoid chain obtained during successive condensations of the growing allylic substrate with IPP is governed by the size of a hydrophobic pocket in the interior of the enzyme. They also showed that replacement of the benzyl groups in Phe112 and Phe113 with smaller side chains gave FPP synthase mutants that synthesized longer isoprenoid products than FPP.[91]

Similarly, Zhang *et al.*[92] and Hirooka *et al.*[93] have shown the important role of the fifth amino acid upstream from the FARM even in the medium-chain (*E*)-prenyltransferase, HepPP synthase, of *B. subtilis*. In this dissociable heterosubunit enzyme system, both subunits cooperatively participate in product chain-length determination.

According to their intensive examinations on *S. acidocaldarius* GGPP synthase, Ohnuma *et al.*[89,94] showed that the ultimate product chain length of (*E*)-prenyltransferases was regulated by bulky residues located at the fifth position upstream to the first DDxxD motif (FARM) in region II, which is one of the most highly conserved regions among (*E*)-prenyltransferases. The library of mutated GGPP synthase genes was screened with a mutant deficient in HexPP synthase. As a result, three mutant enzymes showing catalytic activity to produce large amounts of FGPP with a concomitant formation of small amounts of HexPP were obtained. Sequence analysis revealed that the mutation of Phe77, which is located at the fifth amino acid upstream from the FARM in region II of prenyltransferases, is the most effective for elongating the ultimate product. This fact exactly coincides with the results on FPP synthase mutation.[89,90]

By comparison of the amino acid sequences of many kinds of FPP synthases with those of GGPP synthases, Ohnuma *et al.*[95] noticed that several homologous regions typical for FPP synthase are found around the FARM motif in region II. They then introduced mutations into the region II of the archaeal GGPP synthase. In these mutants, regions around the FARM motif of GGPP synthase were replaced with the corresponding regions of FPP synthases from human,[60] rat,[58] *A. thaliana*,[64] *S. cerevisiae*,[61] *E. coli*,[62] and *B. stearothermophilus*.[25] By analyzing these mutated enzymes, they found that the region around the first DDxxD motif is essential for the product specificity of all FPP synthases and that the sequence in this region of FPP synthase differs between eukaryotic and prokaryotic enzymes. On the basis of these observations, they have proposed that FPP synthases have evolved from a progenitor corresponding to the archaeal GGPP synthases in two ways. For GGPP synthase, compared with the major product C_{30} synthesized by mutant H139A, the products generated by mutant Y107A and F108A are predominantly C_{40} and C_{30}, respectively, suggesting the most important role of Tyr107 in determining the product chain length.[75]

As shown above, in most types of short-chain (all-*E*)-prenyl diphosphate synthases, bulky amino acids at the fourth and/or fifth positions upstream from the FARM motif play a primary role in the product determination mechanism. GGPP synthase from *Pantoea ananatis*, however, lacks such bulky amino acids at these positions. The second position upstream from the G(Q/E) motif has recently been shown to participate in the mechanism of chain-length determination in GGPP synthase. Amino acid substitutions adjacent to the residues upstream from the FARM motif and from the G(Q/E) motif did not affect the chain length of the final product.[96] Bulky amino acids in the α-helix located at the expected subunit interface were replaced with alanine. Two mutants, which gave products with longer chain lengths, suggesting that type GGPP synthase from *P. ananatis* utilizes the mechanism of chain-length determination, which requires subunit interaction in the homooligomeric enzyme.[96]

1.14.3 (*Z*)-Prenyl Diphosphate Synthases

(*Z*)-Prenyltransferase catalyzes the consecutive cis-condensation of IPP with an allylic diphosphate to synthesize long-chain prenyl diphosphates, which often occur in association with the membrane fraction of bacterial cell wall and the microsomal fraction of eukaryotic cells. Despite the similarities of the primer substrates in the prenyl chain-elongation reaction, the final biosynthetic products by (*Z*)-prenyltransferases seem more diverse than (*E*)-prenyltransferase, which varies from *cis*-FPP (C_{15}) to natural rubber whose prenyl chain length extends to several millions (**Table 2**). Among the (*Z*)-prenyltransferases in prokaryotes, UPP synthase has

Table 2 Identified genes and structures of *cis*-prenyl diphosphates in prokaryotes and eukaryotes

Species	*Gene name*	*Enzyme involved*	*Structure*[a]	*Isoprene units*	*Reference(s)*
Eubacteria					
Micrococcus luteus	UPS	Undecaprenyl diphosphate synthase	ω-t_2-c_8-PP	C_{55}	20
Escherichia coli	UPS	Undecaprenyl diphosphate synthase	ω-t_2-c_8-PP	C_{55}	19
Haemophilus influenzae	UPS	Undecaprenyl diphosphate synthase	ω-t_2-c_8-PP	C_{55}	19
Streptococcus pneumoniae	UPS	Undecaprenyl diphosphate synthase	ω-t_2-c_8-PP	C_{55}	19
Mycobacteria					
Mycobacterium tuberculosis	Rv1086	Farnesyl diphosphate synthase	ω-t-c-PP	C_{15}	97
M. tuberculosis	Rv2361c	Decaprenyl diphosphate synthase	ω-t-c_8-PP	C_{50}	97
Archaea					
Sulfolobus acidocaldarius	CPDS	Undecaprenyl diphosphate synthase	ω-t_2-c_8-PP	C_{55}	98
Yeast					
Saccharomyces cerevisiae	RER2	Dehydrodolichyl diphosphate synthase	ω-t_2-c_{11}-S-PP	C_{75}	99
S. cerevisiae	STR1	Dehydrodolichyl diphosphate synthase	ω-t_2-c_{15}-S-PP	C_{95}	110
Plant					
Arabidopsis thaliana	DeDolPS	Dehydrodolichyl diphosphate synthase	ω-t_2-c_{15-20}-S-PP	C_{95}–C_{120}	101, 102
Hevea brasiliensis	HRT	Rubber synthase	–	$C_{\sim 25\ 000}$	103
Human					
	DeDolPS	Dehydrodolichyl diphosphate synthase	ω-t_2-c_{15-16}-S-PP	C_{95}–C_{100}	104

[a] ω, terminal isoprene unit; t, *trans*-isoprene unit; S, α-saturated isoprene unit; c, *cis*-isoprene unit.

been studied the most in detail. UPP synthase (di-*trans*, poly-*cis*-decaprenyl *cis* transferase, EC 2.5.1.31) catalyzes the construction of a (*Z*)-prenyl chain onto (all-*E*)-FPP as a primer to yield a C_{55}-prenyl diphosphate with (*E*)-, (*Z*)-mixed stereochemistry. The roles are to produce the precursors of polyprenyl lipids required as carbohydrate carrier in the biosynthesis of bacterial cell wall or of glycoproteins in eukaryotic cells. Earlier works were performed with partially purified protein samples from several bacteria including *Salmonella newington*,[105] *M. luteus*,[106,107] *B. subtilis*,[108] and *E. coli*,[54] and most extensive studies were performed from *Lactobacillus plantarum*.[109–111,97] The enzyme was shown to be an acidic protein (p*I* = 5.1) and the estimated molecular weight was 56 kDa. The enzyme is associated with periplasmic membrane loosely and is easily solubilized, and requires a phospholipid or detergent such as Triton X-100 for its enzymatic activity. It is interesting that *P. denitrificans*, which is assumed to be an ancestor of mitochondria, has (all-*E*)-farnesyl-(all-*Z*)-hexaprenyl (C_{45}) diphosphate synthase[112] instead of UPP synthase, which is common in general bacteria. Koyama *et al.*[113] extracted and partially purified a polyprenoyl diphosphate synthase from mulberry leaves, *Morus bombysis*. As all (*E*)-prenyltransferases require the divalent ion for the activity, the Mg^{2+} ion is essential for catalysis by the UPP synthase. The optimal concentration for the divalent metal ion is usually lower than that of (*E*)-prenyltransferases.

Unlike the all-α-helix structures of many (*E*)-prenyltransferases, the crystal structure of (*Z*)-prenyl diphosphate synthase from *M. luteus* B-P 26,[18] *E. coli*,[114] and *Mycobacterium tuberculosis*[115] revealed that it forms a homodimer consisting of both α-helix and β-sheet, suggesting the distinctive catalytic and chain-length regulation mechanism.

In this section, we will consider the structural basis for the synthesis of prenyl diphosphate and chain-length regulation mechanism by (*Z*)-prenyltransferases based on the recent studies using X-ray crystallography and site-directed mutagenesis.

1.14.3.1 Classification of (*Z*)-Prenyltransferases

(*Z*)-Prenyltransferases can be classified into three subfamilies depending on their product chain lengths: (1) short-chain-length (*Z*)-prenyltransferases, such as *cis*-FPP (C_{15}) synthase from *M. tuberculosis*, (2) medium-chain-length (*Z*)-prenyltransferases, such as *M. luteus* B-P 26 UPP (C_{55}) synthase, and (3) long-chain-length (*Z*)-prenyltransferases including the dehydrodolichyl diphosphate (DedolPP) synthase, which produces biosynthetic intermediates (C_{80}–C_{120}) of dolichol, glycosyl carrier lipids in the biosynthesis of N-linked glycoproteins (**Figure 2**). Undecaprenyl phosphate (UP, a monophosphate form of UPP), which contains 11 unsaturated isoprene units, is required for carbohydrate carrier lipid in the biosynthesis of bacterial cell wall. In eubacteria and archaea, usually UPP synthase catalyzes the consecutive addition of isoprene unit onto FPP to produce UPP with mixed (*E*)-, (*Z*)-stereochemistry. Sequential condensation of all eight molecules of IPP proceeds in the cis-configuration up to the final product UPP. Mycobacteria are unique in that they contain two types of (*Z*)-prenyltransferases: one catalyzes short-chain C_{15} production and the other synthesizes a longer prenyl chain of C_{50}.[116]

In eukaryotes, DedolPP synthase catalyzes the sequential cis-condensation of IPP to form DedolPP. Dolichyl phosphate (Dol-P), the α-saturated and a single dephosphorylated form of DedolPP, acts as a sugar carrier lipid in the biosynthesis of N-glycosylated proteins and GPI-anchored proteins. The chain length of Dol-P has been shown to be significantly specific to the species. Plant contains dolichols with 13–22 isoprene units,[117–120] whereas yeast possesses dolichols with 14–17 isoprene units,[120] and in mammalian cells, the dolichols with 18–21 isoprene units are predominant.[121] The sugar carrier lipids and the bacterial counterpart are different in terms of the prenyl chain length and the saturation of the isoprene unit. Many higher plants also produce polyprenols with mixed (*E*)-, (*Z*)-stereochemistry, betulaprenols, ficaprenols, and natural rubber in addition to dolichols.

1.14.3.1.1 Short- and long-chain (Z)-prenyltransferases

While the genes for many kinds of prenyltransferases that catalyze (*E*)-type prenyl chain elongations had been identified and characterized, little had been known about the structure of (*Z*)-prenyltransferase at the molecular level until the gene cloning of the UPP synthase from *M. luteus* B-P 26 by Shimizu *et al.*[20] using the colony autoradiography method developed by Raetz.[122] This method requires incubation of a replica

membrane containing a genomic library in the usual incubation mixture with a radiolabeled substrate for the target enzyme. After the incubation, the replica filters were subjected to autoradiography to detect radiolabeled colonies. Surprisingly, the primary structure of the UPP synthase from *M. luteus* B-P 26 is totally different from those of the prenyl diphosphate synthases such as FPP synthase and GGPP synthase catalyzing *E*-prenyl chain elongation, which have characteristic aspartate-rich motifs of DDxxD. Shortly, Apfel *et al.*[19] reported the homologous gene for this enzyme from *E. coli*, *Haemophilus influenzae*, and *Streptococcus pneumoniae*. Two types of (*Z*)-prenyltransferases, Rv1086, a short (C_{15})-chain (*Z*)-prenyltransferase and Rv2361c, a long (C_{50})-chain (*Z*)-prenyltransferase were identified from *M. tuberculosis*.[116] Similarly, some pairs of short-chain (*Z*)-prenyltransferases and long-chain (*Z*)-prenyltransferases were identified from *Corynebacterium efficiens*, *Corynebacterium glutamicum*, and *Thermobifida fusca* in our laboratory.[123] Hemmi *et al.*[99] cloned and characterized a (*Z*)-prenyltransferase from an archaeon, *S. acidocaldarius*.

1.14.3.1.2 Long-chain (Z)-prenyltransferases

In eukaryotes, DedolPP synthase catalyzes the *cis*-prenyl chain elongation. In an earlier work, DedolPP synthase was solubilized and characterized from the membrane fraction of Ehrlich ascites tumor cells,[100] rat tubulin,[101] *Saccharomyces carlsbergensis*,[102] and rat liver.[104] Cloning of the corresponding cDNAs for eukaryotic (*Z*)-prenyltransferase genes has provided more definitive information on the structures at the molecular level. In 1999, Sato *et al.*[103] cloned and characterized the *RER2* gene encoding DedolPP synthase from *S. cerevisiae*. RER2p is involved in the correct localization of endoplasmic reticulum (ER) proteins. Kato *et al.*[124] reported the isolation and characterization of a homologous gene, *SRT1*, encoding another (*Z*)-prenyltransferase, which catalyzes longer *cis*-prenyl diphosphate synthesis in the same organism. Plant DedolPP synthase has been cloned and characterized from *A. thaliana*.[125,126] A cDNA of a human DedolPP synthase has been identified and characterized in our laboratory by Endo *et al.*[127] and a cDNA encoding a (*Z*)-prenyltransferase that may be engaged in rubber biosynthesis in *Hevea brasiliensis* has also been cloned.[128]

1.14.3.2 Stereochemistry of (*Z*)-Prenyltransferase

Since the overall stereochemistry of FPP synthase reaction was first established,[55,56] the stereochemistry of the hydrogen elimination from C-2 of IPP has been elucidated with various prenyltransferases. The stereochemistry of the C–C bond formation was examined only for the pig liver FPP synthase. The stereochemistry of enzymatic C–C bond formation by (*Z*)-prenyltransferase was an interesting problem. The success in the separation of medium- and long-chain prenyl diphosphate synthases of *B. subtilis*,[118,129] enabled us to determine the absolute stereochemistry involving in the (*Z*)-prenyl and (*E*)-prenyl chain elongations catalyzed by UPP and HepPP synthases, respectively.[17] The absolute stereochemistry of UPP synthase is illustrated in **Figure 8**.

1.14.3.3 Three-Dimensional Structure of (*Z*)-Prenyltransferase

Fujihashi *et al.*[18] determined the crystal structure of *M. luteus* B-P 26 UPP synthase as the first 3D structure for any *cis*-prenyl chain-elongating enzymes at 2.2 Å resolution (**Figure 9**). This enzyme acts as a homodimer composed of a set of 29 kDa subunits. Each monomer contains six parallel β-strands forming a central β-sheet core, which is

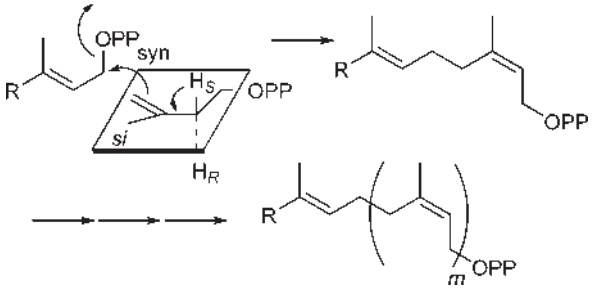


Figure 8 Stereochemistry of the 1′–4 condensation by UPP synthase.

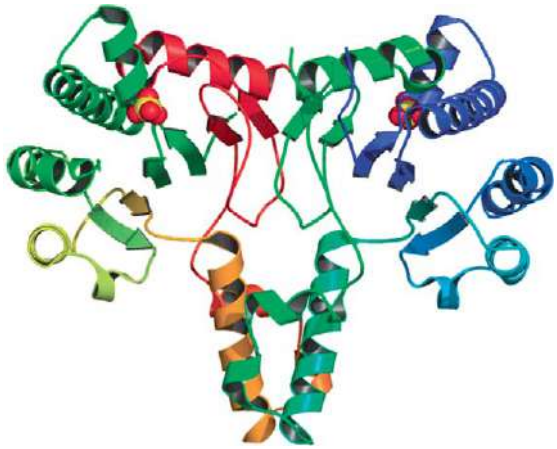


Figure 9 Crystal structure of *Micrococcus luteus* B-P 26 UPP synthase (Protein Data Bank ID, 1F75).[18] Figure was drawn by PyMol.[69]

surrounded by five of the seven α-helices. It is quite interesting that this enzyme shows a novel protein fold, which is totally different from the isoprenoid synthase fold,[74] as suggested for inclusion in most of the enzymes related to isoprenoid biosynthesis such as (*E*)-prenyltransferases (**Figure 7**). The isoprenoid synthase fold contains 10–12 mostly antiparallel α-helices, which can be seen in the structures of the FPP synthase, sesquiterpene cyclase, squalene synthase, and protein farnesyltransferase.[68,70–72] This suggested that the catalytic mechanism of (*Z*)-prenyltransferases, including substrate recognition, is also different from those of (*E*)-prenyltransferases.

Subsequent to the determination of the crystal structure of UPP synthase from *M. luteus* B-P 26, the 3D structure of *E. coli* UPP synthase was revealed,[114] and showed a nearly similar structure to that for *M. luteus* B-P 26 UPP synthase. Furthermore, the structure of the UPP synthase is quite different from that of a common structural motif called isoprenoid synthase fold. The crystal structures of UPP synthase from *E. coli* with allylic or homoallylic substrates were determined.[130,131] These structural analyses also provided insight into the relationship between the structure and catalytic function of *cis*-prenyltransferase. Thus, structural properties are discussed in the following section along with the molecular mechanisms of catalytic function. Recently, the crystal structures of UPP synthase from *Helicobacter pylori* – *E,Z*-FPP (C_{15}) synthase, Rv1086, and C_{50} prenyl diphosphate synthase, Rv2361c – were also determined.[115,132] A large elongated hydrophobic cleft, consisting of highly conserved amino acid residues, surrounded by two α-helices (H2 and H3) and two β-strands (S2 and S4), is found on its molecular surface. The interior of the cleft consists mostly of hydrophobic residues in which most of the conserved amino acid residues in the five conserved regions of *cis*-prenyltransferase are located. This cleft seems to have a suitable volume to keep the large hydrophobic prenyl chain of allylic substrate during the catalytic chain elongation up to C_{55}, the final product of the catalytic reaction. Therefore, this cleft is considered to function in recognition of the allylic substrate, FPP, which has a large hydrophobic prenyl chain, and accommodation of the elongated hydrophobic prenyl intermediates. A sulfate ion used for the crystallization as a precipitant is found to bind with a motif in region I, which is a common motif for phosphate recognition and is called a structural P-loop.[133] It is likely that the negatively charged diphosphate moiety of the allylic substrate FPP is recognized by this motif, where the sulfate ion is located. At the entrance of this cleft, basic amino acid residues including highly conserved amino acid residues (Arg197 and Arg203) form a positively charged cluster near the structural P-loop motif, which is predicted for the recognition of the diphosphate moiety of the homoallylic substrate. In the crystal structure of *M. luteus* B-P 26 UPP synthase without substrates, the residues from 74S to 85V, which belong to one of the conserved regions, region III, could not be incorporated in the final crystal structure because of their invisible electron densities,[18] suggesting that the flexible domain is important for the binding of substrates and catalytic function.

When the two equivalent monomers were superimposed in the dimer structure, it resulted in a good match of the Cα atoms with each other except for those in the region from 86N to 95F, which is located adjacent to the disordered loop (74S to 85V). These highly conserved and flexible regions seem to play an important role in the

catalytic function. The crystal structure of *E. coli* UPP synthase retained the similar structural features as *M. luteus* B-P 26 UPP synthase except for some deviations of some helices, mainly 3, 4, and 6. These helices might be intrinsically flexible and the differences might be owing to the different conditions utilized in the protein crystallization. In the *E. coli* UPP synthase structure, the conserved region III has also been shown as unidentified.[114]

1.14.3.4 Catalytic Mechanism

Figure 10 shows the multialignment of primary structures of the (*Z*)-prenyltransferases so far identified. There are five highly conserved regions (regions I–V) among the *cis*-prenyl chain-elongating enzymes. Comparison of these conserved regions with those of the (*E*)-prenyltransferases revealed no similarities at all between the

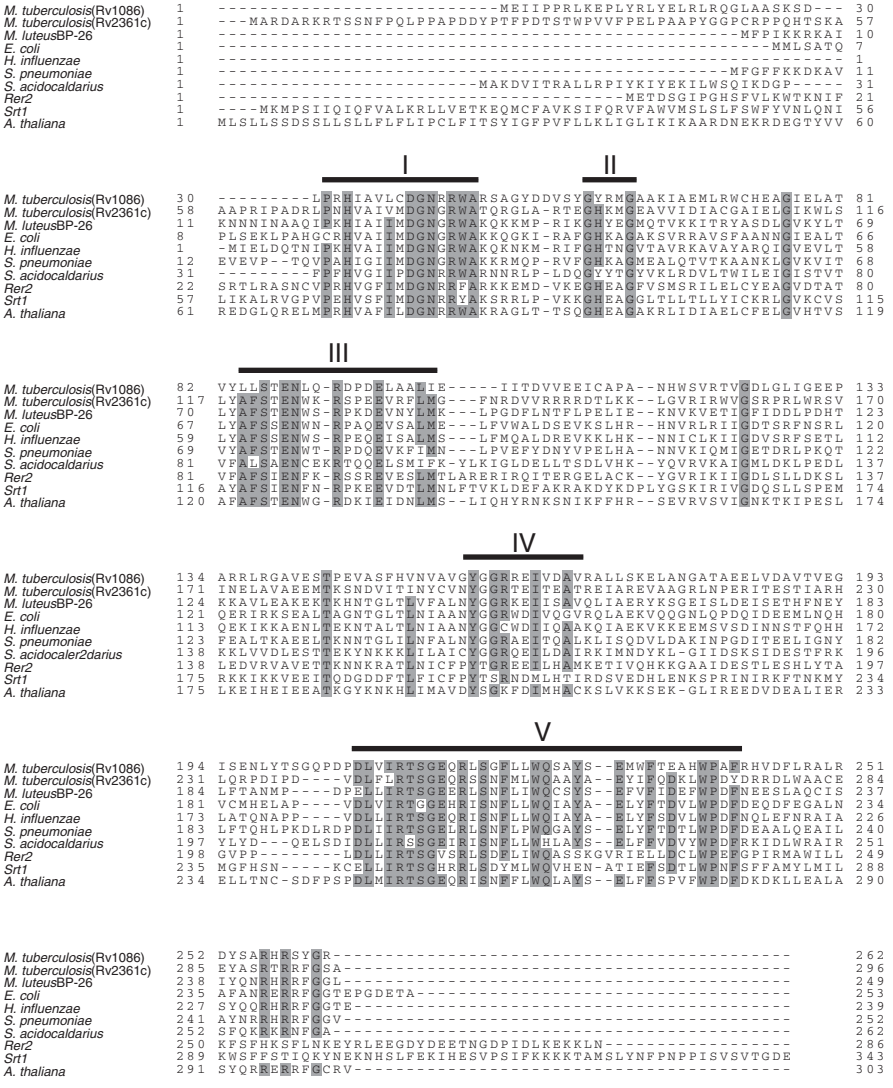


Figure 10 Multialignment of the amino acid sequences of 10 proteins that have been identified as *cis*-prenyltransferases. In the figure, five conserved regions (I–V) are indicated at the top and identical amino acid groups are shaded. Sequences: *Mycobacterium tuberculosis* Rv1086,[116] *M. tuberculosis* Rv2361c,[116] *Micrococcus luteus* B-P 26,[20] *Escherichia coli*,[19] *Haemophilus influenzae*,[19] *Streptococcus pneumoniae*,[19] *Sulfolobus acidocaldarious*,[99] Rer2,[103] Srt1,[134] and *Arabidopsis thaliana*.[125,126]

two enzyme families (**Figure 2**). The absence of DDxxD motifs in any of the conserved regions of (*Z*)-prenyltransferases led us to investigate the similarities and differences in the 3D structures, catalytic mechanisms, and chain-length determination for *cis*-prenyl chain-elongating enzymes.

In contrast to (*E*)-prenyltransferases, which were cloned a decade before (*Z*)-prenyltransferases, the detailed molecular mechanism of (*Z*)-prenyltransferases has not yet been well understood. From the random mutagenesis analysis of *M. luteus* B-P 26 UPP synthase, Fujikura *et al.*[135] suggested that the Asn-Trp/Phe (NW/F) motif plays an important role in the enzymatic catalysis and substrate binding for *cis*-prenyl chain elongation. Asn77 and Trp78 in region III were the highly conserved residues among the (*Z*)-prenyltransferases. Substitutions of Asn77 with Ala, Asp, or Gln resulted in a 10^2–10^3-fold decrease in the k_{cat} values than that of wild type, whereas the mutation did not affect K_m values for IPP and allylic substrates, indicating the significance of Asn77 in the catalytic activity. Replacement of Trp78 with charged amino acid residues leads to a two- to eightfold decrease in the k_{cat} value than that of wild type and a remarkable decrease in K_m for FPP. These results suggest that the aromatic residues of Trp78 form a hydrophobic interaction with the prenyl chain of FPP. This NW/F motif is reminiscent of the Phe-Gln (FQ) motif in region VI of (*E*)-prenyl chain-elongating enzymes, which has been shown to be essential not only for binding of allylic substrate but also for catalytic function.[83] From the viewpoints of the prenyltransferase reactions, the (*Z*)- and (*E*)-prenyltransferases are similar in that they both catalyze sequential condensations between IPP and allylic diphosphates with the concomitant release of inorganic pyrophosphate in the presence of Mg^{2+} ion. These NW/F or FQ motif consisting of carbamoyl and aromatic residues may be common and essential parts of the prenyl chain-elongating machinery that triggers the consecutive condensation of IPP to produce polyprenyl chains stereospecifically. On the other hand, Pan *et al.*[136] performed site-directed mutagenesis of all conserved acidic residues, Asp and Glu to Ala in the five conserved regions of *E. coli* UPP synthase because no typical aspartate-rich motif DDXXD is found in (*Z*)-prenyltransferases. The comprehensive site-directed mutagenesis of the conserved Asp and Glu on *E. coli* UPP synthase revealed the significance of Asp26 in region I for catalytic activity, which corresponds to Asp29 of *M. luteus* B-P 26 UPP synthase. Furthermore, the conserved Asp150 in region IV and Glu213 in region V are essential for IPP binding as well as catalysis. These acidic residues, however, seem to be not so important for the allylic substrate binding in (*Z*)-prenyltransferases as the DDxxD motifs found in (*E*)-prenyltransferases. (*E*)-Prenyltransferases indirectly bind with the allylic substrates through Mg^{2+} ion. Although Mg^{2+} ions are essential for the catalytic functions by all prenyltransferase, the optimum concentration of the divalent metal ion for (*Z*)-prenyltransferases is relatively lower than that for (*E*)-prenyltransferases, and there are no similar conserved motifs between (*Z*)- and (*E*)-prenyltransferase. It seems that (*Z*)-type enzymes bind with allylic substrate in a distinctive manner such as direct binding with the conserved basic residues. Region V consists of five highly conserved negatively charged residues and two completely conserved positively charged residues to form the largest number of amino acid residues among five conserved regions. Site-directed mutagenesis of all the charged residues in region V, which form a high electrostatic potential region, in the crystal structure of *M. luteus* B-P 26 UPP synthase, indicated that the diphosphate moiety of homoallylic substrate is electrostatically recognized by the RX_5RX_nE motif in region V. This suggested a different IPP binding mode of (*Z*)-prenyltransferases from that of (*E*)-prenyl chain-elongating enzymes.[137] Substitution of Arg197 and Arg203 with Ser and Glu216 with Gln in conserved region V resulted in a 7–11-fold increase of K_m for IPP and a 18–2000-fold decrease of k_{cat}, suggesting the critical roles of these charged residues for both enzymatic catalysis and IPP binding. These results together with the crystal structure of UPP synthase led to a hypothetical binding model for IPP to the UPP synthase being proposed.[137] The diphosphate moiety of IPP appears to be recognized directly with the side chains of Arg197 and Arg203 and indirectly with the carbonyl group of Glu21 from another monomer of the homodimer through a Mg^{2+} bridge. The Phe-Ser motif in region III, which is also highly conserved, was substituted with Ala because the Phe-Ser motif is located at the position adjacent to the high electrostatic potential region in region V according to the 3D structure of *M. luteus* B-P 26 UPP synthase.[18] Phe73 in the Phe-Ser motif may function in the holding of the hydrocarbon moiety of homoallylic substrate IPP, while Ser74 may bind to the diphosphate group of IPP with hydrogen bonding. Mutational analysis of the FS motif revealed their importance in substrate binding as well as catalytic function.[137] A critical function of the structural P-loop motif as well as Asp29 and Arg33 in the binding of pyrophosphate group of allylic substrate FPP, which is proposed by the crystal structure of *M. luteus* B-P 26 UPP synthase, was also confirmed by studies of site-directed mutagenesis.[135–137] We proposed a substrate-binding model for *M. luteus* B-P 26 UPP synthase (**Figure 11**). In this model, the hydrocarbon chain and diphosphate moiety of FPP are recognized by the hydrophobic cleft and structural P-loop motif of the

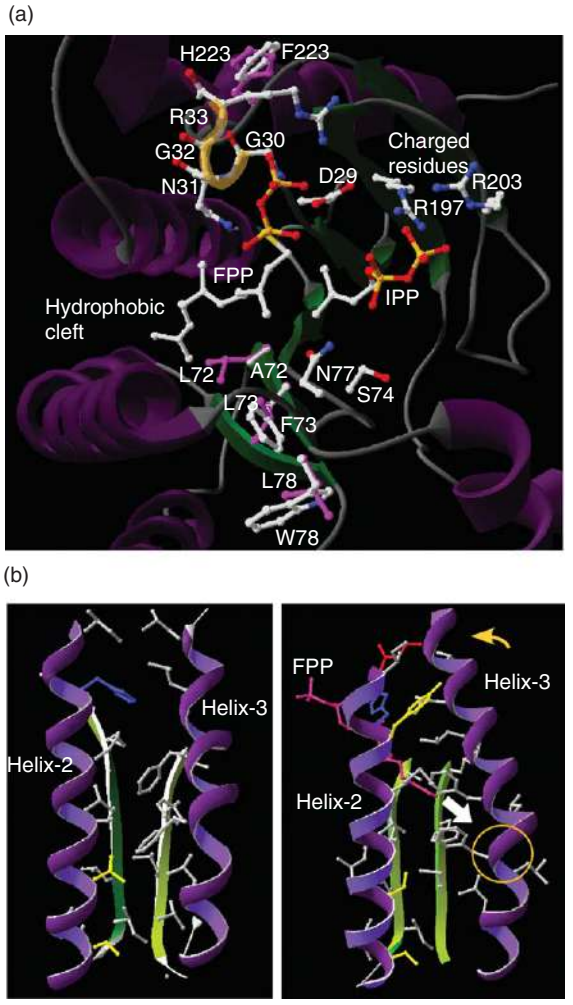


Figure 11 (a) A structural model for *Micrococcus luteus* B-P 26 UPP synthase.[138] The model was built based on the crystal structure of *E*,*E*-FPP complex of UPP synthase from *Escherichia coli*.[130] Then, structures of *E*,*E*-FPP and IPP, which were identified from the complex structure of IPP and the D26A mutant of *E. coli* UPP synthase,[130] were superimposed on the structural models. Amino acid residues mutated in the A72L/F73L/W78L/F223H mutant are indicated in red and are overlapped with residues in the wild-type UPP synthase. The structural P-loop motif, proposed to function in the binding of allylic substrates such as *E*,*E*-FPP, and charged residues including Arg197 and Arg203 are indicated. (b) Large hydrophobic cleft of *M. luteus* B-P 26 UPP synthase with or without FPP. Models were built based on the crystal structure of *E*,*E*-FPP complex of UPP synthase from *E. coli*.[130] Then, the structure of *E*,*E*-FPP was superimposed on the models. Side chains are colored as follows: nonpolar, gray; uncharged polar, yellow; acidic, red; basic, blue. Left: large hydrophobic cleft of *M. luteus* B-P 26 UPP synthase[18] consisting of helix-2, helix-3, sheet-2, and sheet-4. Only side chains facing toward the inside of the hydrophobic cleft are indicated. Right: structural model of the large hydrophobic cleft of *M. luteus* B-P 26 UPP synthase with *E*,*E*-FPP. Only side chains facing toward the inside of the hydrophobic cleft are indicated. The white arrow indicates the predicted direction of the prenyl chain elongation. The circle indicates regions in which the extra amino acid residues, EKE or RAKDY, were introduced. Reproduced from Y. Kharel; S. Takahashi; S. Yamashita; T. Koyama, *FEBS J.* **2006**, 273, 647–657, with permission from Wiley-Blackwell.

enzyme, respectively. Similarly, the model indicates the importance of five residues, F73, S74, R197, R203, and E216, which bind the IPP molecule in the proper direction suitable for stereospecific condensation with FPP. However, the amino acid residues involved in the removal of hydrogen from the 2-position of IPP, pro-*R* for (*E*)-prenyltransferase and pro-*S* for (*Z*)-prenyltransferase, have not yet been reported. To enhance the understanding of the entire mechanism of *cis*-prenyl chain elongation, further X-ray crystallography studies on substrate-bound (*Z*)-prenyltransferases are needed.

The crystal structure of *E. coli* UPP synthase in complex with the allylic substrate FPP has been reported.[130] Guo *et al.*[131] determined the complex structure of *E. coli* UPP synthase with Mg^{2+}, IPP, and farnesyl thiopyrophosphate, a low reactive analogue of FPP in which the bridging oxygen atom between farnesyl and diphosphate group is substituted with a sulfur atom. These complex structures are consistent with the models for the catalytic mechanism of UPP synthase, which were proposed by the studies of site-directed mutagenesis (**Figure 11**). Comparison of the crystal structures of UPP synthase with and without substrates revealed the conformational change between the open (apoenzyme and product-bound) and closed (substrate-bound) forms. The flexible loop downstream of sheet-2, which was disordered in the crystal structure of apoenzyme because of high flexibility,[18] becomes more rigid and visible in the new structures with both allylic and homoallylic substrates. The pyrophosphate group of FPP interacts with the residues in the structural P-loop motif, and the hydrocarbon tail of FPP is stabilized by the hydrophobic residues in helix-2 and helix-3, which form the hydrophobic interior of the large hydrophobic cleft. Binding of the allylic substrate results in the conformational change of *E. coli* UPP synthase from the open to the closed forms, in which helix-3 is kinked to be closer to the allylic substrate-binding domain compared with the structure of apoenzyme.[114] Hydrophobic residues in kinked helix-3 (Leu85, Leu88, and Phe89) interact with the FPP, which is considered to be critical for the regulation of UPP synthase activity. In the crystal structure of *E. coli* UPP synthase with FPP and IPP, Mg^{2+} ion is octahedrally coordinated by the diphosphate group of FPP, three water molecules, and the carboxylate residue of Asp26 (Asp29 for *M. luteus* B-P 26 UPP synthase). The crystal structure of the D26A mutant with substrates,[131] revealed that Mg^{2+} does not contact directly with the diphosphate group of FPP but is associated with the binding site for the diphosphate group of IPP. These results strongly suggest Asp26 plays a critical role in the catalytic mechanism by assisting the Mg^{2+} migration from IPP to FPP, which may initiate the condensation reaction by ionization of the pyrophosphate group of FPP. From the complex crystal structure, the H_S atom at the 2-position of IPP, which is removed in the (*Z*)-condensation, faces outside of the catalytic center. Asn74 and Ser71 (Asn77 and Ser74 for UPP synthase from *M. luteus* B-P 26) are the candidates for the basic residues function in the stereospecific removal of the proton. The site-directed mutagenesis of these residues resulted in a significant decrease in the k_{cat} values,[114,136] which indicate the significant role of these residues in the catalytic activity of UPP synthase. Taken together, the binding mode of the substrates, FPP and IPP, to (*Z*)-prenyltransferase is totally different from that of (*E*)-prenyltransferase, in which the Asp-rich motif is critical for substrate binding through the Mg^{2+} bridge.

1.14.3.5 Product Chain-Length Regulation Mechanism

Many reports have shown that the product chain length of (*Z*)-prenyltransferases varies depending on the origin of organisms. Bacterial enzyme produces C_{55} chain-length product UPP, whereas yeast enzyme synthesizes C_{75}–C_{95} products DedolPP, and mammals produce dolichols with the chain length up to C_{110}. Plants produce even longer dolichols and polyprenyl alcohols up to C_{120} and natural rubber's carbon chain length reaches up to several millions. The chain length of dolichol in mammals was different in various tissues such as liver[139] (C_{90}–C_{95}) and testis[38] (C_{85}–C_{90}). Some plants show an age dependence of product length, such as young leaves and old leaves of ginko, which produce C_{70}–C_{90} and C_{80}–C_{95} compounds, respectively.[98]

The (*Z*)-prenyltransferases identified so far were classified into three subfamilies depending on its final product chain length: (1) short-chain (C_{15}), (2) medium-chain (C_{50}–C_{55}), and (3) long-chain (C_{70}–C_{120}) (*Z*)-prenyltransferases (**Figure 2**). (*Z*,*E*)-FPP (C_{15}) synthase from *M. tuberculosis*, Rv1086, is the first enzyme identified as a short-chain (*Z*)-prenyltransferase, which catalyzes (*Z*)-condensation of one IPP with GPP (C_{10}).[116] Recently, three novel short-chain (*Z*)-prenyltransferases were cloned and characterized from *C. efficiens*, *C. glutamicum*, and *T. fusca* in our laboratory.[123] Medium-chain (*Z*)-prenyltransferases are represented by UPP synthase, which catalyzes (*Z*)-condensation of eight molecules of IPP with *E*,*E*-FPP (C_{15}). This enzyme is responsible for supplying the allylic substrate to UPP synthase for the biogenesis of UP, an indispensable glycosyl carrier lipid in bacterial cell wall biosynthesis. (*Z*,*E*)-Mixed decaprenyl diphosphate (*Z*,*E*-DecPP, C_{50}) synthase from *M. tuberculosis*, Rv2361, is also categorized in this subfamily.[116] Most DedolPP synthases in eukaryotes catalyzing synthesis of the precursors of the sugar carrier lipid dolichol during biosynthesis of glycoproteins are categorized as long-chain (*Z*)-prenyltransferases. SRT1p and RER2p from *S. cerevisiae*[127,103] and DeDolPP synthase from human[134] are included in this subfamily.

Distribution patterns of chain length are also affected by some reaction conditions *in vitro*. Pan *et al.*[136] showed that UPP synthase from *E. coli* produces longer chains if Triton X-100 is removed from the reaction mixture. The enzyme produces shorter chain lengths in excess of substrates IPP or FPP. Eukaryotic (*Z*)-prenyltransferases also have shown similar results. Matsuoka *et al.*[104] have shown that product chain lengths of DedolPP synthase in microsomal fractions of rat liver are significantly affected by detergents and phospholipids such as phosphatidylcholine and deoxycholate.

Since the primary and the 3D structure of (*Z*)-prenyltransferases are totally different from (*E*)-prenyltransferases, the molecular mechanisms to regulate the final product chain length of (*Z*)-prenyltransferases appeared to be different from that of (*E*)-prenyltransferases. Although the reaction conditions may only partly affect the product chain lengths, the diversity in the primary structure of (*Z*)-prenyltransferases, revealed by the cloning and characterization from various organisms, suggested that there are some specific regions responsible for the chain-length regulation for each subfamily of (*Z*)-prenyltransferases. Unlike the regions responsible for the catalytic mechanism of (*Z*)-prenyltransferases, which are well characterized by mutational and X-ray studies by using UPP synthase,[33,34] the regions to define the final product chain length of (*Z*)-prenyltransferase remained to be elucidated.

In order to investigate regions important for determination of product chain length in (*Z*)-prenyltransferases, Kharel *et al.*[138] explored the characteristic residues of short- or long-chain (*Z*)-prenyltransferase subfamilies based on comparisons between the primary structures of (*Z*)-prenyltransferases identified so far and the crystal structures of UPP synthases. They introduced mutations in regions of *M. luteus* B-P 26 UPP synthase, which correspond to the characteristic residues (**Figure 10**).

Short-chain-length *cis*-prenyltransferases[116,123] have conserved Leu residues at positions 84, 85, and 90 instead of the corresponding Ala, Phe, and Leu/Trp in the conserved region III of medium- and long-chain-length *cis*-prenyltransferases. The corresponding residues in *M. luteus* B-P 26 UPP synthase, that is, Ala72, Phe73, and Trp78, in the highly conserved region III are located at the edge of the large hydrophobic cleft. When *E*,*E*-FPP and GPP were used as allylic substrates, the *M. luteus* B-P 26 UPP synthase mutants, A72L/F73L (LL) and A72L/F73L/W78L (LLL), resulted in shorter (C_{25}–C_{40} as major products) and even shorter (C_{20}–C_{35} as major products) polyisoprenoids, respectively, than did wild-type UPP synthase, which produces C_{55}–C_{60} polyisoprenoids. In the crystal structure of *E. coli* UPP synthase bound with the allylic substrate *E*,*E*-FPP,[130,131] Ala69, which corresponds to Ala72 of *M. luteus* B-P 26 UPP synthase, is located close to the ω-end carbon, C-14 of *E*,*E*-FPP at a distance of 3.2 Å. This suggests that the bulky alkyl group of Leu72 in the mutants or Leu84 in Rv1086 might interfere with cis-addition of IPP onto *E*,*E*-FPP or *Z*,*E*-FPP (**Figure 11**).

In contrast, Phe73 and Trp78 of *M. luteus* B-P 26 UPP synthase, corresponding to Leu85 and Leu90 of Rv1086, respectively, are not close to the binding site for *E*,*E*-FPP but are placed on a flexible loop.[18,130,131] Substitutions of the highly conserved Phe73 and Ser74 of *M. luteus* B-P 26 UPP synthase into Ala resulted in 32- and 16-fold increases in the K_m value for IPP, and 16- and 12-fold decreases in the k_{cat} value, respectively,[137] indicating that the flexible domain in the conserved region III is important both for the binding of IPP in the proper direction and for catalytic function. The double mutant A72L/F73L also shows an 88-fold higher K_m value for IPP compared with the wild-type enzyme. However, the triple mutant A72L/F73L/W78L shows a sixfold increase in the K_m value for IPP. These results suggest that the replacement of Trp78 with Leu was necessary for the creation of a proper IPP binding domain when Ala72 and Phe73 were replaced with Leu residues, which corresponded to Leu84, Leu85, and Leu90 in Rv1086, respectively.

Furthermore, His237 in Rv1086 downstream of the conserved region V was found at the corresponding position for the Leu/Phe residue in other *cis*-prenyltransferases. Substitution of Phe223 of *M. luteus* B-P 26 UPP synthase, which corresponds to His237 in Rv1086, with His dramatically decreased the catalytic activity of UPP synthase when *E*,*E*-FPP was used as allylic substrate. However, the quadruple mutations A72L/F73L/W78L/F223H (LLLH) can accept GPP as an allylic substrate, producing shorter prenyl products, C_{45}–C_{55} and C_{15}. According to the crystal structure of UPP synthase, His237 of Rv1086 is considered located near the structural P-loop motif, which recognizes the diphosphate group of allylic substrates,[18,37] suggesting that the His237 might affect the binding affinity for the allylic substrate with the structural P-loop motif.

Multiple alignment of (*Z*)-prenyltransferases also revealed that long-chain *cis*-prenyltransferases, such as SRT1p and RER2p from *S. cerevisiae*[103,134] and the DedolPP synthase from human,[127] have three to seven extra amino acid residues downstream of the conserved region III (**Figure 10**). This position corresponds to helix-3

of the *M. luteus* B-P 26 UPP synthase. Helix-3 participates in the creation of the hydrophobic cleft, which is considered to accommodate the elongated prenyl intermediates.[18] The mutant of *M. luteus* B-P 26 UPP synthase, EKE, which contains insertion corresponding exactly to the extra amino acid residues of DedolPP synthases from human (HDS, positions 107–109), produce relatively longer prenyl products with carbon chain lengths of C_{55}–C_{70}. Moreover, the mutant RAKDY, which contains insertion corresponding exactly to the extra amino acid residues of DedolPP synthases from yeast (SRT1p, positions 148–152), gives even longer prenyl products with chain lengths of C_{60}–C_{75}. Insertion of five Ala residues instead of the peptide RAKDY does not cause a significant change in the length of prenyl products, suggesting the significance of some charged or polar residues at the proper positions rather than expansion of the interior space of the hydrophobic cleft. As mentioned earlier, in the crystal structure of *E. coli* UPP synthase bound with *E,E*-FPP, helix-3 is kinked to be closer to the *E,E*-FPP binding domain compared with the structure without substrates.[130,131] The interior of the large hydrophobic cleft, surrounded by sheet-2, sheet-4, helix-2, and helix-3, consists mostly of hydrophobic residues. Hydrophobic residues on kinked helix-3 in the closed conformation may function as a guide rail to introduce the elongation of the hydrophobic prenyl chain in the proper direction (**Figure 11(b)**). Moreover, the domain in which we introduced extra amino acid residues corresponded to the hinge region of the kinked helix-3.

Chang *et al.*[130] reported the crystal structure of UPP synthase from *E. coli* complexed with Triton X-100. Interestingly, in this structure, two molecules of Triton X-100 are bound in the catalytic center and the large hydrophobic cleft, folding with a suitable torsion angle for accommodation in the hydrophobic cleft. They supposed that the conformation of Triton X-100 molecules, which interact with the hydrophobic residues in the cleft, may mimic that of the hydrophobic moiety of UPP, and proposed the conformation of UPP folded in the cleft. In the structure, the polyprenyl chain is also folded back at the position corresponding to the hinge region of the kinked helix-3. Therefore, it is postulated that charged residues inserted at the hinge region of helix-3 might control the bending direction of the growing hydrophobic prenyl chain along the hydrophobic interior of helix-3 so that the hydrophobic cleft could accommodate the bulk of the prenyl chain to fit a suitable size during enzymatic elongation. The fact that residues localized on helix-3 show a wide diversity while most residues constituting the hydrophobic cleft are highly conserved among *cis*-prenyltransferases also supports the hypothesis.

Substitution of Leu137, which is located at the bottom of the hydrophobic cleft of *E. coli* UPP synthase, with Ala, resulted in one-unit elongation of the chain length of the final product of (*Z,E*)-mixed polyisoprenoids, producing C_{55} and C_{60} as major products in the presence of 0.1% Triton X-100 and C_{70} and C_{75} in the absence of Triton X-100.[114] It was proposed that the bulky Leu137 functioned as the floor of the tunnel to block further elongation of polyprenyl products. This proposed model seems to be similar to the chain-length determination mechanism of (*E*)-prenyltransferases, in which bulky residues located upstream of FARM play an important role in determining the final prenyl chain length, by forming a suitable size for the pocket for elongation of the isoprenoid chain. According to the model, we constructed and analyzed a mutant of *M. luteus* B-P 26 UPP synthase L140A, which corresponded to the *E. coli* UPP synthase mutant L137A, with the result that the mutant also produced C_{55} and C_{60} as major products in the presence of Triton X-100. Elongation of the ultimate product to C_{70}–C_{75} in the reaction of the mutant L137A without Triton X-100 seems to depend on the reaction conditions *in vitro*.

1.14.4 Conclusions

With the molecular cloning of UPP synthase from *M. luteus* B-P 26 in 1998 as a turning point, the study of prenyl chain-elongation enzymes turned to a new direction. Detailed analyses of the relationships between protein structure and catalytic mechanism of a new family of isoprenoid biosynthetic enzymes, *cis*-prenyltransferase, and comparative studies of both types of prenyl chain-elongation enzymes are of great interest as aspects of molecular evolution. In addition, the chain-length determination mechanism of *cis*-prenyltransferase is quite important because biological materials composed by the polymer of IPP with cis-configuration, such as natural rubber, can be used for the development of novel functional materials. Recently, our group reported the molecular cloning of two cDNAs, HRT1 and HRT2, encoding

cis-prenyltransferases from the latex of *H. brasiliensis*, which is almost the sole source for commercial natural rubber production.[128] The predominant expression of these genes in latex suggests a specialized function of HRTs in latex, that is, cis-1,4-polymerization for the carbon backbone of natural rubber. *In vitro* prenyltransferase assay using recombinant HRT protein revealed the catalytic activity for the production of polyisoprenoids with a carbon chain length of C_{100}. However, in the presence of a latex fraction, HRT shows high prenyltransferase activity, producing polyisoprenoids with higher molecular size, which corresponds to natural rubber. These results suggest the existence of mechanism for functional conversion in the chain-length determination of *cis*-prenyltransferase. Further investigations might provide a general role for the regulation of the carbon chain length of products from *cis*-prenyltransferase, and may enable us to manipulate the chain-length determination mechanism of the enzyme so as to synthesize polyisoprenoids with various chain lengths.

Glossary

FARM The first aspartate-rich motif, DDXXD, in (*E*)-prenyltransferases.

References

1. T. Koyama, *Biosci. Biotechnol. Biochem.* **1999**, *63*, 1671.
2. K. Ogura; T. Koyama, *Chem. Rev.* **1998**, *98*, 1263.
3. S. Takahashi; T. Koyama, *Chem. Rec.* **2006**, *6*, 194.
4. E. D. Beytia; J. W. Porter, *Annu. Rev. Biochem.* **1976**, *45*, 113.
5. M. Rohmer, *Nat. Prod. Rep.* **1999**, *16*, 565.
6. M. Rohmer; M. Knani; P. Simonin; B. Sutter; H. Sahm, *Biochem. J.* **1993**, *295* (Pt. 2), 517.
7. M. Rohmer; M. Seemann; S. Horbach; S. Bringer-Meyer; H. Sahm, *J. Am. Chem. Soc.* **1996**, *118*, 2564.
8. S. David; B. Estramareix; J. C. Fischer; M. Therisod, *J. Am. Chem. Soc.* **1981**, *103*, 7341.
9. J. H. Julliard; R. Douce, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2042.
10. K. Himmeldirk; B. G. Sayer; I. D. Spenser, *J. Am. Chem. Soc.* **1998**, *120*, 3581.
11. L. Charon; J. F. Hoeffler; C. Pale-Grosdemange; L. M. Lois; N. Campos; A. Boronat; M. Rohmer, *Biochem. J.* **2000**, *346* (Pt. 3), 737.
12. T. Kuzuyama, *Biosci. Biotechnol. Biochem.* **2002**, *66*, 1619.
13. W. Eisenreich; A. Bacher; D. Arigoni; F. Rohdich, *Cell. Mol. Life Sci.* **2004**, *61*, 1401.
14. W. Eisenreich; B. Menhard; P. J. Hylands; M. H. Zenk; A. Bacher, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6431.
15. J. Schwender; M. Seemann; H. K. Lichtenthaler; M. Rohmer, *Biochem. J.* **1996**, *316* (Pt. 1), 73.
16. M. Ito; M. Kobayashi; T. Koyama; K. Ogura, *Biochemistry* **1987**, *26*, 4745.
17. M. Kobayashi; M. Ito; T. Koyama; K. Ogura, *J. Am. Chem. Soc.* **1985**, *107*, 4588.
18. M. Fujihashi; Y. W. Zhang; Y. Higuchi; X. Y. Li; T. Koyama; K. Miki, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4337.
19. C. M. Apfel; B. Takacs; M. Fountoulakis; M. Stieger; W. Keck, *J. Bacteriol.* **1999**, *181*, 483.
20. N. Shimizu; T. Koyama; K. Ogura, *J. Biol. Chem.* **1998**, *273*, 19476.
21. L. Heide; U. Berger, *Arch. Biochem. Biophys.* **1989**, *273*, 331.
22. N. L. Eberhardt; H. C. Rilling, *J. Biol. Chem.* **1975**, *250*, 863.
23. A. Joly; P. A. Edwards, *J. Biol. Chem.* **1993**, *268*, 26983.
24. K. Asai; S. Fujisaki; Y. Nishimura; T. Nishino; K. Okada; T. Nakagawa; M. Kawamukai; H. Matsuda, *Biochem. Biophys. Res. Commun.* **1994**, *202*, 340.
25. T. Koyama; S. Obata; M. Osabe; A. Takeshita; K. Yokoyama; M. Uchida; T. Nishino; K. Ogura, *J. Biochem.* **1993**, *113*, 355.
26. S. Ohnuma; M. Suzuki; T. Nishino, *J. Biol. Chem.* **1994**, *269*, 14792.
27. A. Chen; C. D. Poulter, *Arch. Biochem. Biophys.* **1994**, *314*, 399.
28. H. Sagami; Y. Morita; K. Ogura, *J. Biol. Chem.* **1994**, *269*, 20561.
29. L. Blanchard; F. Karst, *Gene* **1993**, *125*, 185.
30. A. Tachibana, *FEBS Lett.* **1994**, *341*, 291.
31. H. Fujii; T. Koyama; K. Ogura, *J. Biol. Chem.* **1982**, *257*, 14610.
32. I. Yoshida; T. Koyama; K. Ogura, *Biochim. Biophys. Acta* **1989**, *995*, 138.
33. Y. Kharel; T. Koyama, *Nat. Prod. Rep.* **2003**, *20*, 111.
34. P. H. Liang; T. P. Ko; A. H. Wang, *Eur. J. Biochem.* **2002**, *269*, 3339.
35. A. Koike-Takeshita; T. Koyama; K. Ogura, *J. Biochem.* **1998**, *124*, 790.
36. I. Takahashi; K. Ogura; S. Seto, *J. Biol. Chem.* **1980**, *255*, 4539.
37. K. Fujikura; Y. W. Zhang; M. Fujihashi; K. Miki; T. Koyama, *Biochemistry* **2003**, *42*, 4035.
38. K. Yamada; S. Abe; T. Suzuki; K. Katayama; T. Sato, *Anal. Biochem.* **1986**, *156*, 380.
39. S. Ohnuma; T. Koyama; K. Ogura, *J. Biol. Chem.* **1991**, *266*, 23706.
40. K. Ishii; H. Sagami; K. Ogura, *Biochim. Biophys. Acta* **1985**, *835*, 291.
41. S. Takahashi; T. Nishino; T. Koyama, *Biochem. Eng. J.* **2003**, *16*, 183.
42. Y. Ye; M. Fujii; A. Hirata; M. Kawamukai; C. Shimoda; T. Nakamura, *Mol. Biol. Cell* **2007**, *18*, 3568.

43. R. Croteau; P. T. Purkett, *Arch. Biochem. Biophys.* **1989**, *271*, 524.
44. F. Bouvier; C. Suire; A. d'Harlingue; R. A. Backhaus; B. Camara, *Plant J.* **2000**, *24*, 241.
45. C. Burke; R. Croteau, *Arch. Biochem. Biophys.* **2002**, *405*, 130.
46. T. Endo; T. Suga, *Phytochemistry* **1992**, *31*, 2273.
47. C. C. Burke; M. R. Wildung; R. Croteau, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13062.
48. A. B. Gilg; J. C. Bearfield; C. Tittiger; W. H. Welch; G. J. Blomquist, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 9760.
49. L. Jeffries; M. A. Cawthorne; M. Harris; A. T. Diplock; J. Green; S. A. Price, *Nature* **1967**, *215*, 257.
50. H. Fujii; T. Koyama; K. Ogura, *FEBS Lett.* **1983**, *161*, 257.
51. I. Yoshida; T. Koyama; K. Ogura, *Biochemistry* **1987**, *26*, 6840.
52. I. Yoshida; T. Koyama; K. Ogura, *Biochem. Biophys. Res. Commun.* **1989**, *160*, 448.
53. H. Sagami; K. Ogura; S. Seto, *Biochemistry* **1977**, *16*, 4616.
54. S. Fujisaki; T. Nishino; H. Katsuki, *J. Biochem.* **1986**, *99*, 1327.
55. J. W. Cornforth; R. H. Cornforth; C. Donninger; G. Popjak, *Proc. R. Soc. Lond. B Biol. Sci.* **1966**, *163*, 492.
56. J. W. Cornforth; R. H. Cornforth; G. Popjak; L. Yengoyan, *J. Biol. Chem.* **1966**, *241*, 3970.
57. D. J. Hosfield; Y. Zhang; D. R. Dougan; A. Broun; L. W. Tari; R. V. Swanson; J. Finn, *J. Biol. Chem.* **2004**, *279*, 8526.
58. C. F. Clarke; R. D. Tanaka; K. Svenson; M. Wamsley; A. M. Fogelman; P. A. Edwards, *Mol. Cell. Biol.* **1987**, *7*, 3138.
59. B. T. Sheares; S. S. White; D. T. Molowa; K. Chan; V. D. Ding; P. A. Kroon; R. G. Bostedor; J. D. Karkas, *Biochemistry* **1989**, *28*, 8129.
60. D. J. Wilkin; S. Y. Kutsunai; P. A. Edwards, *J. Biol. Chem.* **1990**, *265*, 4607.
61. M. S. Anderson; J. G. Yarger; C. L. Burck; C. D. Poulter, *J. Biol. Chem.* **1989**, *264*, 19176.
62. S. Fujisaki; H. Hara; Y. Nishimura; K. Horiuchi; T. Nishino, *J. Biochem.* **1990**, *108*, 995.
63. S. Attucci; S. M. Aitken; P. J. Gulick; R. K. Ibrahim, *Arch. Biochem. Biophys.* **1995**, *321*, 493.
64. N. Cunillera; M. Arro; D. Delourme; F. Karst; A. Boronat; A. Ferrer, *J. Biol. Chem.* **1996**, *271*, 7774.
65. K. Adiwilaga; A. Kush, *Plant Mol. Biol.* **1996**, *30*, 935.
66. B. C. Reed; H. C. Rilling, *Biochemistry* **1975**, *14*, 50.
67. H. Nakane; T. Koyama; S. Obata; M. Osabe; A. Takeshita; T. Nishino; K. Ogura; K. Miki, *J. Mol. Biol.* **1993**, *233*, 787.
68. L. C. Tarshis; M. Yan; C. D. Poulter; J. C. Sacchettini, *Biochemistry* **1994**, *33*, 10871.
69. W. L. DeLano, *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, 2002.
70. C. A. Lesburg; G. Zhai; D. E. Cane; D. W. Christianson, *Science* **1997**, *277*, 1820.
71. K. U. Wendt; K. Poralla; G. E. Schulz, *Science* **1997**, *277*, 1811.
72. C. M. Starks; K. Back; J. Chappell; J. P. Noel, *Science* **1997**, *277*, 1815.
73. H. W. Park; S. R. Boduluri; J. F. Moomaw; P. J. Casey; L. S. Beese, *Science* **1997**, *275*, 1800.
74. J. C. Sacchettini; C. D. Poulter, *Science* **1997**, *277*, 1788.
75. T. H. Chang; R. T. Guo; T. P. Ko; A. H. Wang; P. H. Liang, *J. Biol. Chem.* **2006**, *281*, 14991.
76. K. L. Kavanagh; J. E. Dunford; G. Bunkoczi; R. G. Russell; U. Oppermann, *J. Biol. Chem.* **2006**, *281*, 22004.
77. N. Shimizu; T. Koyama; K. Ogura, *J. Bacteriol.* **1998**, *180*, 1578.
78. A. Koike-Takeshita; T. Koyama; S. Obata; K. Ogura, *J. Biol. Chem.* **1995**, *270*, 18396.
79. Y. W. Zhang; T. Koyama; D. M. Marecak; G. D. Prestwich; Y. Maki; K. Ogura, *Biochemistry* **1998**, *37*, 13411.
80. L. Song; C. D. Poulter, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3044.
81. T. Koyama; M. Tajima; H. Sano; T. Doi; A. Koike-Takeshita; S. Obata; T. Nishino; K. Ogura, *Biochemistry* **1996**, *35*, 9533.
82. P. F. Marrero; C. D. Poulter; P. A. Edwards, *J. Biol. Chem.* **1992**, *267*, 21873.
83. T. Koyama; M. Tajima; T. Nishino; K. Ogura, *Biochem. Biophys. Res. Commun.* **1995**, *212*, 681.
84. D. A. Dougherty, *Science* **1996**, *271*, 163.
85. T. Koyama; K. Saito; K. Ogura; S. Obata; A. Takeshita, *Can. J. Chem.* **1994**, *72*, 75.
86. T. Koyama; S. Obata; K. Saito; A. Takeshita-Koike; K. Ogura, *Biochemistry* **1994**, *33*, 12644.
87. R. M. Phan; C. D. Poulter, *J. Org. Chem.* **2001**, *66*, 6705.
88. N. Misawa; M. Nakagawa; K. Kobayashi; S. Yamano; Y. Izawa; K. Nakamura; K. Harashima, *J. Bacteriol.* **1990**, *172*, 6704.
89. S. Ohnuma; T. Nakazawa; H. Hemmi; A. M. Hallberg; T. Koyama; K. Ogura; T. Nishino, *J. Biol. Chem.* **1996**, *271*, 10087.
90. S. Ohnuma; K. Narita; T. Nakazawa; C. Ishida; Y. Takeuchi; C. Ohto; T. Nishino, *J. Biol. Chem.* **1996**, *271*, 30748.
91. L. C. Tarshis; P. J. Proteau; B. A. Kellogg; J. C. Sacchettini; C. D. Poulter, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 15018.
92. Y. W. Zhang; X. Y. Li; T. Koyama, *Biochemistry* **2000**, *39*, 12717.
93. K. Hirooka; S. Ohnuma; A. Koike-Takeshita; T. Koyama; T. Nishino, *Eur. J. Biochem.* **2000**, *267*, 4520.
94. S. Ohnuma; K. Hirooka; H. Hemmi; C. Ishida; C. Ohto; T. Nishino, *J. Biol. Chem.* **1996**, *271*, 18831.
95. S. Ohnuma; K. Hirooka; C. Ohto; T. Nishino, *J. Biol. Chem.* **1997**, *272*, 5192.
96. M. Noike; T. Katagiri; T. Nakayama; T. Koyama; T. Nishino; H. Hemmi, *FEBS J.* **2008**, *275*, 3921.
97. J. D. Muth; C. M. Allen, *Arch. Biochem. Biophys.* **1984**, *230*, 49.
98. S. Tateyama; R. Wititsuwannakul; D. Wititsuwannakul; H. Sagami; K. Ogura, *Phytochemistry* **1999**, *51*, 11.
99. H. Hemmi; S. Yamashita; T. Shimoyama; T. Nakayama; T. Nishino, *J. Bacteriol.* **2001**, *183*, 401.
100. W. L. Adair, Jr.; N. Cafmeyer; R. K. Keller, *J. Biol. Chem.* **1984**, *259*, 4441.
101. Z. Chen; C. Morris; C. M. Allen, *Arch. Biochem. Biophys.* **1988**, *266*, 98.
102. Y. E. Bukhtiyarov; Y. A. Shabalin; I. S. Kulaev, *J. Biochem.* **1993**, *113*, 721.
103. M. Sato; K. Sato; S. Nishikawa; A. Hirata; J. Kato; A. Nakano, *Mol. Cell. Biol.* **1999**, *19*, 471.
104. S. Matsuoka; H. Sagami; A. Kurisaki; K. Ogura, *J. Biol. Chem.* **1991**, *266*, 3464.
105. J. G. Christenson; S. K. Gross; P. W. Robbins, *J. Biol. Chem.* **1969**, *244*, 5436.
106. T. Kurokawa; K. Ogura; S. Seto, *Biochem. Biophys. Res. Commun.* **1971**, *45*, 251.
107. T. Baba; C. M. Allen, *Arch. Biochem. Biophys.* **1980**, *200*, 474.
108. I. Takahashi; K. Ogura, *J. Biochem.* **1982**, *92*, 1527.
109. M. V. Keenan; C. M. Allen, Jr., *Arch. Biochem. Biophys.* **1974**, *161*, 375.
110. C. M. Allen; M. V. Keenan; J. Sack, *Arch. Biochem. Biophys.* **1976**, *175*, 236.

111. T. Baba; C. M. Allen, Jr., *Biochemistry* **1978**, *17*, 5598.
112. K. Ishii; H. Sagami; K. Ogura, *Biochem. J.* **1986**, *233*, 773.
113. T. Koyama; T. Kokubun; K. Ogura, *Phytochemistry* **1988**, *27*, 2005.
114. T. P. Ko; Y. K. Chen; H. Robinson; P. C. Tsai; Y. G. Gao; A. P. Chen; A. H. Wang; P. H. Liang, *J. Biol. Chem.* **2001**, *276*, 47474.
115. W. Wang; C. Dong; M. McNeil; D. Kaur; S. Mahapatra; D. C. Crick; J. H. Naismith, *J. Mol. Biol.* **2008**, *381*, 129.
116. M. C. Schulbach; P. J. Brennan; D. C. Crick, *J. Biol. Chem.* **2000**, *275*, 22876.
117. E. Skoczylas; E. Swiezewska; T. Chojnacki; Y. Tanaka, *Plant Physiol. Biochem.* **1994**, *32*, 825.
118. V. Kulcitsky; J. Hertel; E. Skoczylas; E. Swiezewska; T. Chojnacki, *Acta Biochim. Pol.* **1996**, *43*, 707.
119. M. Wanke; T. Chojnacki; E. Swiezewska, *Acta. Biochim. Pol.* **1998**, *45*, 811.
120. G. J. Quellhorst, Jr.; J. S. Piotrowski; S. E. Steffen; S. S. Krag, *Biochem. Biophys. Res. Commun.* **1998**, *244*, 546.
121. J. W. Rip; C. A. Rupar; K. Ravi; K. K. Carroll, *Prog. Lipid Res.* **1985**, *24*, 269.
122. C. R. Raetz, *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 2274.
123. T. Ambo; M. Noike; H. Kurokawa; T. Koyama, *Biochem. Biophys. Res. Commun.* **2008**, *375*, 536.
124. J. Kato; S. Fujisaki; K. Nakajima; Y. Nishimura; M. Sato; A. Nakano, *J. Bacteriol.* **1999**, *181*, 2733.
125. S. K. Oh; K. H. Han; S. B. Ryu; H. Kang, *J. Biol. Chem.* **2000**, *275*, 18482.
126. N. Cunillera; M. Arro; O. Fores; D. Manzano; A. Ferrer, *FEBS Lett.* **2000**, *477*, 170.
127. S. Endo; Y. W. Zhang; S. Takahashi; T. Koyama, *Biochim. Biophys. Acta* **2003**, *1625*, 291.
128. K. Asawatreratanakul; Y. W. Zhang; D. Wititsuwannakul; R. Wititsuwannakul; S. Takahashi; A. Rattanapittayaporn; T. Koyama, *Eur. J. Biochem.* **2003**, *270*, 4671.
129. I. Takahashi; K. Ogura, *J. Biochem.* **1981**, *89*, 1581.
130. S. Y. Chang; T. P. Ko; A. P. Chen; A. H. Wang; P. H. Liang, *Protein Sci.* **2004**, *13*, 971.
131. R. T. Guo; T. P. Ko; A. P. Chen; C. J. Kuo; A. H. Wang; P. H. Liang, *J. Biol. Chem.* **2005**, *280*, 20762.
132. C. J. Kuo; R. T. Guo; I. L. Lu; H. G. Liu; S. Y. Wu; T. P. Ko; A. H. Wang; P. H. Liang, *J. Biomed. Biotechnol.* **2008**, *2008*, 841312.
133. K. Kinoshita; K. Sadanami; A. Kidera; N. Go, *Protein Eng.* **1999**, *12*, 11.
134. M. Sato; S. Fujisaki; K. Sato; Y. Nishimura; A. Nakano, *Genes Cells* **2001**, *6*, 495.
135. K. Fujikura; Y. W. Zhang; H. Yoshizaki; T. Nishino; T. Koyama, *J. Biochem.* **2000**, *128*, 917.
136. J. J. Pan; L. W. Yang; P. H. Liang, *Biochemistry* **2000**, *39*, 13856.
137. Y. Kharel; Y. W. Zhang; M. Fujihashi; K. Miki; T. Koyama, *J. Biol. Chem.* **2001**, *276*, 28459.
138. Y. Kharel; S. Takahashi; S. Yamashita; T. Koyama, *FEBS J.* **2006**, *273*, 647.
139. K. Yamada; H. Yokohama; S. Abe; K. Katayama; T. Sato, *Anal. Biochem.* **1985**, *150*, 26.

Biographical Sketches

Hirofumi Kurokawa was born in Ehime Prefecture, Japan, in 1968. He received his B.S. (1991), M.S. (1993), and Ph.D. (1996) degrees from Kyoto University. After undertaking postdoctoral work in the laboratories of Professor Mitsu Ikura at the University of Toronto and Professor Masatsune Kainosho at Tokyo Metropolitan University, he became an assistant professor at Tohoku University in 2002. He was promoted to associate professor at the Tohoku University Graduate School of medicine in 2009. His current research interest is the structural biology of prenyltransferases.

Tanetoshi Koyama was born in 1945 in Shizuoka, Japan, and graduated from Tohoku University in 1968, where he obtained his doctorate in 1973. After postdoctoral work at Stanford University with E. E. van Tamelen, he joined the Chemical Research Institute of Non-Aqueous Solutions, Tohoku University as a research associate in 1974. He was promoted to associate professor at the Department of Biochemistry and Engineering, Faculty of Engineering, Tohoku University in 1994, to professor at the Institute for Chemical Reaction Science in 1997, and then to professor of the Institute of Multidisciplinary Research for Advanced Materials, Tohoku University in 2001. His interest focuses on molecular mechanisms of enzyme catalysis involved in the biosynthesis of isoprenoid compounds.

1.15 Advances in the Enzymology of Monoterpene Cyclization Reactions

Edward M. Davis, Washington State University, Pullman, WA, USA

© 2010 Elsevier Ltd. All rights reserved.

1.15.1	**Introduction**	585
1.15.2	**Cyclization Chemistry**	588
1.15.2.1	Monoterpene Cyclases and Related Catalysts	588
1.15.2.2	Monoterpene Cyclase Enzymology	588
1.15.2.3	Ionization and Isomerization of Geranyl Diphosphate	590
1.15.2.4	Cyclization Reaction Stereochemistry	591
1.15.2.5	Formation of Monoterpene Structural Types	593
1.15.2.6	Molecular Biology of the Monoterpene Cyclases	594
1.15.3	**Structural Studies**	595
1.15.3.1	Structural Commonalities	595
1.15.3.2	(−)-(4*S*)-Limonene Synthase	598
1.15.3.3	(+)-Bornyl Diphosphate Synthase	600
1.15.3.4	1,8-Cineole Synthase	603
1.15.4	**Conclusions**	605
References		606

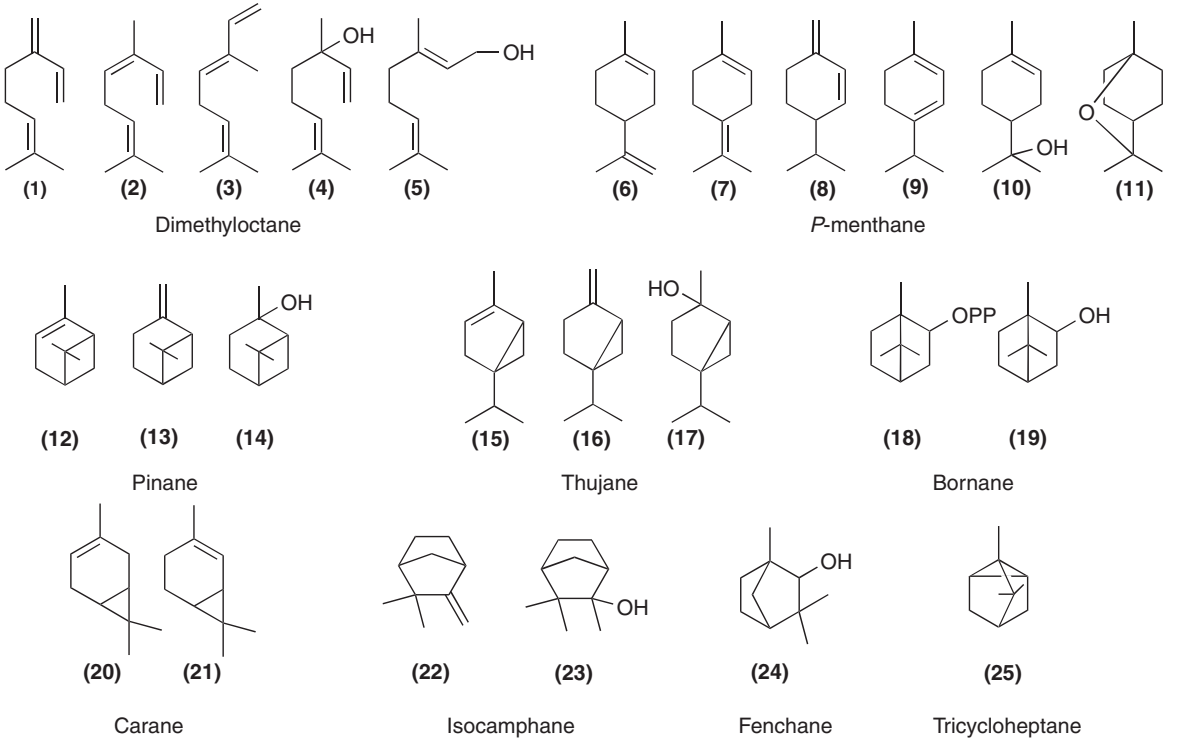
1.15.1 Introduction

Monoterpenes are widely occurring, likely ubiquitous natural products found in the plant kingdom and are the primary contributors to the organoleptic properties associated with various herbs, spices, citrus, conifers, and most flowers and fruits. These 10-carbon, short-chain terpenes, often in combination with sesquiterpenes (C15) and diterpenes (C20), exhibit many ecological roles, including broad-spectrum antimicrobial, allelopathic, herbivore deterrent (indirect and direct), and pollinator attractant properties.[1–4] The diversity of monoterpenes generated from the cyclization of the predominant monoterpene precursor, geranyl diphosphate (GPP (**28**)), or its *cis*-isomer, neryl dliphosphate (NPP), results in several general monoterpene types (**Scheme 1**), with greater multiplicity achieved by the cyclase through stereochemical and regiochemical alterations. Downstream enzymatic transformations of these initial manifolds through hydroxylations, dehydrogenations, double-bond and/or carbonyl reductions, isomerizations, and conjugations can lead to increased biodiversity as exemplified in the mint family (**Scheme 2**).[5–7] The primary subject of this review is the enzyme-mediated generation of the initial monoterpene scaffolds by monoterpene cyclases (or synthases).

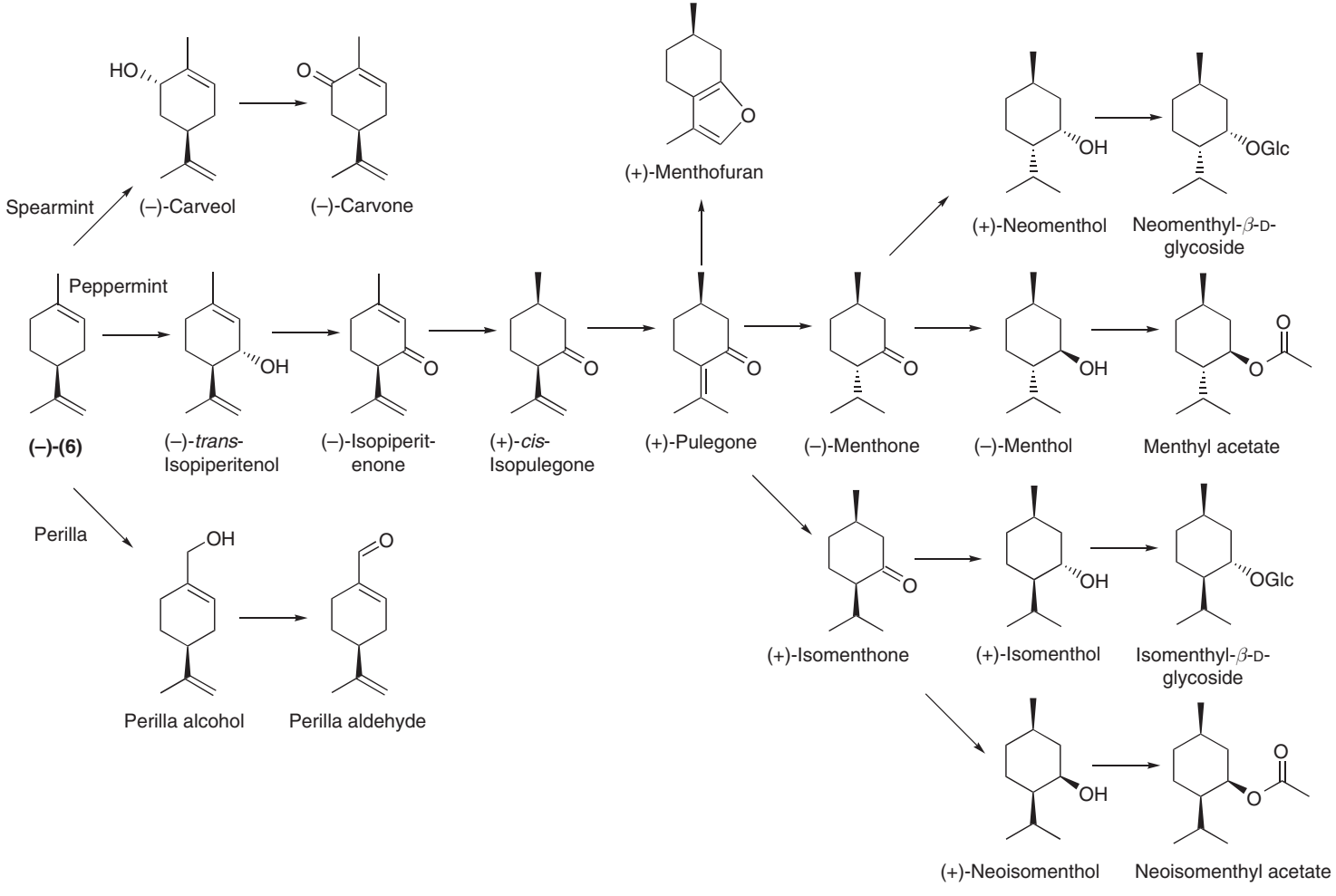
Comprehensive and historical reviews of monoterpene biosynthesis were published in 1987[8] and in this series in 1999;[9] the present contribution will provide an overview of the mechanistic and structural advances that have been made since that time including recent enzyme inhibitor studies, structural studies, and cyclase mutagenesis in the context of a more comprehensive monoterpene cyclization model. While highlighting the recent advances made in the past decade, I would be remiss if I did not acknowledge and build upon the seminal findings made by the terpene giants of the past and present, from Kekulé, Wallach, and Rüzicka to Croteau, Coates, and Cane.

Because many aspects of monoterpene chemistry were revealed by previous work, and thoroughly reviewed in a variety of contexts,[8–12] a judicious appraisal of these principles will supplement the latest findings in the framework of our current state of understanding. This review is divided into two main sections. The first section is an overview of monoterpene cyclization chemistry including the general properties of cyclases and isomerization, cyclization, and carbocation rearrangements that precede reaction termination. A brief consideration of the molecular biology of monoterpene cyclases will provide a segue to the second section, which will consider the mechanistic specificity of three monoterpene cyclases that catalyze ostensibly different reaction types, and provide the structural basis for each enzyme.

(1) Myrcene
(2) *Z*-β-ocimene
(3) *E*-β-ocimene
(4) Linalool
(5) Geraniol
(6) Limonene
(7) Terpinolene
(8) β-phellandrene
(9) α-terpinene
(10) α-terpineol
(11) 1,8-cineole
(12) α-pinene
(13) β-pinene
(14) Pinan-2-ol
(15) α-thujene
(16) Sabinene
(17) Sabinene hydrate
(18) Bornyl diphosphate
(19) Borneol
(20) 3-carene
(21) 2-carene
(22) Camphene
(23) Camphene hydrate
(24) Fenchol
(25) Tricyclene



Scheme 1


HO
O
Spearmint
(–)-Carveol
(–)-Carvone
(+)-Menthofuran
OH
OGlc
(+)-Neomenthol
Neomenthyl-β-D-glycoside
Peppermint
OH
O
O
O
O
OH
O
O
(–)-(6)
(–)-trans-Isopiperitenol
(–)-Isopiperit-enone
(+)-cis-Isopulegone
(+)-Pulegone
(–)-Menthone
(–)-Menthol
Menthyl acetate
Perilla
OH
O
OH
OGlc
(+)-Isomenthone
(+)-Isomenthol
Isomenthyl-β-D-glycoside
Perilla alcohol
Perilla aldehyde
OH
O
(+)-Neoisomenthol
Neoisomenthyl acetate

Scheme 2

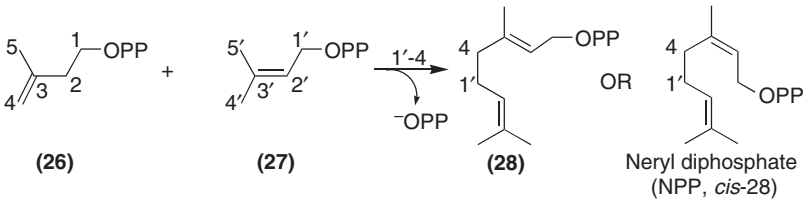
1.15.2 Cyclization Chemistry

1.15.2.1 Monoterpene Cyclases and Related Catalysts

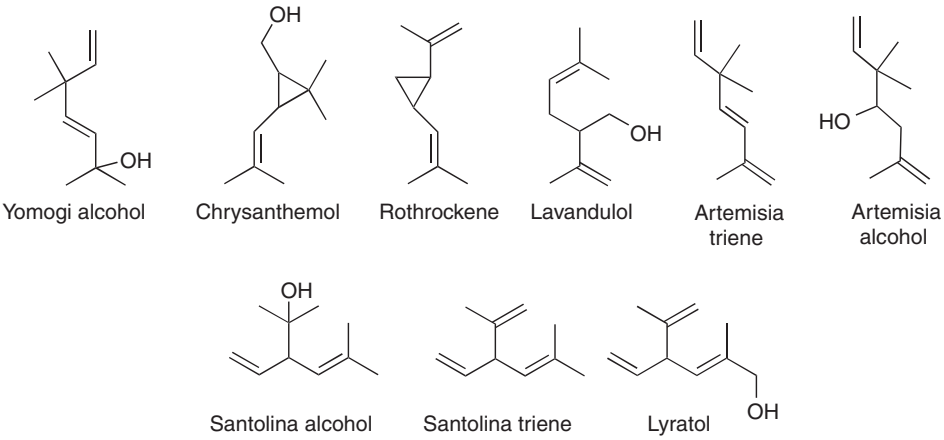
Terpene cyclases catalyze intramolecular, electrophilic C–C bond formation of allylic prenyl diphosphate substrates (GPP (**28**), NPP, farnesyl diphosphate (FPP (**29**)), and geranylgeranyl diphosphate (GGPP)) in a reaction analogous to the intermolecular condensations of the prenyltransferase family of bisubstrate enzymes (isoprenyl diphosphate synthases) in which an allylic Δ^2 prenyl diphosphate (**27**) and a Δ^3 prenyl diphosphate (**26**) are combined in a 1′-4 linkage (**Scheme 3**).[13,14] This linkage is prototypical in the chain elongation prenyltransferases including GPP synthase, but other condensation reactions that result in cyclopropyl (1′-2-3), branched (1′-2), or cyclobutyl (1′-2-3-2′) diphosphates have undergone recent characterization using prenyltransferase chimeras;[15] yet little is known regarding the downstream enzymology of the biosynthesis of the so-called irregular monoterpenes (**Scheme 4**). The cyclases and *trans*-prenyltransferases are structurally (discussed in Section 1.15.3.1) and mechanistically similar,[16,17] and it is notable that examples of prenyltransferases that catalyze cyclization-type reactions[18–20] and a cyclase that catalyzes a coupled prenyltransferase–cyclase reaction[21] have been reported (**Scheme 5**). Common to both enzyme types is the generation of highly reactive carbocation intermediates by diphosphate ionization (or in some diterpene cyclases double-bond protonation) with subsequent carbon–carbon bond formation resulting from electrophilic addition to a π-system, and reaction termination.[8,17] In contrast to the straightforward prenyltransferase reaction, terpene cyclase catalysis often involves myriad rearrangements and further cyclizations facilitated by guided carbocation movements throughout the initial and subsequently formed ring systems before termination by multiple nucleophile capture strategies, leading to an amazing structural diversity of terpenes.[8–10]

1.15.2.2 Monoterpene Cyclase Enzymology

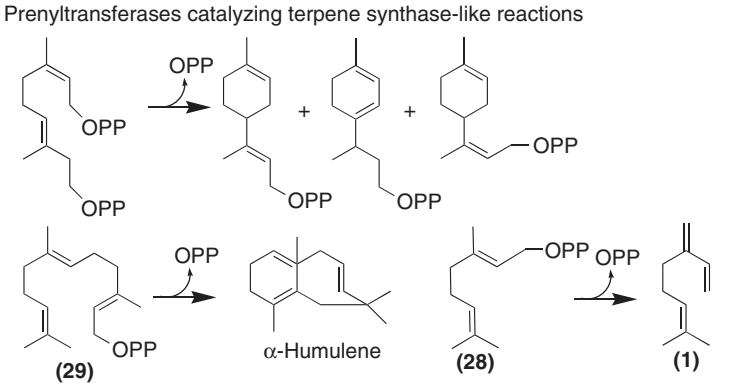
The monoterpene cyclases are quite similar to the sesquiterpene and diterpene cyclases in physicochemical properties[8–10,22] and have adopted largely similar strategies for substrate binding and product formation.[16,22]



Scheme 3



Scheme 4



Scheme 5

The very nature of the cyclization reactions of these three cyclase types and the common electrophilic carbocation attack on remote π-bonds suggest that the greater flexibility of the longer prenyl chains (C10 vs. C15 and C20) and the additional double bonds in their respective substrates will increase potential product diversity. Although essentially true, it is equally clear that the cyclization chemistry of the monoterpene cyclases is no less complex than that of their sesquiterpene and diterpene cyclase counterparts and, in that context, the monoterpene cyclases provide an adequate and simplified model for terpene cyclase chemistry.

Regarding the general properties of plant monoterpene cyclases, they are functionally soluble, 50–70 kDa monomers or homodimers, require a divalent cation (usually Mg^{2+} or Mn^{2+}), sometimes monovalent cation (K^+), have K_m values in the low micromolar range, and have relatively low turnover numbers ($<1\ s^{-1}$).[8–10,17] Most monoterpene cyclases are multiproduct enzymes with promiscuity profiles (level of minor products) ranging from trace to greater than 60% of the total, with product diversity often including several monoterpene structural types (**Scheme 1**) generally of essentially one enantiomeric series but near racemic cyclase products are known.[23–24] Yet, despite the apparent active site flexibility needed to generate such product diversity, monoterpene cyclases are generally specific in using only C10 prenyl diphosphate substrates, GPP (**28**), linalyl diphosphate (LPP (**30**)), or NPP (*cis*-isomer of **28**).[8–10]

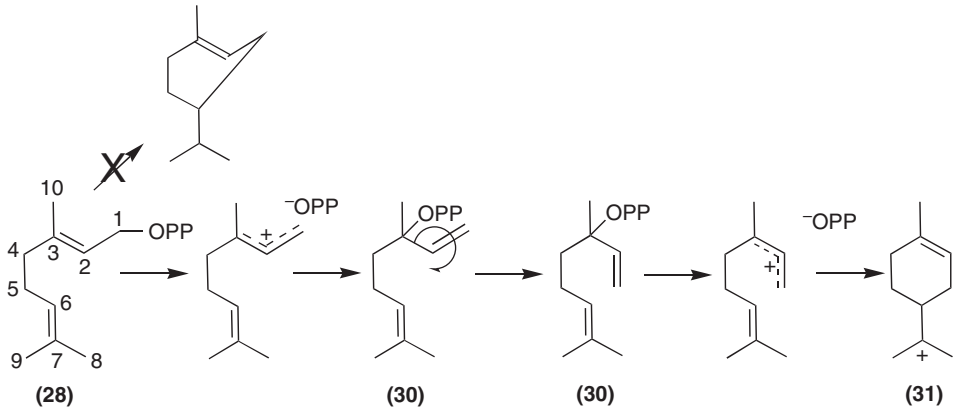
GPP is generally considered the natural C10 substrate of monoterpene cyclases[8–10] and early cell-free attempts to define substrate preference utilized a variety of cyclases representative of diverse plant families and nearly all monoterpene skeletal types and, despite the mechanistically facile, direct cyclization of the *cis*-Δ^2 prenyl diphosphate (NPP), most monoterpene cyclases preferred the *trans*-Δ^2 prenyl diphosphate (GPP)[25–29]. These studies were confounded by the preferential enzymatic hydrolysis of GPP (over NPP and LPP) by the contaminating phosphatases in the crude preparations; a situation remedied by further purification and separation of the competing activities and resulting in near universal preference for GPP.[30] However, recent characterization of the two recombinant enzymes, a prenyl transferase and a monoterpene cyclase, from cultivated tomato have clearly identified that the prenyl transferase is an NPP (**30**) synthase (presumably by

typical 1′-4 elongation, **Scheme 3**) and that the cyclase (β-phellandrene synthase) catalytic efficiency for NPP is 10^{14}-fold higher than for GPP.[31] As this is the first authenticated NPP synthase and the first, dedicated NPP cyclase the significance of this finding is considerable in the context of substrate availability and preference, and has mechanistic implications for substrate binding and ionization. It is notable that cyclization of NPP is an isomerization-independent reaction yet this cyclase retains the ability to catalyze the normally, tightly coupled isomerization–cyclization of GPP to β-phellandrene (**8**), albeit at reduced rates; an observation that may have evolutionary implications. This finding is consistent with the discovery in wild tomato of a *cis*-FPP synthase and a corresponding *cis*-sesquiterpene cyclase,[32] a cyclase also capable of cyclization of NPP but not GPP to α-terpineol (**10**). Regarding substrate preference, it is notable that ample studies have identified authentic GPP synthases from multiple, diverse plant families,[33–39] and as such, the taxonomic extent and significance of this finding await further biochemical characterization of enzymes from diverse plant families. It should also be noted that the universality of the α-terpinyl carbocation and the subsequent series of internal additions, cyclizations, rearrangements, and reaction terminations leading to the characteristic mono-, bi-, and tricyclic monoterpenes and monoterpenols (and diphosphate esters) catalyzed by both the *cis*- and *trans*-cyclases is in accord with the mechanistic model presented herein and reviewed elsewhere.[8–10]

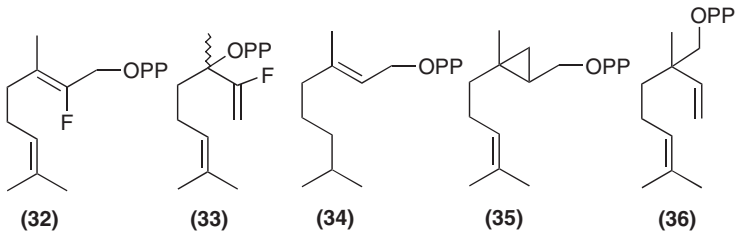
Methodical characterization of monoterpene cyclases from their native plant sources has afforded details on the mechanistic aspects of the cyclization cascade for enzymes that catalyze products of each structural type and stereochemistry and, as a body of work, has provided a detailed and comprehensive model of monoterpene cyclization.[8–10] Building on these initial cyclase studies, characterization of monoterpene cyclases from liverworts, ferns, and algae[40–42] has demonstrated that these 'lower eukaryote' cyclases resemble their higher plant counterparts in basic enzymological characteristics and suggests a common plant monoterpene cyclase progenitor. In contrast, recent cloning and characterization of the first microbial monoterpene cyclase[43,44] bolsters earlier evidence (from sesquiterpene cyclases) that microbial cyclases evolved catalytic function independently from plant cyclases yet have adopted similar strategies for binding the prenyl diphosphate substrate and catalyzing the electrophilic reaction.[16,22]

1.15.2.3 Ionization and Isomerization of Geranyl Diphosphate

The predominant precursor to monoterpene cyclization, GPP (**28**), is geometrically precluded from direct cyclization due to the *trans* geometry of the C2–C3 double bond, and hence requires ionization and isomerization to the linalyl system to accommodate C1–C2 bond rotation (**Scheme 6**, carbon numbering is based on **28**).[8] In support of the intermediacy of the tertiary intermediate, cyclization of LPP (**30**) by the monoterpene cyclases is ubiquitous, and product profiles thus generated are usually similar to those obtained with **28**; catalytic efficiency favors the cyclization of **30** over **28** by at least twofold,[45–49] whereas geraniol (**5**), linalool (**4**), and nerol (and their monophosphate esters) are not catalytically competent.[25,50,51] Successful demonstration of the electrophilic nature (and thus ionization) of the cyclization reaction was achieved through studies



Scheme 6



Scheme 7

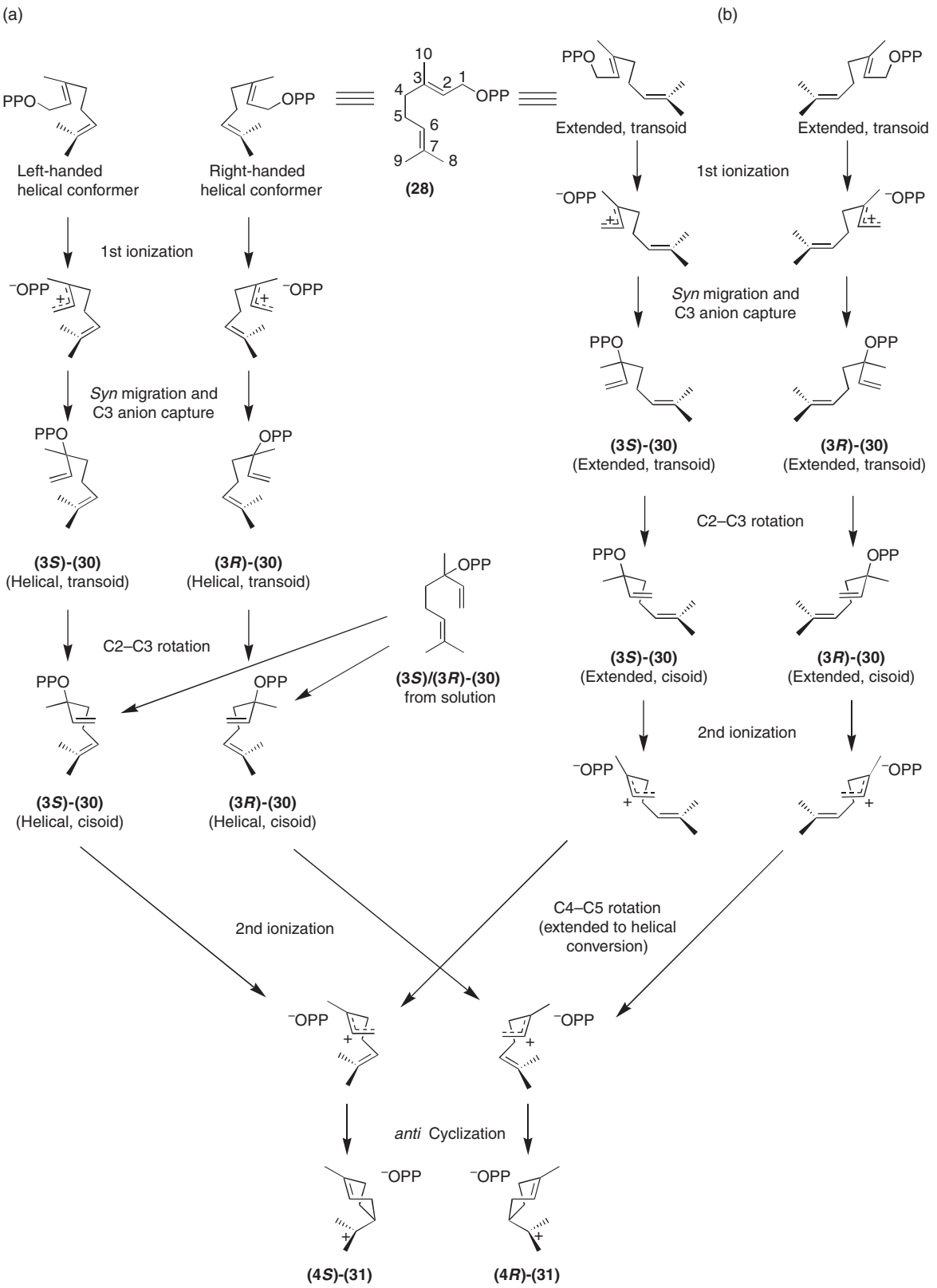
using the fluorinated, electron-withdrawing analogs 2-F-GPP (**32**) and 2-F-LPP (**33**) (**Scheme 7**) with the monoterpene cyclases[52,53] that showed minimally 100-fold rate suppressions and further provided evidence that the cyclase reaction is initiated by separate ionizations of **28** and **30**. The electrophilic nature of the reactions of sesquiterpene cyclases and *trans*-prenyltransferases was also demonstrated using appropriate allylic, fluorinated substrate analogs.[54,55] Recent kinetic analyses using these monoterpene analogs with purified, recombinant (−)-limonene synthase ((−)-LS) and (+)-bornyl diphosphate synthase ((+)-BS) have bolstered this early work with native cyclases and demonstrated that both fluoroanalogs are effective competitive inhibitors with inhibition constants in the low micromolar range (6–79 $\mu mol\,l^{-1}$).[56] Furthermore, the competitive nature of the inhibition is evidence that both ionizations occur in the same active site.

The intermediacy of LPP (**30**) in the reaction cycle had never been directly observed and, coupled with the greater cyclization efficiency of **30** versus **28** by the cyclases, suggests a tightly coupled reaction sequence.[8,9] The noncyclizable analogs 6,7-dihyroGPP (**34**) and 2,3-cyclopropyl GPP (**35**) (**Scheme 7**) were used with native cyclases in an effort to directly observe the release of the corresponding LPP analogs and, in all cases tested, products were identified that were entirely consistent with the formation of **30** including preliminary identification of 3-homoLPP (**36**), the expected LPP analog derived from the isomerization of 2,3-cyclopropyl GPP (**35**).[57,58] This work was repeated with purified, recombinant (−)-LS, (+)-BS, and (−)-pinene synthase ((−)-PS) in an attempt to generate products at increased quantities for unambiguous identification and stereochemical assignment.[59] Unfortunately, no detectable diphosphorylated monoterpenes were generated, though product type and stereochemistry did permit identification of several mechanistic distinctions between these different cyclase types, most notably in the biosynthesis of significant quantities of monoterpenols from both analogs with (+)-BS but not from the olefin cyclases (−)-PS and (−)-LS. Structural evidence for these apparent differences will be discussed in Section 1.15.3.3.

Although the studies with the substrate analogs (described above) failed to detect the formation and release of the cryptic tertiary intermediate, and despite the low turnover demonstrated by (−)-LS of 2-F-GPP (**32**) and 2-F-LPP (**33**) to fluorolimonene,[56] these analogs were successfully used in the crystallization of (−)-LS (see Section 1.15.3.2).[60] Fortuitously, (−)-LS crystallized in the presence of **32** catalyzed the isomerization of **32** to **33**, wherein the electron-withdrawing properties of the allylic fluorogroup prevented the second ionization and subsequent cyclization. This represents the first direct demonstration of LPP as an intermediate in monoterpene cyclization observed in the context of the active site, and presumably in a conformation relevant to cyclization.

1.15.2.4 Cyclization Reaction Stereochemistry

During the course of ionization and isomerization of GPP (**28**), the first chiral center is established at C3 of LPP (**30**), and its significance in the stereochemistry of subsequent transformations and in product formation is absolute. The ability to successfully separate this two-component reaction pathway using exogenous substrates of enantiomeric purity has provided the opportunity to elucidate several mechanistic and stereochemical aspects of monoterpene cyclization (these experiments were reviewed by Croteau,[8] Wise and Croteau,[9] and Davis and Croteau[10]), and has led to the development of a comprehensive model of monoterpene cyclization.[8] The model suggests that the stereochemical outcome is likely a consequence of initial binding of **28** in the cyclase active site in either a right-handed or a left-handed helical orientation and, following ionization, there occurs suprafacial migration of the diphosphate across the C1–C2–C3 plane to produce (3*S*)-LPP or (3*R*)-LPP, exclusively (**Scheme 8(a)**). A transoid to cisoid transition of the enantiomers of **30** is dependent on C1–C2 bond


(a)
(b)
PPO
OPP
10
3
2
1
4
5
6
7
9
8
(28)
Left-handed helical conformer
Right-handed helical conformer
Extended, transoid
1st ionization
⁻OPP
Syn migration and C3 anion capture
(3S)-(30) (Helical, transoid)
(3R)-(30) (Helical, transoid)
(3S)-(30) (Extended, transoid)
(3R)-(30) (Extended, transoid)
C2–C3 rotation
(3S)-(30) (Extended, cisoid)
(3R)-(30) (Extended, cisoid)
2nd ionization
(3S)-(30) (Helical, cisoid)
(3R)-(30) (Helical, cisoid)
(3S)/(3R)-(30) from solution
C4–C5 rotation (extended to helical conversion)
anti Cyclization
(4S)-(31)
(4R)-(31)

Scheme 8

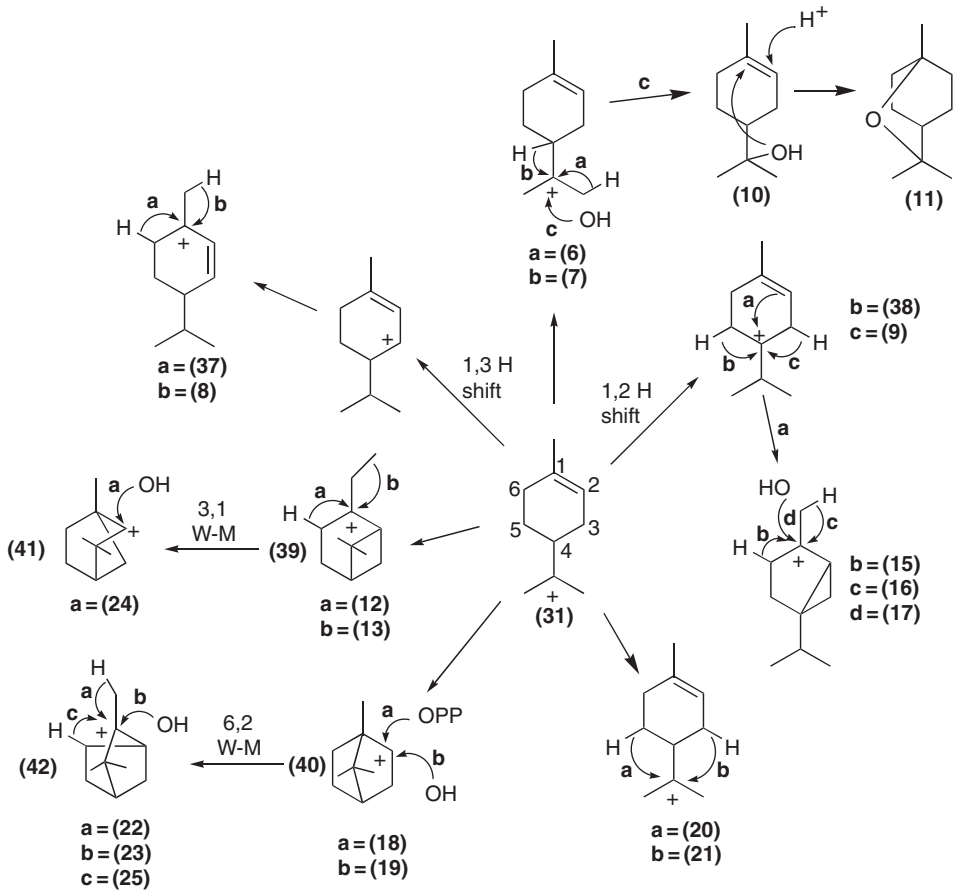
rotation, precedes ionization of **30**, and is necessary for proper juxtaposition of C1 with C6 to facilitate π-cation cyclization. Following the ionization of **30**, cyclization occurs *anti*, *endo* (*anti* refers to the face of C1 bond formation relative to the face of C1 diphosphate departure and *endo* refers to the spatial orientation of the olefinic chain) to produce the (4*S*)-α-terpinyl cation from (3*S*)-(**30**) and the (4*R*)-α-terpinyl cation from (3*R*)-(**30**) with subsequent formation of products of the (−)- and (+)-antipodal series, respectively. The consequence of *syn*-migration of the diphosphate and *anti*, *endo* cyclization is that the overall retention of configuration at C1 will be observed in the product. Retention of configuration has been clearly established for numerous cyclases in separate assays using (1*R*)- and (1*S*)-[^{3}H]-GPP; following cyclization of the respective substrate, the product was converted to a ketone wherein a stereoselective, base-mediated proton exchange permitted stereochemical determination and deduction of configuration at C1 of the substrate.[46,47,61–63] These results, coupled with the stereospecificity of cyclization of either (3*R*)- or (3*S*)-(**30**), fully support the model outlined.[62,63]

As noted in the previous section, enzymatically formed 2-F-LPP (**33**) was observed in the (−)-LS crystal structure, and it is noteworthy that electron density was consistent with **33** in an extended, transoid orientation, which suggests that GPP (**28**) binds similarly.[60] Thus, a refinement of the cyclization model was presented (**Scheme 8(b)**) in which **28** binds and isomerizes to **30** in the extended, transoid conformer, an orientation from which the electrophile is too distant from the π-system for cyclization to occur, and after C2–C3 bond rotation attains the extended, cisoid conformation. Finally, ionization of **30** and C4–C5 bond rotation presents the helical, cisoid linyl cation for cyclization. It is notable that 2-F-LPP (**33**), when crystallized with (−)-LS, is bound in the helical, cisoid conformer (an orientation competent for cyclization), and thus likely represents an artifact of presentation to the cyclase in an energetically favored form (from solution) rather than as an intermediate in the dynamic reaction cascade but, nonetheless, suggests that the C4–C5 bond rotation occurs after the second ionization, or the rotation is sufficiently slow such that no interchange occurs in the crystalline state.[60]

Having established stereochemical considerations for GPP isomerization to specific stereoisomers of LPP, and their cyclization to (4*R*)- and (4*S*)-α-terpinyl cation by a common monoterpene cyclase mechanism, it is clear that further manipulations are enzyme specific and, thus, the central importance of the α-terpinyl carbocation (**31**) in product formation cannot be overemphasized. This universal cyclohexenoid carbocation, as a consequence of its formation from either (*R*)- or (*S*)-LPP, determines the stereochemical outcome of the products, and its precise orientation in, and its interactions with, the cyclase active site determines if and how the course of carbocation migration through hydride shifts, secondary electrophilic cyclizations, and C–C bond rearrangements will proceed prior to reaction termination.

1.15.2.5 Formation of Monoterpene Structural Types

The acyclic dimethyloctane monoterpenes result from premature deprotonation of, or water capture of, the geranyl or linalyl cations, and afford *cis*-/*trans*-ocimene (**2**)/(**3**) or myrcene (**1**), or linalool (**4**) and geraniol (**5**). The simplest set of cyclic monoterpenes are the *p*-menthane type, which result from alternate, direct deprotonations of the α-terpinyl carbocation (**31**) (**Scheme 9**) to generate limonene (**6**) or terpinolene (**7**) or, involve water capture to generate α-terpineol (**10**) (which following protonation of the C–C double bond (of α-terpineol) and internal addition leads to 1,8-cineole (**11**)). A 1,2- or a 1,3-hydride shift (delocalization of charge on the latter secondary allylic cation leads to the more stable tertiary cation) and subsequent deprotonations produce α- and γ-terpinene (**9** and **38**, respectively) and α- and β-phellandrene (**37** and **8**, respectively). The thujane monoterpenes proceed by a 1,2-hydride shift followed by cyclopropyl ring closure and tertiary cation quench, by deprotonation in the case of α-thujene (**15**) and sabinene (**16**), or by water capture to produce sabinene hydrate (**17**). Carane monoterpenes undergo opposing 1,3-proton eliminations from C3 or C5 to generate the cyclopropyl ring of 2-carene (**21**) and 3-carene (**20**), respectively. The pinane and bornane monoterpenes undergo bicyclization at opposing ends of the remaining double bond of the α-terpinyl cation (**31**) to generate the favored tertiary cation (pinyl cation (**39**)) or a less-favored, anti-Markovnikov secondary cation (bornyl cation (**40**)). Alternate deprotonations of the former generate α-pinene (**12**) and β-pinene (**13**), whereas diphosphate or water capture leads to bornyl diphosphate (BPP (**18**)) and borneol (**19**), respectively. Wagner–Meerwein (W–M) rearrangements of the bornyl cation (**40**) (6,2-rearrangement) and pinyl cation (**39**) (3,1-rearrangement) lead to



Scheme 9

the isocamphyl (**42**) and fenchyl (**41**) cations, respectively. In the latter case, water capture leads to fenchol (**24**), whereas termination in the former case by deprotonation, water capture, or a deprotonation-mediated ring closure results in camphene (**22**), camphene hydrate (**23**), and tricyclene (**25**), respectively.

It is significant that the above descriptions, meant to provide a survey of cyclic monoterpene chemistry, should in no way be considered complete. Each carbocation may undergo alternate electrophile capture (i.e., water capture rather than deprotonation), and capture may occur from different faces of the carbocation, situations that expand the cyclase repertoire and clearly emphasize potential cyclase versatility.

1.15.2.6 Molecular Biology of the Monoterpene Cyclases

Although it is clear that a limited number of monoterpenes are generated as a result of monoterpene cyclases (estimated at 30–40[8,9]), the promiscuous nature of these cyclases suggests an infinite number of permutations in product distributions, reflected by changes in amino acid sequence and thereby structure.

Consistent with this observation, recent genome projects suggest that terpene synthase (TPS) genes are present in very high numbers: *Arabidopsis*, not previously recognized as a prolific terpene-producing plant, has at least 40 TPS genes;[64] black cottonwood, *Populus trichocarpa*, has nearly 50;[65] and a wine grape sequencing project has identified 89 TPS genes (of these, the authors estimate that 35 are monoterpene cyclases).[66] Currently, at least 50 recombinant monoterpene cyclases from multiple, evolutionarily conserved and diverse plant sources have been characterized, and as a group they generate most of the major monoterpene structural types, each with a different degree of reaction promiscuity (based on the types and ratios of minor products) that provides a vast resource for structural and mechanistic studies previously not possible.

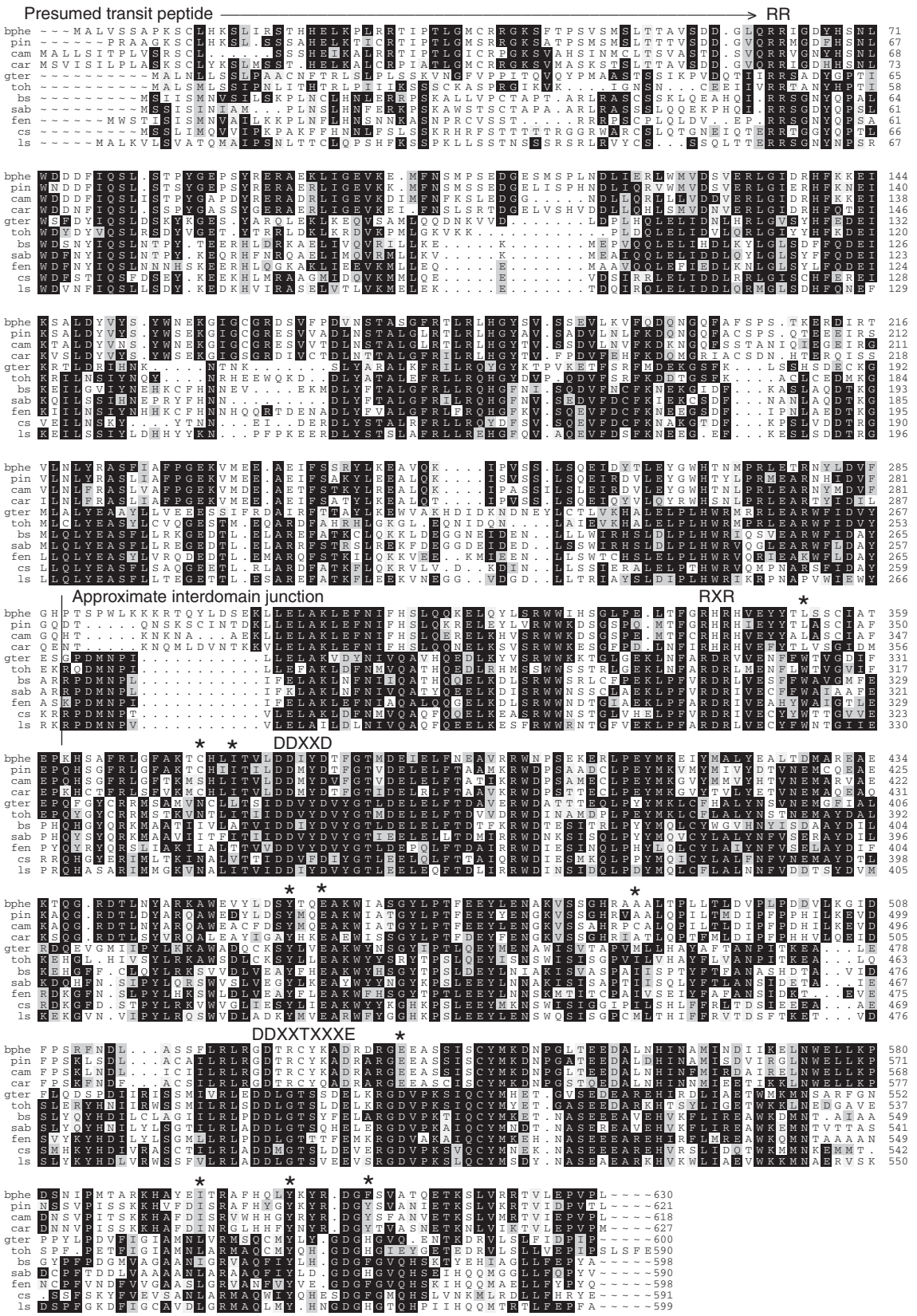
An alignment of representative monoterpene cyclases indicates that enzymes from quite diverse plant sources nonetheless retain high conservation, and several highly conserved sequence motifs are apparent (**Figure 1**). The TPSs are classified according to sequence relatedness and thus phylogenetic origin, and fall into seven subfamilies, TPS-a through TPS-g.[67–69] Angiosperm monoterpene synthases fall largely into the TPS-b and TPS-g subfamilies, whereas their gymnosperm counterparts are members of TPS-d, a complex family with further subdivisions.[69] The subfamilies b and d are not exclusive to monoterpene cyclases nor are they completely inclusive of all C10 cyclases, an observation that suggests that the evolution of terpene cyclases has proceeded along multiple paths; subfamilies g and f contain mostly acyclic monoterpene synthases.[69] A complementary classification system considers the inton–exon structure and intron number of these cyclase genes, and a proposal for evolutionary descent from a presumed progenitor cyclase was developed based on intron gains and losses through evolutionary time.[70,71] A recent *in silico* survey of *Arabidopsis* TPS genes uncovered evidence for multiple gene duplication events and significant clustering of cyclases and prenyltransferases, and suggests a mechanism for the observed diversity of terpene cyclases and their respective products and pathways.[68] Genome analysis and cyclase gene categorization systems are instructive and, in the context of the immense sequence data available, provide a means for gene characterization. However, it must be stressed that terpene cyclase activity is not predictable from sequence; characterization, coupled with enzymological and structural studies, is far more informative.

1.15.3 Structural Studies

1.15.3.1 Structural Commonalities

As noted earlier, the mechanistic resemblance between the prenyltransferases and terpene cyclases is striking, though little sequence homology is shared. Not surprisingly, there is considerable sequence homology among the plant terpene cyclases, and common evolutionary lineages have been noted;[16,67,68] however, the microbial cyclases, which catalyze similar reaction types and generate structurally related products, share no sequence similarity with their plant counterparts (nor among themselves),[9,10,22] and neither group shares significant sequence similarity with the prenyltransferases.[72] In spite of these primary structural differences, a common α-helical fold has been adopted by these enzyme types to accommodate similar chemistries.[16,22,60,73–75] This active site fold has been termed the 'class I TPS fold' and contrasts with the structurally unrelated, α-helical 'class II TPS fold' common to the C30 triterpene cyclases.[16,72] The latter catalysts initiate carbocation formation by C–C double bond or epoxide protonation instead of initiation by diphosphate ester ionization, but cyclization chemistry proceeds through electrophilic addition to π-systems similar to the short-chain terpene cyclases. Notably, the plant diterpene cyclase, abietadiene synthase, catalyzes a protonation-initiated cyclization and an ionization-initiated cyclization in two separate active sites,[76] and although no structure has been solved for this bifunctional cyclase, homology suggests that it will adopt a fold characteristic of the known plant terpene cyclases.[10,16]

Structures are available for four plant terpene cyclases: the sesquiterpene cyclase, 5-*epi*-aristolochene synthase,[77] and three monoterpene cyclases (discussed below).[60,73,78] The structural similarities between these cyclases are consistent with their largely shared mechanistic and physicochemical properties and permit several generalizations to be made (**Figure 2**). The typical plant TPS is a bidomain structure with N- and C-terminal domains of approximately 8 and 12 α-helices, respectively. Structures obtained using a variety of substrate, intermediate, and product analogs suggest that the N-terminal domain has no direct role in catalysis and has adopted a gluconase/glucoamylase-like fold with similarity to the class II fold of triterpene cyclases.[16,72] No known function has been attributed to this domain but its absence from the microbial cyclases, and the experimental evidence that catalytic specificity of plant monoterpene and sesquiterpene cyclases resides in the C-terminal domain,[79–81] suggests that it may be relictual; yet its considerable sequence conservation suggests otherwise. A notable element of the N-terminal domain consists of a largely unstructured loop at the terminus that becomes ordered upon substrate binding and caps the active site cavity, presumably to prevent solvent access to the highly reactive carbocations. A potential role for residues of this loop in monoterpene cyclase catalysis is discussed below.


Presumed transit peptide
RR
Approximate interdomain junction
RXR
DDXXD
DDXXTXXXE

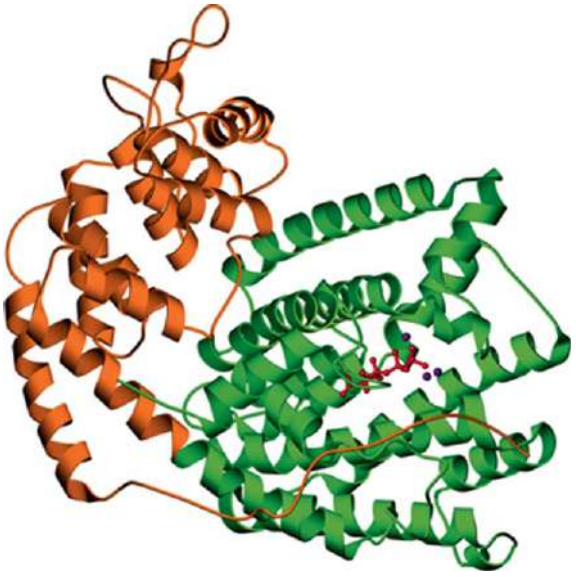


Figure 2 Ribbon plot of a monomer chain of (−)-LS showing the helical N-terminal domain (orange) and C-terminal active site domain (green). The analog 2-F-LPP (red) and the divalent cation Mn^{2+} (purple) are shown and were used in cocrystallization. Note the orientation of the N-terminal loop over the active site opening. Reproduced with permission from D. C. Hyatt; B. Youn; Y. Zhao; B. Santhamma; R. M. Coates; R. B. Croteau; C.Kang, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*(13), 5360–5365. Copyright 2007 National Acadamy of Sciences, USA.[60]

Analog-bound structures provide additional evidence that cyclization chemistry occurs exclusively within the C-terminal domain, in a cavity commonly composed of a handful of α-helices and adjoining loops. Typical active site features have been noted, including a polar, active site opening for diphosphate binding and a relatively deep, hydrophobic prenyl-binding pocket replete with aromatic residues.[16,60,73,77,78] The strategy for allylic diphosphate binding is shared among the plant terpene cyclases, and has many notable similarities with FPP synthase[10,16] and the microbial cyclases,[16,22,82] including a conserved DDxx(D/E) motif involved in diphosphate coordination through divalent cations. In addition to this highly conserved aspartate-rich motif, plant cyclases have an adjacent (in space) RxR motif and an opposing (D/N)Dxx(T/S)xxxE motif (designated the DTE/NSE motif), which provide all the requisite features for diphosphate recognition and ionization by the typical plant cyclases (**Figure 1**).[16] A conspicuous lack of DTE/NSE motif conservation is apparent in the gymnosperm monoterpene cyclases and suggests that these enzymes have adopted an alternate diphosphate-binding feature. Gymnosperm cyclases also exhibit monovalent cation dependence (K^+) and a conserved serine, several residues downstream from the modeled location of the angiosperm NSE/DTE motif, was implicated in K^+-binding by site-directed mutagenesis of a spruce (−)-PS (corresponding to residue S541 in bphe, **Figure 1**).[92] The precise nature of the divalent and monovalent cation interactions with the alkyl diphosphate in the gymnosperm monoterpene cyclase active sites awaits the determination of a representative structure.

Figure 1 Alignment of representative monoterpene cyclase amino acid sequences. The plastid-targeting sequence is shown with an approximate cleavage site indicated. Conserved sequence motifs are noted above their respective locations. The positions of selected active site residues of (+)-BS, (−)-LS, and CS are identified with an asterisk. The key to abbreviations (with accession numbers and plant sources): bphe, β-phellandrene synthase (AF139205), *Abies grandis*;[83] pin, (−)-α/(−)-β-pinene synthase (AAS47692), *Picea abies*;[84] cam, (−)-camphene synthase (AAB70707), A. grandis;[85] car, (+)-3-carene synthase (AAO73863), *P. abies*;[86] gter, γ-terpinene synthase (AAM53943), *Citrus limon*;[87] toh, α-terpineol synthase (AAS79352), *Vitis vinefera*;[88] bs, (+)-bornyl diphosphate synthase (AAC26017), *Salvia officinalis*;[89] sab, (+)-sabinene synthase (AAC26018), *S. officinalis*;[89] fen, fenchol synthase (AAV63790), *Ocimum basilicum*;[90] cs, 1,8-cineole synthase (ABH07677), *Salvia fruticosa*;[78] ls, (−)-limonene synthase (AAC37366), *Mentha spicata*.[91]

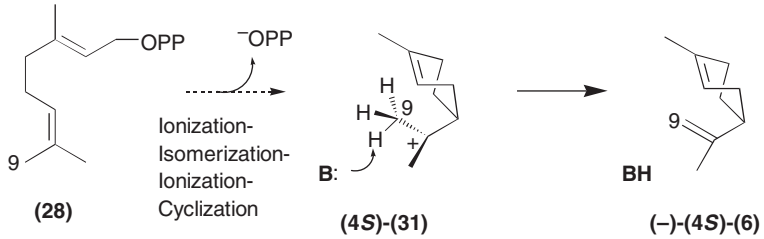
The three mechanistically unique monoterpene cyclase structures are presented in the context of the mechanistic and stereochemical analyses of their respective reaction cascades (from the α-terpinyl cation (**31**) to product) to provide substantive details. Therefore, it is important to consider each active site cavity as promoting ostensibly different reaction cascades, and in that context to seek out structural differences. The first enzyme presented, (−)-LS, catalyzes a simple deprotonation to (−)-limonene (**6**) and generates only minor amounts of side products, all of which are olefinic and of the (−)-antipodal series. In contrast, (+)-BS terminates its reaction three-quarters of the time by diphosphate capture, produces largely (+)-series monoterpenes, and is relatively promiscuous in terminating its remaining reaction cycles by deprotonation. The final structure considered is that of 1,8-cineole synthase (CS), a cyclase that has a stable monoterpenol intermediate (that debinds occasionally), requires π-bond protonation before cyclic ether formation, and terminates one-third of its reaction cycles by deprotonation.

1.15.3.2 (−)-(4*S*)-Limonene Synthase

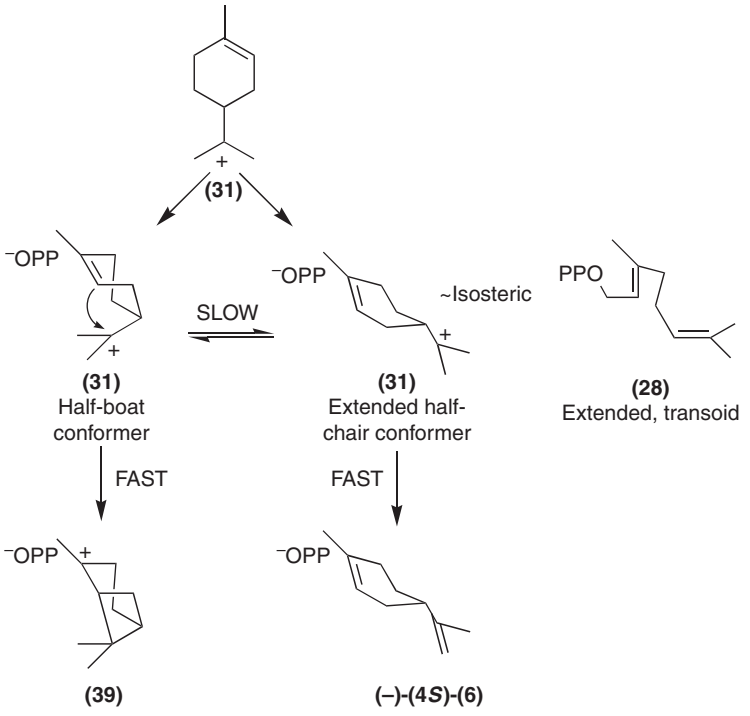
Limonene (**6**) is a common component of many plant essential oils, and biotransformation of the (−)-antipode in the mints leads to a variety of scented compounds (**Scheme 2**), whereas in *Citrus* the (+)-antipode is largely responsible for the characteristic 'peel aroma' associated with oranges. Investigations using (−)-LS as a model monoterpene cyclase have provided a wealth of information regarding mechanistic (reviewed by Croteau,[8] Wise and Croteau,[9] and Davis and Croteau[10]) and, recently, structural aspects of this, the simplest of all monoterpene cyclizations.[60] Considerable enzymological evidence obtained with (−)-LS has provided support for the general model of isomerization and cyclization of GPP to the (4*S*)-α-terpinyl cation (as outlined in the previous section) and supports the general application of this model.[8–10]

Several informative studies were undertaken with antipodal LSs, and with two sage PSs that generate either (+)- or (−)-limonene (**6**) in significant side reactions. One study sought to determine the nature of methyl deprotonation of the α-terpinyl cation (**31**) by (+)- and (−)-LS and the PSs and used [8-^{3}H]-GPP with chemical derivitization-assisted product analysis to determine that methyl deprotonation occurs exclusively from the *cis*-methyl (C9) for the antipodal cyclases from *Citrus* and *Mentha*[93–95] (an observation consistent with natural abundance NMR[96]), whereas the PSs showed little discrimination between *cis*- and *trans*-methyl deprotonation. These results suggest that dedicated LSs (of either enantiomer) may be more restrictive in α-terpinyl cation (**31**) binding than a cyclase that is more committed to other reaction types (i.e., bicyclization by PS). Stereochemical determination of the methyl-to-methylene elimination was accomplished with assays of the mint (−)-LS using C9-trilabeled GPP (^{1}H, ^{2}H, and ^{3}H). The NMR-based analysis demonstrated *anti*-elimination (relative to C1–C6 bond formation) and suggests a precisely positioned isopropyl group with clearly defined contacts that are maintained through the cyclization cascade.[95] The above experimental evidence provides a comprehensive stereochemical model for the deprotonation of the (4*S*)-(**31**) to yield limonene (**6**) (**Scheme 10**).

Intramolecular isotope effects associated with the methyl-methylene elimination (in limonene (**6**) production) were determined using the mint (−)-LS and the PSs and [8,8,9,9-$^2H_4$]-GPP.[95] Significant effects were demonstrated for all three cyclases as expected, and are consistent with primary kinetic isotope effects associated with deuterium elimination. These experiments provided a means (using [8,9-$^2H_6$]-GPP) to observe isotopically sensitive branching effects with these multiproduct cyclases, since the branch point between **6** and pinene production requires C8- or C9-deprotonation or bicyclization, respectively.[95] Quite unexpectedly, no significant isotope effect was observed for all three cyclases, and it was suggested that the lack of competition between monocyclization and bicyclization is due to the presence of conformational isomers of the α-terpinyl cation (**31**),



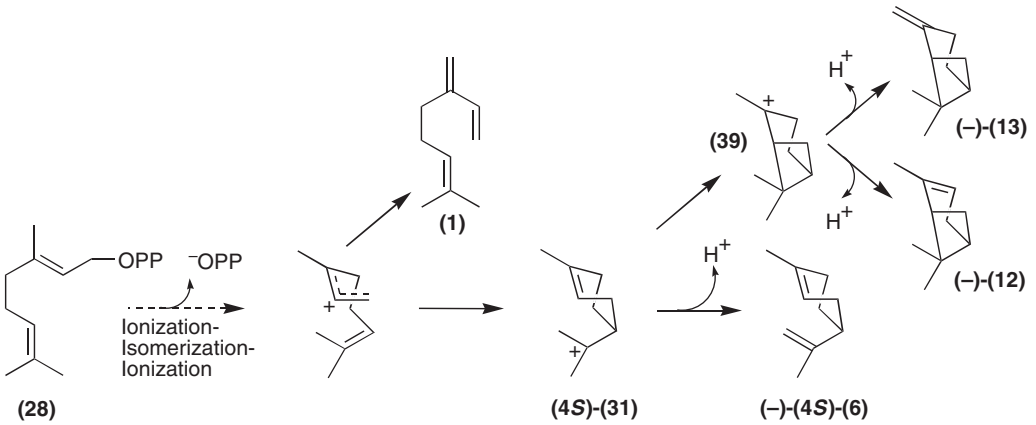
Scheme 10



Scheme 11

the extended half-chair and the half-boat, respectively, whose interchange in the active site is slow compared with deprotonation or bicyclization (**Scheme 11**).[95] The prescience of this proposal made in the absence of cyclase structural information is notable in the light of the recent crystallographic studies of (−)-LS, in which 2-F-GPP (**32**) was isomerized to 2-F-LPP (**33**) and adopted an extended, transoid orientation that is nearly isosteric to the extended, half-chair orientation of the predicted α-terpinyl conformer that leads to limonene (**6**) (**Scheme 11**). It is tempting to speculate that the extended, transoid conformation of **33**, observed in the active site of (−)-LS, is reflective of an active site contour that promotes the extended half-chair conformer of the **31** for generation of **6**; the half-boat conformer required for pinene production occurs rarely in a dedicated LS.

(−)-LS was cloned and characterized from *Mentha spicata*[91] and has typical terpene cyclase properties,[97,98] including reaction kinetics and a preference for Mg^{2+}. This enzyme is minimally promiscuous in generating largely **6** and essentially only products of the (−)-antipodal series (**Scheme 12**). The latter observation is



Scheme 12

entirely consistent with the mechanistic studies presented that suggest a very restrictive active site cavity. Native (−)-LS has been localized to the plastid[99] and its modest expression from the full-length cDNA construct suggested that removal of the N-terminal targeting peptide for expression and characterization might be fruitful. A series of N-terminal truncations (using the cDNA) and kinetic analyses of the resultant recombinant proteins identified a pseudomature form of (−)-LS that included the highly conserved, tandem arginine motif (**Figure 1**, (−)-LS residues 58 and 59) and provided a cyclase with kinetics comparable to native (−)-LS.[100] Mutagenesis or removal of either arginine resulted in isomerization-incompetent cyclases whose cyclization capacity was intact (modest cyclization of LPP was demonstrated with truncation to residue 89), suggesting that the tandem arginine motif is involved in the binding or recognition of GPP, or in the isomerization of GPP to LPP, but not in cyclization chemistry, *per se*.

As noted earlier, the structure of (−)-LS from *M. spicata* was obtained using the substrate and intermediate analogs 2-F-GPP (**32**) and 2-F-LPP (**33**), respectively, and **32** was isomerized in the (−)-LS crystal to **33**, which adopted an extended, transoid orientation.[60] On the contrary, **33**, presented from solution, bound in the helical, cisoid conformer (**Schemes 8(a) and 8(b)**).[60] The (−)-LS structure is typical of plant cyclases (**Figure 2**, described above) and assumes a homodimer in the crystalline state, in contrast to studies on the native enzyme that suggest monomer formation in solution.[98] Binding of the substrate and intermediate analogs and three divalent metal ions occurs exclusively in the C-terminal domain; no direct interaction between substrate or intermediate analogs occurs in the N-terminal domain. The active site is in closed conformation with the terminal, tandem arginine residues providing bonding interactions with the C-terminal residues, E363, Y435, and the backbone carbonyl of V357. It is probable that active site closure occurs upon diphosphate binding, and that initiation of the isomerization step cannot happen without this interaction. Three Mn^{2+} ions (Mn^{2+} was used for crystallization rather than the preferred cation Mg^{2+}) and the diphosphate are clearly positioned in the active site lip but subunit differences exist and suggest that alternate diphosphate binding orientations arise even in the highly restrictive (−)-LS active site. The prenyl-binding cavity of (−)-LS is largely hydrophobic with two prominent aromatic residues, Y573 and W324, positioned below the isopropyl group. These are perhaps involved in π-carbocation stabilization of the α-terpinyl cation (**31**), a stabilization of mechanistic significance given the regio- and stereospecificity of the terminating methyl-methylene elimination. Potential termination nucleophiles are plentiful (i.e., backbone carbonyls, active site waters, and the diphosphate anion), and, given the reactivity of the carbocations, any of several proximal residues could also fulfill this role (B: in **Scheme 10**); a suitably positioned asparagine (N345) and histidine (H579) are attractive candidates. An active site histidine was predicted by substrate-mediated protection in the presence of a histidine-directed reagent.[101] Clearly, the (−)-LS structure provides critical information on the structural determinants of a simple monoterpene cyclase for which there exists extensive mechanistic information, but perhaps of greater significance are the mechanistic implications uncovered by the use of fluoroanalogs with this enzyme, which provides details of substrate binding and conformation in the context of the isomerization component of the monoterpene cyclase reaction cycle. It is not clear at this time whether the binding conformation of the enzyme-generated 2-F-LPP (**33**) is specific to (−)-LS or is a more universal aspect of monoterpene cyclization.

1.15.3.3 (+)-Bornyl Diphosphate Synthase

The (+)- and (−)-antipodes of camphor are especially fragrant components in the oils of sage and tansy, respectively, and the committed step in their biosynthesis is catalyzed by the antipodal BSs from the respective plants. An immense body of work has contributed to the understanding of many aspects of the reactions of these cyclases and to the more general monoterpene cyclase mechanism discussed above and previously reviewed in detail.[8,9] It is instructive, however, to discuss several aspects of the cyclization cascade that are specific to (+)-BS and contribute to the context of the (+)-BS crystal structure.

Displacement of the diphosphate from C1 to C3 (isomerization) and from C3 to C2 (termination) suggests that (+)-BS must have unique structural features compared to other cyclase types to accommodate this apparent movement of the diphosphate anion. A series of positional isotope effect studies were conducted that used [1-^{18}O]-GPP, [α-^{32}P]-GPP, [β-^{32}P]-GPP, and the correspondingly labeled LPP compounds to determine the extent of diphosphate scrambling in the course of the reaction to BPP (**18**).[102,103] Product

analysis involved derivitization and MS analysis for the [^{18}O]-BPP and selective monophosphate hydrolysis and radio analysis for the [^{32}P]-BPP. The cumulative results of these studies unambiguously demonstrated that the phosphoester oxygens of substrate and product are unchanged, suggesting an ordered removal and movement of the diphosphate without P–O–P bond rotation or end-to-end exchange.

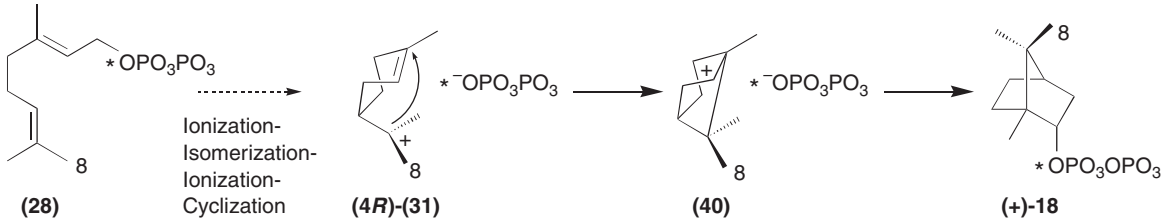
A comparison of inhibitory properties between (+)-BS and (−)-LS using 2-F-GPP (**32**) and 2-F-LPP (**33**) found that these competitive inhibitors had largely similar K_i values for both cyclases except in the case of **32** with (+)-BS (10-fold lower inhibition constant).[56] This contrast in inhibition may be reflective of differences in active site architecture that are specific to the different reaction chemistries of these cyclases. Such differences are also supported by inhibitor studies conducted with (+)-BS, (−)-LS, and a gymnosperm (−)-PS using the analogs 6,7-dihydroGPP (**34**) and 2,3-cyclopropylGPP (**35**).[59] The strictly olefin cyclases, (−)-LS and (−)-PS, generated largely hydrocarbon products from these analogs, whereas (+)-BS generated mostly monoterpenols, suggesting access and involvement of water in the (+)-BS reaction, which is not observed in the normal reaction cycle. The involvement of this normally unreactive water in the (+)-BS reaction will be addressed below, in the context of the structure.

Bicyclization in the BS reaction cascade may proceed through two different pathways, direct cyclization to the bornyl cation (**40**) (anti-Markovnikov) and cyclization to the pinyl cation (**39**) followed by rearrangement to **40**. Evidence gained from isotopically sensitive branching experiments (using [10-$^2H_3$]-GPP) with two sage PSs (both cyclases generate camphene (**22**) and α-pinene (**12**)) demonstrated that bicyclization to **40** (en route to **22**) proceeded directly (anti-Markovnikov cyclization) in one case and through the transient **39** with subsequent rearrangement to **40** in the other case.[104,105] Thus, precedent was established for both mechanisms of bicyclization by these olefin cyclases. The bicyclization mechanism for (+)-BS is not known but it has been suggested that the localization of the diphosphate over C2 could polarize the C1–C2 double bond, thus leading directly to **40** formation, as a consequence of diphosphate positional effects.[106]

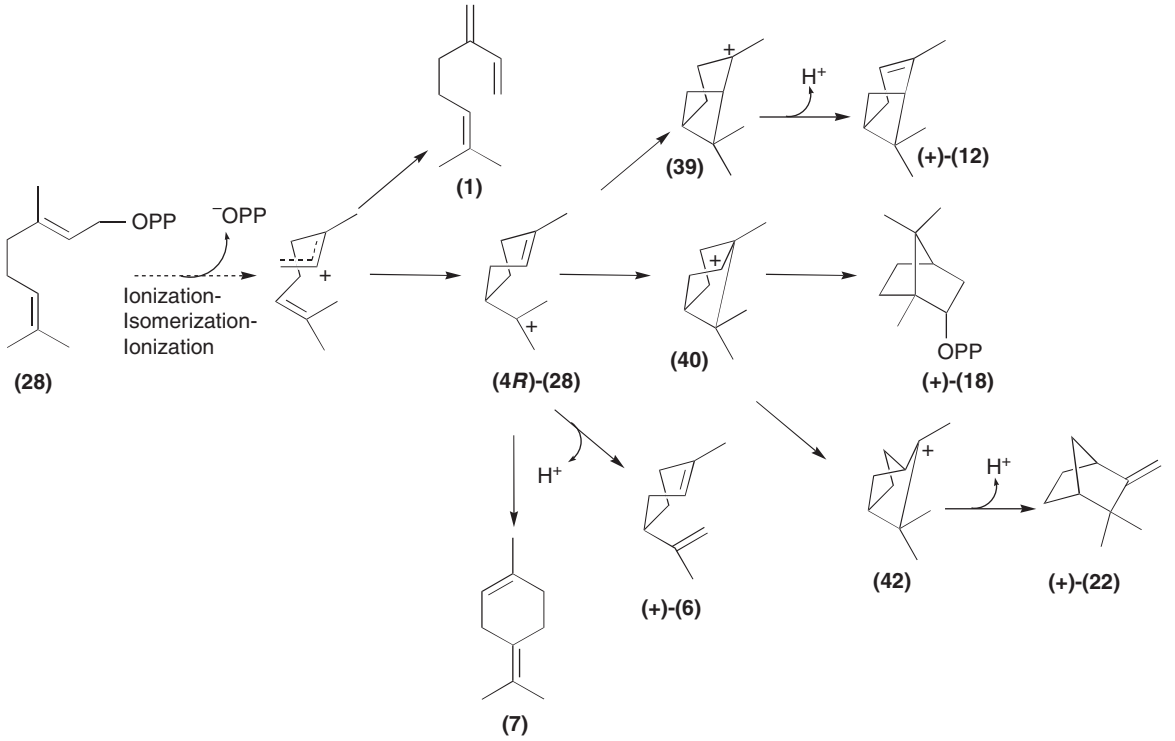
Studies using [8-^{2}H]-GPP and [9-^{2}H]-GPP, coupled with NMR analysis of products, were able to localize the position of the *trans*-methyl of GPP as the *syn* geminal methyl of BPP, thus demonstrating that bicyclization occurs on the same face of the C6–C7 bond of GPP on which primary C1–C6 cyclization had occurred.[106] The studies outlined above have established considerable details regarding different aspects of the (+)-BS reaction cascade and provide a stereochemical model of bicyclization and anion capture (**Scheme 13**).

(+)-BS was cloned and characterized from *Salvia officinalis* and it is quite similar to (−)-LS in physicochemical properties.[89] Based on the studies described above for (−)-LS, (+)-BS localization to the plastid is presumed. Poor expression of the soluble preprotein, and the presence of N-terminal tandem arginines, led to the production of a 'mature' recombinant cyclase truncated immediately upstream of this motif. This pseudomature form resembles the native enzyme in all properties evaluated including product profile (**Scheme 14**).[89]

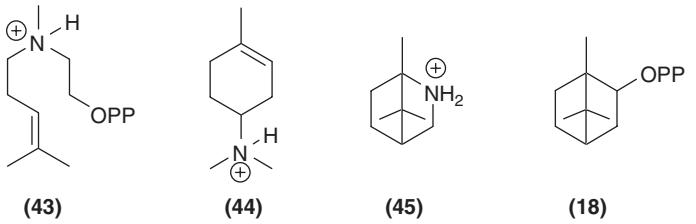
(+)-BS was the first monoterpene cyclase for which structures were determined.[73] Judicious selection of analogs that mimic all intermediates and the product of the reaction (**Scheme 15**) has provided structures for which binding interactions during the isomerization–cyclization cascade can be inferred.[73] In general, results obtained were consistent with the plant cyclase structures described above. The structure of the native (+)-BS (without analog) had considerable disorder in the N-terminal loop and two active site regions, one that includes part of the DTE motif (active site lip) and the other that is deep within the prenyl-binding cavity (F578-S583). Analog or product binding promotes significant conformational changes that result in ordering through two significant interdomain interactions (involving hydrogen bonds between the penultimate R (of the tandem R-motif) and a downstream Y (of the N-terminal loop) with the first and third D of the DDxxD motif,



Scheme 13



Scheme 14



Scheme 15

respectively) that cap the active site cavity. The first and third D residues also participate in Mg^{2+} coordination and coordinate a tightly held water; the second D (of the DDxxD motif) is salt-linked to R314 of the RxR motif. The conspicuous presence of a water molecule near the active site lip is consistent with monoterpenol production by (+)-BS observed during turnover of the noncyclizable analogs 2,3-cyclopropylGPP (**34**) and 6,7-dihydroGPP (**35**), as discussed earlier (**Scheme 7**). In the normal reaction cascade of (+)-BS, no monoterpenols are formed and this suggests that the tightly held water performs a structural or catalytic role.[73] The orientation of the diphosphate in all of the structures is largely unchanged, and this group is held in position by Mg^{2+} coordination and three hydrogen bonds (residues R493, R314, and K512), thereby suggesting that the requisite and precise repositioning of the diphosphate observed during the reaction cycle (C1–C3–C2 migration) is facilitated by corresponding movements of the prenyl carbocations. This suggestion is supported by the structure with the 2-aza-bornane analog (**45**, bornyl cation mimic), which is positioned with the charged aza group within 3 Å of the oxygen that would become the phosphoester oxygen of product (described above). Interestingly, the 7-aza-7,8-dihydrolimonene analog (**44**) (an α-terpinyl cation mimic) bound upside down, likely an indication that cation interactions with the diphosphate counterion are important. The unproductive binding of this analog emphasizes the difficulties that may be encountered when intermediates are presented from solution rather than generated at the active site during which numerous protein conformational changes

may occur. This phenomenon is reminiscent of (−)-LS binding exogenous 2-F-LPP (**33**) in a helical, cisoid conformer versus the extended, transoid orientation adopted by **33** produced at the active site by isomerization of 2-F-GPP (**32**). A similar phenomenon was observed with the sage PSs, which generated altered product mixtures when presented with either LPP enantiomer from solution (compared to GPP-generated product profiles).[107]

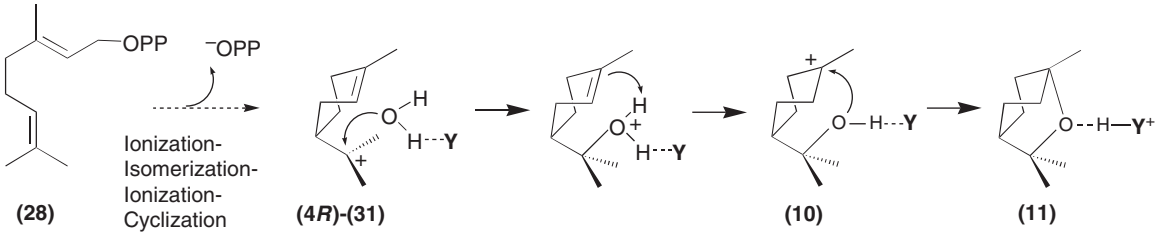
The prenyl group of the analogs participates in van der Waals forces with three hydrophobic residues in the active site pocket, W323, I344, and F578. Such aromatic residues are similarly positioned in (−)-LS, and likely contribute, in both cases, to α-terpinyl carbocation (**31**) stabilization through π-cation interactions. Although termination by anion capture is regio- and stereochemically very precise, (+)-BS is moderately promiscuous during 25% of its reaction cycles in producing olefin co-products. Thus, orchestrated 'tumbling and folding' of the olefinic chain within the active site must occur to allow rearrangements before deprotonation from at least six different carbons (**Scheme 14**).

1.15.3.4 1,8-Cineole Synthase

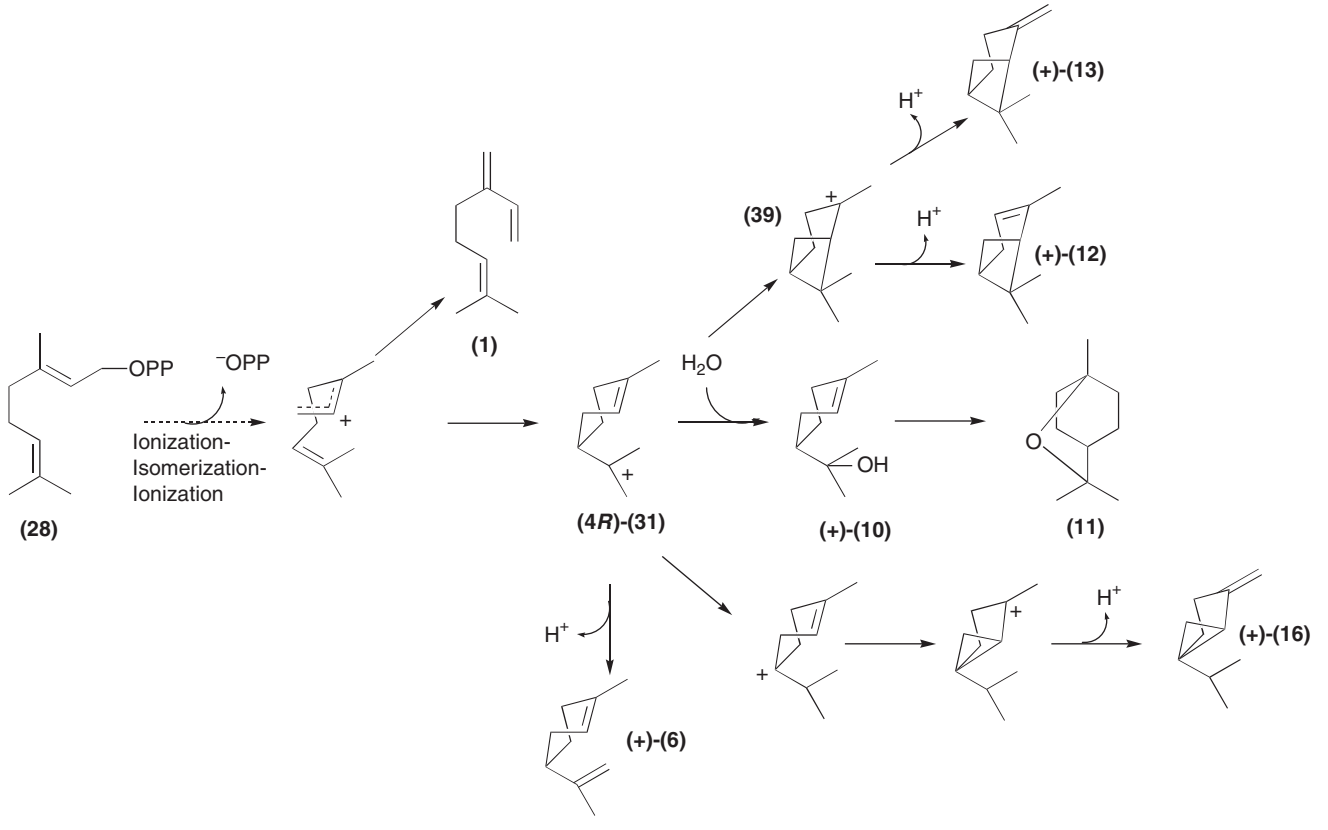
1,8-Cineole (**11**, eucalyptol) is an achiral aromatic component of many plants, including *Salvia* and *Eucalyptus* leaves, has recently been identified as a rhizosphere volatile in *Arabidopsis*,[108] is well known as an allelopathic agent from *Salvia* and *Artemisia*,[109,110] and has folivore deterrent activity.[111] The generation of **11** by the namesake cyclase is atypical in that a second, stable (and chiral) intermediate (α-terpineol (**10**)) is formed at the active site and requires protonation of the endocyclic double bond to proceed to end product. Stereochemical studies with the 1,8-CS from *S. officinalis* demonstrated a kinetic preference of (*R*)-LPP over (*S*)-LPP, thus supporting the intermediacy of the (4*R*)-α-terpinyl cation and (4*R*)-(**10**).[112] Confirmation of (4*R*) stereochemistry was achieved by enzymatic cyclization of [1-^{3}H]-GPP with chemical conversion of the resultant (1*R*)-[3-^{3}H]-1,8-cineole to the corresponding butanediol ketal for stereochemical assignment.[112]

The origin of the ether oxygen was evaluated using [1-^{18}O]-GPP, [3-^{18}O]-LPP, and $H_2^{18}O$ in three separate experiments, and significant enrichment in the bridge oxygen occurred only with water as the nucleophile.[112] Recombinant enzyme (provided from an *S. officinalis* clone[89]) provided the means to generate sufficient product for NMR analysis, and the *syn* stereochemical disposition of water capture by the α-terpinyl cation (**31**) (addition on the inside face) was thus verified using [1,1,8,8,8-$^{2}H_5$]-GPP and [1,1,9,9,9-$^{2}H_5$]-GPP as substrates.[113] Enzymatic cyclization in the presence of D_2O and NMR analysis of the product confirmed *syn* addition at C6 (of the endocyclic double bond) and, thus, suprafacial ether bridge formation.[113] This work has firmly established the nature of cyclization of GPP to the achiral product by CS from *S. officinalis* (**Scheme 16**).

CS was cloned and characterized from *Salvia fruticosa*, and a pseudomature form of this cyclase (presumed transit peptide removed) was utilized for crystallization.[78] A single structure was obtained that displayed typical TPS features as described for (+)-BS and (−)-LS, and displayed only minor backbone differences with these cyclases in the active site region. The enzyme had a disordered N-terminal loop and two disordered C-terminal regions, similar to the (+)-BS native structure, which is likely a consequence of assuming the 'unbound conformation'. It should be noted that the structural studies (described below) and the mechanistic studies (described above) are from two closely related CSs that share 95% amino acid identity and have similar reaction profiles (**Scheme 17**). However, relatedness and high homology do not imply similar stereochemistry,



Scheme 16


OPP
$^{-}$OPP
Ionization-
Isomerization-
Ionization
(28)
(1)
(4R)-(31)
H_2O
(39)
(+)-(13)
(+)-(12)
(+)-(10)
(11)
(+)-(6)
(+)-(16)
H^+

Scheme 17

a situation encountered with two enantiomeric maize sesquiterpene cyclases.[114] Nonetheless, the mechanistic studies provide ample detail for this enzyme type and provide enhanced context of this structural study.

Sequence alignments of the related *Salvia* cyclases (two sabinene synthases (SSs), two CSs, and (+)-BS) coupled with careful examination of the active site cavity identified an active site asparagine, located near the bottom of the hydrophobic cavity, that affects product outcome.[78] This residue, N338, coordinates an active site water in the CS structure (**Scheme 16**, Y = N338) that is notably absent in the (+)-BS (different location from the water discussed above in the (+)-BS structure) and (−)-LS structures. The N338I mutant (CS scaffold and SS residue) successfully eliminated all cation capture by water and increased sabinene (**16**) production by 15-fold relative to wild-type CS activity (with minimal kinetic effects); the reciprocal change in SS had little effect on product generation. Additional mutagenesis of the neighboring residue (and residues on the opposing face of the active site cavity) provided mild synergistic effects on the CS-to-SS conversion but had little effect on the reverse conversion. Further substitutions at N338 (L, V, A, C, and S) had pronounced effects on product ratios of the CS, with only small changes in catalytic efficiency.[78] Sesquiterpene synthase activity (normally not present in wild-type CS) was observed in three of the mutants (A, C, and S), consistent with the smaller overall side-chain size of the substitutions (relative to N), the presumed absence of water, and with the position near the bottom of the prenyl-binding cavity. The C15 products were largely analogous in structure to the C10 products generated by these mutants, and this suggests that the reaction cascades were similar, irrespective of chain length, such that the 'extra' five-carbon tail was not involved in the reaction cascade but was accommodated by the larger active site cavity.

The near-complete CS-to-SS conversion is significant and is entirely consistent with the role of N338 in positioning a water as the α-terpinyl cation (**31**) capture nucleophile. The lack of reciprocity in the SS-to-CS conversion suggests that the active site structure of SS is quite different, and the role of this N in the SS mutant must be considered in the context of the surrounding cavity. The function of a similarly placed N in the olefinic cyclase (−)-LS is certainly also different from CS. The subtlety of active site structure–function relationships was demonstrated in directed mutagenesis studies performed on two sets of closely related conifer monoterpene cyclases that showed product outcome to require the collective interaction of multiple amino acid changes.[115,116] Clearly, the relationship between structure and function in the terpene cyclases is often difficult to assess; the reaction sequences are complex and defining the role of active site residues in the reaction cascades is rarely as straightforward as in the case of CS.

1.15.4 Conclusions

Notable progress has been made in the past decade in elucidating monoterpene cyclase reaction stereochemistry and mechanism. Molecular biology has greatly expanded the availability of, and access to, scores of terpene cyclases, including monoterpene cyclases that produce most structural types. Crystal structures have been determined for three catalytically unique monoterpene cyclases, and several important mechanistic features of these enzymes have been confirmed by this structural work. Progress made during the past decade is substantial and informative, and it is clear that structural analysis of cyclases bound with nonreactive analogs has been particularly fruitful when coupled with structure-guided site-directed mutagenesis. Acquiring structures of alternate cyclases of multiple types and of different stereochemistries is a priority for the future. When coupled with mechanistic, biochemical, and mutagenic studies of these enzymes, structural evaluation can provide enormous insight into monoterpene cyclization cascades of considerable complexity.

Acknowledgments

I would like to thank Professor Rodney Croteau for critical reading of this manuscript and for helpful comments. Financial support was provided by the NIH.

References

1. J. Gershenzon; N. Dudareva, *Nat. Chem. Biol.* **2007**, *3*, 408.
2. J. B. Harborne, Recent Advances in the Ecological Chemistry of Plant Terpenoids. In *Ecological Chemistry and Biochemistry of Plant Terpenoids*; J. B. Harborne, F. A. Tomas-Barberan, Eds.; Clarendon: Oxford, 1991; Vol. 31, p 399.
3. J. Gershenzon; R. Croteau, Terpenoids. In Herbivores: *Their Interaction with Secondary Metabolites*; G. A. Rosenthal, M. Berenbaum, Eds.; Academic Press: New York, 1991; p 165.
4. M. A. Phillips; R. Croteau, *Trend Plant Sci.* **1999**, *4*, 184.
5. R. Croteau; E. M. Davis; K. L. Ringer; M. R. Wildung, *Naturwissenschaften* **2005**, *92*, 562.
6. R. Croteau; C. L. Hooper, *Plant Physiol.* **1978**, *61*, 737.
7. R. Croteau; C. Martinkus, *Plant Physiol.* **1979**, *64*, 169.
8. R. Croteau, *Chem. Rev.* **1987**, *87*, 929.
9. M. L. Wise; R. Croteau, Monoterpene Biosynthesis. In *Comprehensive Natural Products Chemistry: Isoprenoids Including Steroids and Carotenoids*; D. E. Cane, Ed.; Elsevier Science: Oxford, 1999; Vol. 2, p 97.
10. E. M. Davis; R. Croteau, Cyclization Enzymes in the Biosynthesis of Monoterpenes, Sesquiterpenes and Diterpenes. In *Topics in Current Chemistry: Biosynthesis – Aromatic Polyketides, Isoprenoids, Alkaloids*; F. Leeper, J. C. Vederas, Eds.; Springer-Verlag: Heidelberg, Germany, 2000; Vol. 209, p 53.
11. D. B. Little; R. Croteau, Biochemistry of Essential Oil Terpenes. A Thirty Year Overview. In *Flavor Chemistry: Thirty Years of Progress*; I. Hornstein, E. L. Wick, R. Teranashi, Eds.; Kluwer Academic, Plenum Press: New York, 1999; p 239.
12. R. Croteau; T. M. Kutchan; N. G. Lewis, Natural Products (Secondary Metabolites). In *Biochemistry and Molecular Biology of Plants*; B. B. Buchanan, W. Gruissem, R. L. Jones, Eds.; American Society of Plant Physiologists: Rockville, MD, 2000; p 1250.
13. T. Koyama; K. Ogura, Isopentenyl Diphosphate Isomerase and Prenyltranferases. In *Comprehensive Natural Products Chemistry: Isoprenoids Including Steroids and Carotenoids*; D. E. Cane, Ed.; Elsevier Science: Oxford, 1999; Vol. 2, p 69.
14. C. D. Poulter; H. C. Rilling, Prenyl Transferases and Isomerase. In *Biosynthesis of Isoprenoid Compounds*; J. W. Porter, S. L. Spurgeon, Eds.; Wiley: New York, 1981; Vol. 1, p 162.
15. H. V. Thulasiram; H. K. Erickson; C. D. Poulter, *Science* **2007**, *316*, 73.
16. D. W. Christianson, *Chem. Rev.* **2006**, *106*, 3412.
17. R. Croteau; D. E. Cane, Monoterpene and Sesquiterpene Cyclases. In *Methods Enzymology*; J. H. Law, H. C. Rilling, eds.; Academic Press: New York, 1985; Vol. 110, p 383.
18. A. Saito; H. C. Rilling, *Arch. Biochem. Biophys.* **1981**, *208*, 508.
19. V. J. Davisson; T. R. Neal; C. D. Poulter, *J. Am. Chem. Soc.* **1985**, *107*, 5277.
20. A. B. Gilg; C. Tittiger; G. J. Blomquist, *Naturwissenschaften* **2009**, *96*, 731.
21. S. Green; E. N. Friel; A. Matich; L. L. Beuning; J. M. Cooney; D. D. Rowan; E. MacRae, *Phytochemistry* **2007**, *68*, 176.
22. D. E. Cane, Sesquiterpene Biosynthesis: Cyclization Mechanisms. In *Comprehensive Natural Products Chemistry: Isoprenoids Including Steroids and Carotenoids*; D. E. Cane, Ed.; Elsevier Science: Oxford, 1999; Vol. 2, p 155.
23. M. A. Phillips; M. R. Wildung; D. C. Williams; D. C. Hyatt; R. Croteau, *Arch. Biochem. Biophys.* **2003**, *411*, 267.
24. C. G. Jones; C. I. Keeling; E. L. Ghisalberti; E. L. Barbour; J. A. Plummer; J. Bohlmann, *Arch. Biochem. Biophys.* **2008**, *477*, 121.
25. R. Croteau; F. Karp, *Arch. Biochem. Biophys.* **1977**, *179*, 257.
26. R. Croteau; F. Karp, *Arch. Bioch. Biophys.* **1979**, *198*, 512.
27. R. Croteau; M. Felton; R. C. Ronald, *Arch. Bioch. Biophys.* **1980**, *200*, 534.
28. H. Gambliel; R. Croteau, *J. Biol. Chem.* **1982**, *257*, 2335.
29. R. Kjonaas; R. Croteau, *Arch. Bioch. Biophys.* **1983**, *220*, 79.
30. R. Croteau; F. Karp, *Arch. Bioch. Biophys.* **1979**, *198*, 523.
31. A. L. Schilmiller; I. Schauvinhold; M. Larson; R. Xu; A. L. Charbonneau; A. Schmidt; C. Wilkerson; R. L. Last; E. Pichersky, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 10865.
32. C. Sallaud; D. Rontein; S. Onillon; F. Jabès; P. Duffé; C. Giacalone; S. Thoraval; C. Escoffier; G. Herbette; M. Leonhardt; A. Causse; A. Tissier, *Plant Cell* **2009**, *21*, 301.
33. C. C. Burke; M. R. Wildung; R. Croteau, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13062.
34. F. Bouvier; C. Suire; A. d'Harlingue; R. A. Backhaus; B. Camara, *Plant J.* **2000**, *24*, 241.
35. C. Burke; R. Croteau, *Arch. Biochem. Biophys.* **2002**, *405*, 130.
36. D. Tholl; C. M. Kish; I. Orlova; D. Sherman; J. Gershenzon; E. Pichersky; N. Dudareva, *Plant Cell* **2004**, *16*, 977.
37. Y.-Y. Hsiao; M.-F. Jeng; W.-C. Tsai; Y.-C. Chuang; C.-Y. Li; T.-S. Wu; C.-S. Kuoh; W.-H. Chen; H.-H. Chen, *Plant J.* **2008**, *55*, 719.
38. A. Schmidt; J. Gershenzon, *Phytochemistry* **2008**, *69*, 49.
39. G. Wang; R. A. Dixon, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 9914.
40. K. P. Adams; R. Croteau, *Phytochemistry* **1998**, *49*, 475.
41. K. P. Adams; J. E. Crock; R. Croteau, *Arch. Biochem. Biophys.* **1996**, *332*, 352.
42. M. Wise; G. L. Rorrer; J. J. Polzin; R. B. Croteau, *Arch. Biochem. Biophys.* **2002**, *400*, 125.
43. C.-M. Wang; D. E. Cane, *J. Am. Chem. Soc.* **2008**, *130*, 8908.
44. M. Komatsu; M. Tsuda; S. Omura; H. Oikawa; H. Ikeda, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 7422.
45. T. W. Hallahan; R. Croteau, *Arch. Biochem. Biophys.* **1988**, *264*, 618.
46. R. Croteau; N. M. Felton; C. J. Wheeler, *J. Biol. Chem.* **1985**, *260*, 5956.
47. D. M. Satterwhite; C. J. Wheeler; R. Croteau, *J. Biol. Chem.* **1985**, *260*, 13901.
48. T. W. Hallahan; R. Croteau, *Arch. Biochem. Biophys.* **1989**, *269*, 313.
49. T. J. Savage; R. Croteau, *Arch. Biochem. Biophys.* **1993**, *305*, 581.
50. G. Portilla; M. C. Rojas; L. Chayet; O. Cori, *Arch. Biochem. Biophys.* **1982**, *218*, 614.
51. L. Chayet; M. C. Rojas; E. Cardemil; A. M. Jabalquinto; R. Vicuña; O. Cori, *Arch. Biochem. Biophys.* **1977**, *180*, 318.
52. R. Croteau, *Arch. Biochem. Biophys.* **1986**, *251*, 777.
53. F. Karp; R. Croteau, *Arch. Biochem. Biophys.* **1994**, *309*, 184.

54. C. D. Poulter; J. C. Argyle; E. A. Mash, *J. Biol. Chem.* **1978**, *253*, 7227.
55. C. D. Poulter; D. M. Satterwhite, *Biochemistry* **1977**, *16*, 5470.
56. F. Karp; Y. Zhao; B. Santhamma; B. Assink; R. M. Coates; R. B. Croteau, *Arch. Biochem. Biophys.* **2007**, *468*, 140.
57. C. J. Wheeler; R. Croteau, *Arch. Biochem. Biophys.* **1986**, *246*, 733.
58. C. J. Wheeler; R. Croteau, *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4856.
59. W. Schwab; D. C. Williams; E. M. Davis; R. Croteau, R, *Arch. Biochem. Biophys.* **2001**, *392*, 123.
60. D. C. Hyatt; B. Youn; Y. Zhao; B. Santhamma; R. M. Coates; R. B. Croteau; C. H. Kang, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104,* 5360.
61. R. Croteau; D. M. Satterwhite; D. E. Cane; C. C. Chang, *J. Biol. Chem.* **1986**, *261*, 13438.
62. R. Croteau; D. M. Satterwhite; C. J. Wheeler; N. M. Felton, *J. Biol. Chem.* **1989**, *264*, 2075.
63. R. Croteau; D. M. Satterwhite; D. E. Cane; C. C. Chang, *J. Biol. Chem.* **1988**, *263*, 10063.
64. S. Aubourg; A. Lecharny; J. Bohlmann, *Mol. Genet. Genomics* **2002**, *267*, 730.
65. G. A. Tuskan; S. Difazio; S. Jansson; J. Bohlmann; I. Grigoriev; U. Hellsten; N. Putnam; S. Ralph; S. Rombauts; A. Salamov; J. Schein; L. Sterck; A. Aerts; R. R. Bhalerao; R. P. Bhalerao; D. Blaudez; W. Boerjan; A. Brun; A. Brunner; V. Busov; M. Campbell; J. Carlson; M. Chalot; J. Chapman; G. L. Chen; D. Cooper; P. M. Coutinho; J. Couturier; S. Covert; Q. Cronk; R. Cunningham; J. Davis; S. Degroeve; A. Déjardin; C. Depamphilis; J. Detter; B. Dirks; I. Dubchak; S. Duplessis; J. Ehlting; B. Ellis; K. Gendler; D. Goodstein; M. Gribskov; J. Grimwood; A. Groover; L. Gunter; B. Hamberger; B. Heinze; Y. Helariutta; B. Henrissat; D. Holligan; R. Holt; W. Huang; N. Islam-Faridi; S. Jones; M. Jones-Rhoades; R. Jorgensen; C. Joshi; J. Kangasjärvi; J. Karlsson; C. Kelleher; R. Kirkpatrick; M. Kirst; A. Kohler; U. Kalluri; F. Larimer; J. Leebens-Mack; J. C. Leplé; P. Locascio; Y. Lou; S. Lucas; F. Martin; B. Montanini; C. Napoli; D. R. Nelson; C. Nelson; K. Nieminen; O. Nilsson; V. Pereda; G. Peter; R. Philippe; G. Pilate; A. Poliakov; J. Razumovskaya; P. Richardson; C. Rinaldi; K. Ritland; P. Rouzé; D. Ryaboy; J. Schmutz; J. Schrader; B. Segerman; H. Shin; A. Siddiqui; F. Sterky; A. Terry; C. J. Tsai; E. Uberbacher; P. Unneberg; J. Vahala; K. Wall; S. Wessler; G. Yang; T. Yin; C. Douglas; M. Marra; G. Sandberg; Y. Van de Peer; D. Rokhsar, *Science* **2006**, *313*, 1596.
66. O. Jaillon; J. M. Aury; B. Noel; A. Policriti; C. Clepet; A. Casagrande; N. Choisne; S. Aubourg; N. Vitulo; C. Jubin; A. Vezzi; F. Legeai; P. Hugueney; C. Dasilva; D. Horner; E. Mica; D. Jublot; J. Poulain; C. Bruyère; A. Billault; B. Segurens; M. Gouyvenoux; E. Ugarte; F. Cattonaro; V. Anthouard; V. Vico; C. Del Fabbro; M. Alaux; G. Di Gaspero; V. Dumas; N. Felice; S. Paillard; I. Juman; M. Moroldo; S. Scalabrin; A. Canaguier; I. Le Clainche; G. Malacrida; E. Durand; G. Pesole; V. Laucou; P. Chatelet; D. Merdinoglu; M. Delledonne; M. Pezzotti; A. Lecharny; C. Scarpelli; F. Artiguenave; M. E. Pè; G. Valle; M. Morgante; M. Caboche; A. F. Adam-Blondon; J. Weissenbach; F. Quétier; P. Wincker; *Nature* **2007**, *449*, 463.
67. J. Bohlmann; G. Meyer-Gauen; R. Croteau, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4126.
68. S. Aubourg; A. Lecharny; J. Bohlmann, *Mol. Genet. Genomics* **2002**, *267*, 730.
69. D. M. Martin; J. Fäldt; J. Bohlmann, *Plant Physiol.* **2004**, *135*, 1908.
70. S. C. Trapp; R. B. Croteau, *Genetics* **2001**, *158*, 811.
71. B. S. Lee; J. Chappell, *Plant Physiol.* **2008**, *147*, 1017.
72. K. U. Wendt; G. E. Schulz, *Structure* **1998**, *6*, 127.
73. D. A. Whittington; M. L. Wise; M. Urbansky; R. M. Coates; R. B. Croteau; D. W. Christianson, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15375.
74. L. C. Tarshis; M. Lan; C. D. Poulter; J. C. Sacchettini, *Biochemistry* **1994**, *33*, 10871.
75. C. A. Lesburg; G. Zhai; D. E. Cane; D. W. Christianson, *Science* **1997**, *277*, 1820.
76. R. J. Peters; M. M. Ravn; R. M. Coates; R. B. Croteau, *J. Amer. Chem. Soc.* **2001**, *123*, 8974.
77. C. M. Starks; K. Back; J. Chappell; J. P. Noel, *Science* **1997**, *277*, 1815.
78. S. C. Kampranis; D. Ioannidis; A. Purvis; W. Mahrez; E. Ninga; N. A. Katerelos; S. Anssour; J. M. Dunwell; J. Degenhardt; A. M. Makris; P. W. Goodenough; C. B. Johnson, *Plant Cell* **2007**, *19*, 1994.
79. M. K. El Tamer; J. Lücker; D. Bosch; H. A. Verhoeven; F. W. A. Verstappen; W. Schwab; A. J. van Tunen; A. G. J. Voragen; R. A. de Maagd; H. J. Bouwmeester, *Arch. Biochem. Biophys.* **2003**, *411*, 196.
80. R. J. Peters; R. B. Croteau, *Arch. Biochem. Biophys.* **2003**, *417*, 203.
81. K. Back; J. Chappell, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6841.
82. E. Y. Shishova; F. Yu; D. J. Miller; J. A. Faraldos; Y. Zhao; R. M. Coates; R. K. Allemann; D. E. Cane; D. W. Christianson, *J. Biol. Chem.* **2008**, *283*, 15431.
83. J. Bohlmann; M. Phillips; V. Ramachandiran; S. Katoh; R. Croteau, *Arch. Biochem. Biophys.* **1999**, *368*, 232.
84. D. M. Martin; J. Fäldt; J. Bohlmann, *Plant Physiol.* **2004**, *136*, 3724.
85. J. Bohlmann; C. L. Steele; R. Croteau, *J. Biol. Chem.* **1997**, *272*, 21784.
86. J. Fäldt; D. M. Martin; B. Miller; S. Rawat; J. Bohlmann, *Plant Mol. Biol. Rep.* **2003**, *51*, 119.
87. J. Lucker; M. K. El Tamer; W. Schwab; F. W. A. Verstappen; L. H. W. van der Plas; H. J. Bouwmeester; H. A. Verhoeven, *Eur. J. Biochem.* **2002**, *269*, 3160.
88. D. M. Martin; J. Bohlmann, *Phytochemistry* **2004**, *65*, 1223.
89. M. L. Wise; T. J. Savage; E. Katahira; R. Croteau, *J. Biol. Chem.* **1998**, *273*, 14891.
90. Y. Iijima; R. Davidovich-Rikanati; E. Fridman; D. R. Gang; E. Bar; E. Lewinsohn; E. Pichersky, *Plant Physiol.* **2004**, *136*, 3724.
91. S. M. Colby; W. R. Alonso; E. J. Katahira; D. J. McGarvey; R. Croteau, *J. Biol. Chem.* **1993**, *268*, 23016.
92. S. Green; C. J. Squire; N. J. Nieuwenhuizen; E. N. Baker; W. Laing, *J. Biol. Chem.* **2009**, *284*, 8661.
93. T. Suga; Y. Hiraga; W. A. Mie; S. Isumi, *J. Chem. Soc. Chem. Commun.* **1992**, 1556.
94. Y. Hiraga; W. Shi; D. I. Ito; S. Ohta; T. Suga, *J. Chem. Soc. Chem. Commun.* **1993**, 1370.
95. H.-J. Pyun; R. M. Coates; K. C. Wagschal; P. McGeady; R. B. Croteau, *J. Biol. Chem.* **1993**, *58*, 3998.
96. M. F. Leopold; W. W. Epstein; D. M. Grant, *J. Am. Chem. Soc.* **1988**, *110*, 616.
97. J. I. M. Rajaonarivony; J. Gershenzon; R. Croteau, *Arch. Biochem. Biophys.* **1992**, *296*, 49.
98. W. R. Alonso; J. I. M. Rajaonarivony; J. Gershenzon; R. Croteau, *J. Biol. Chem.* **1992**, *267*, 7582.
99. G. Turner; J. Gershenzon; E. E. Nielson; J. E. Froehlich; R. Croteau, *Plant Physiol.* **1999**, *120*, 879.
100. D. C. Williams; D. J. McGarvey; E. J. Katahira; R. Croteau, *Biochemistry* **1998**, *37*, 12213.

101. J. I. M. Rajaonarivony; J. Gershenzon; J. Miyazaki; R. Croteau, *Arch. Biochem. Biophys.* **1992**, *299*, 77.
102. R. Croteau; J. J. Shaskus; B. Renstrom; N. M. Felton; D. E. Cane; A. Saito; C. Chang, *Biochemistry* **1985**, *24*, 7077.
103. D. E. Cane; A. Saito; R. Croteau; J. Shaskus; N. M. Felton, *J. Am. Chem. Soc.* **1982**, *104*, 5831.
104. K. C. Wagschal; H.-J. Pyun; R. M. Coates; R. Croteau, *Arch. Biochem. Biophys.* **1994**, *308*, 477.
105. R. Croteau; C. J. Wheeler; D. E. Cane; R. Ebert; H.-J. Ha, *Biochemistry* **1987**, *26*, 5383.
106. M. L. Wise; H.-J. Pyun; G. Helms; B. Assink; R. M. Coates; R. B. Croteau, *Tetrahedron* **2001**, *57*, 5327.
107. R. Croteau; D. M. Satterwhite, *J. Biol. Chem.* **1989**, *264*, 15309.
108. F. Chen; D.-K. Ro; J. Petri; J. Gershenzon; J. Bohlmann; E. Pichersky; D. Tholl, *Plant Physiol.* **2004**, *135*, 1956.
109. J. G. Romagni; S. N. Allen; F. E. Dayan, *J. Chem. Ecol.* **2000**, *26*, 303.
110. J. P. Halligan, *Ecology* **1975**, *56*, 999.
111. I. R. Lawler; J. Stapley; W. J. Foley; B. M. Eschler, *J. Chem. Ecol.* **1999**, *25*, 401.
112. R. Croteau; W. R. Alonso; A. E. Koepp; M. A. Johnson, *Arch. Biochem. Biophys.* **1994**, *309*, 184.
113. M. L. Wise; M. Urbansky; G. L. Helms; R. M. Coates; R. Croteau, *J. Am. Chem. Soc.* **2002**, *124*, 8546.
114. T. G. Kolner; C. Schnee; J. Gershenzon; J. Degenhardt, *Plant Cell* **2004**, *16*, 1115.
115. D. C. Hyatt; R. Croteau, *Arch. Biochem. Biophys.* **2005**, *439*, 222.
116. S. Katoh; D. Hyatt; R. Croteau, *Arch. Biochem. Biophys.* **2004**, *425*, 65.

Biographical Sketch

Edward M. Davis was born in Buffalo, NY, but spent his youth in Michigan City, IN. He is currently an assistant scientist in the laboratory of Professor Rodney Croteau (Institute of Biological Chemistry, Washington State University) working on monoterpene biosynthesis in mint and the structure–function relationship in monoterpene cyclases. He earned his B.S. in 1987 (Biological Sciences) and his Ph.D. in 1999 (Biochemistry and Molecular Biology) from Oklahoma State University in Stillwater, OK. He did his doctoral work with Dr. Margaret Essenberg where he studied a sesquiterpene cyclase that catalyzes the first committed step in phytoalexin production in *Gossypium hirsutum* (cotton).

1.16 Sesquiterpenes

Joe Chappell, University of Kentucky, Lexington, KY, USA

Robert M. Coates, University of Illinois, Urbana, IL, USA

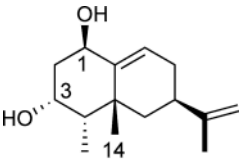
© 2010 Elsevier Ltd. All rights reserved.

1.16.1	**Introduction**	609
1.16.2	**Classification of Sesquiterpenes**	610
1.16.2.1	Carotenoid Cleavage Products	610
1.16.2.2	Farnesanes	611
1.16.2.3	Drimanes	613
1.16.2.4	Cyclic Sesquiterpenes	614
1.16.2.4.1	The bisabolane series	617
1.16.2.4.2	The cuparane series	619
1.16.2.4.3	The cadinane series	622
1.16.2.4.4	The humulane series	624
1.16.2.4.5	The germacrane series	626
1.16.3	**Decorating the Sesquiterpene Scaffolds**	630
1.16.4	**Future Challenges**	635
References		637

1.16.1 Introduction

Sesquiterpenes, like so many other classes of terpenes and other natural products, have attracted significant interest because of the roles they play in biological systems and their utility for human uses. However, one of the features that serves to distinguish sesquiterpenes from many of the other families of natural products is the thousands of different sesquiterpene compounds that have been identified. The source of this diversity arises in two ways. First is the structural diversity that arises in the assembly of the 15-carbon skeletons making up the backbone of all sesquiterpenes. The second source of diversity is due to the layering of functional groups and substituents upon the structural scaffolds in distinct regio- and stereospecific manners. Hence, while some recent reviews of sesquiterpenes have focused on the number of distinct and new sesquiterpenes identified,[1] the current chapter will initially discuss theoretical considerations for how the sesquiterpene scaffolds arise, then attempt to rationalize this information with what is now known about the enzymes mediating these key biosynthetic reactions in a manner similar to the mechanistic approaches of Cane,[2,3] Hohn,[4] and Christianson.[5] We will then turn our attention to emerging details about the biochemical mechanisms responsible for elaborating the sesquiterpene scaffolds with substituent and functional groups. Lastly, we will raise the question of how well do we understand sesquiterpene biosynthesis. That is, do we understand the enzyme mechanisms and the physiological context of how these compounds are made in cells to actually engineer the biosynthesis of novel or unique sesquiterpene compounds?

The structural complexity of sesquiterpenes is readily illustrated by what initially appears to be a simple structure. Capsidiol (**1**), an eremophilane-type sesquiterpenoid first isolated from pepper,[6] *Capsicum annum*, contains two hydroxyl substituents. Hence, the trivial name capsidiol. However, more important to note is the complexity inherent in this structure. For a 15-carbon or C_{15} compound, capsidiol contains five chiral centers and three different substituents (hydroxyl, methyl and isopropenyl groups) arrayed precisely (regio- and stereospecifically) around a bicyclic scaffold. What we intend to suggest in this chapter is that if we can truly understand the mechanisms responsible for the biosynthesis of molecules like capsidiol, then we should be able to modify these enzymes so they, for example, yield the inverted stereochemistry of the C14 methyl substituent, or alter the regio- or stereochemical positioning of the hydroxyl at C1, thus creating novel molecules, which may have benefits to the organisms producing them as well as man-made applications.



Capsidiol (**1**)

1.16.2 Classification of Sesquiterpenes

1.16.2.1 Carotenoid Cleavage Products

The issue of how to classify sesquiterpenes is not as straightforward as one might imagine. While the majority of sesquiterpenes arise directly from farnesyl diphosphate (FPP), a key C_{15} diphosphorylated intermediate of the mevalonate (MVA) biosynthetic pathway, a number of biologically important sesquiterpenes arise as breakdown products from terpenes synthesized by the nonmevalonate or methylerythritol phosphate (MEP) pathway, which in plants operates simultaneously and in parallel to the MVA pathway except in the plastid compartment (**Figure 1**). Perhaps the most familiar of these carotenoid breakdown products is abscisic acid (ABA, **4**, **Scheme 1**), a well-known plant growth regulator.[7] An important discovery in elucidating the biosynthetic pathway for ABA was the functional mapping of the viviparous mutants of maize. The *vp* mutants and especially *vp*14 were distinguished by having corn kernels that sprouted prematurely during corn cob development.[8] The premature sprouting was subsequently correlated with the developing seeds having abnormally low levels of ABA, which typically arrests premature seed germination.

Functional characterization of the wild-type VP14 locus demonstrated that it encoded for a carotenoid cleavage dioxygenase (CCD) that cleaves 9-*cis*-neoxanthoxin or 9-*cis*-violaxanthoxin (**2**) at the C11═C12 double bond, introducing oxygens from molecular O_2 into both cleavage products and liberating xanthoxin (**3**), which undergoes successive oxidations and an epoxide ring opening to yield ABA (**4**, **Scheme 1**).[9] The CCD genes are now well recognized as part of a much larger and diverse gene family, widely distributed in bacteria, plants, and animals, and characterized for their specificity of cleavage of various carotenoids substrates at the 7,8; 9,10; 11,12; and 15,15′ double bonds distributed across the conjugated polyene chains. The CCD

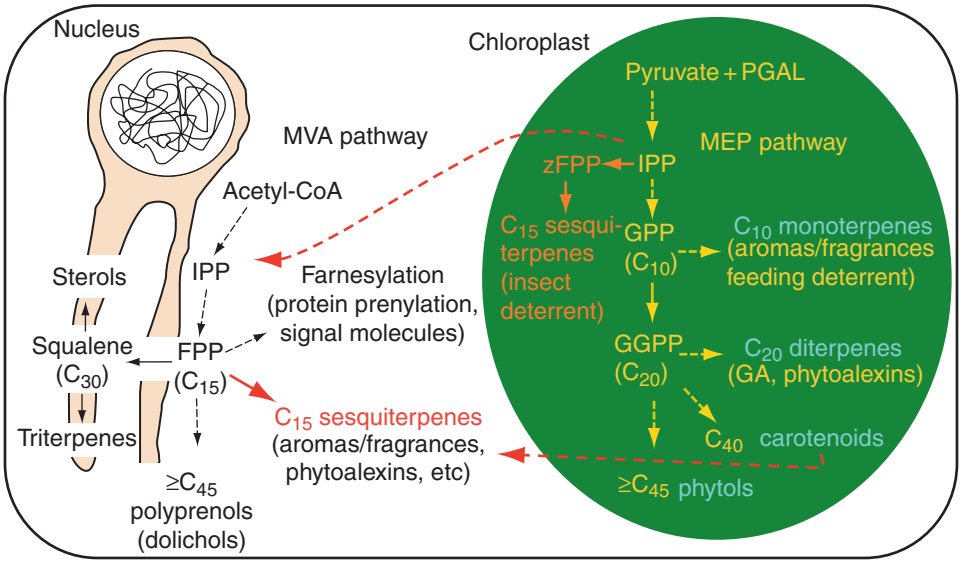
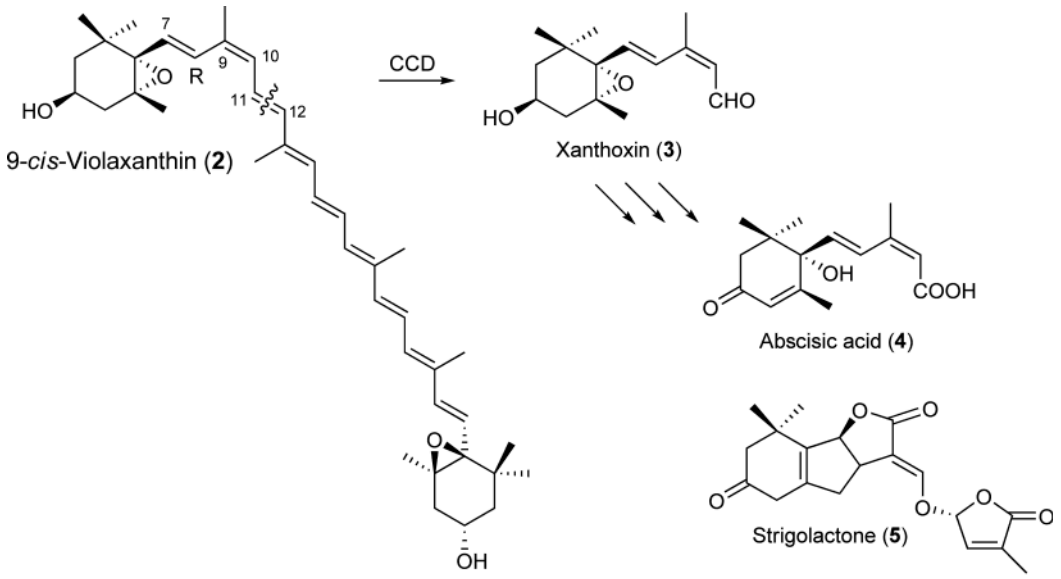


Figure 1 Cartoon depiction of terpene metabolism in a typical plant cell with an emphasis on the different carbon sources contributing to sesquiterpene biosynthesis and showing their biological functions.



Scheme 1

homologs found in animals have been associated with the conversion of dietary carotenoids into essential retinal pigments important for visual perception.[10] In plants, several of the CCDs have been found associated with the biosynthesis of strigolactones, (e.g., **5**) secreted sesquiterpenoids that serve as signal molecules for the establishment of mycorrhizal relationships[11] as well as their colonization by parasitic plants like *Striga* and *Orobanche* species.[12,13]

More recently, new members of the CCD family of enzymes have been associated with the biosynthesis of a new class of hormones controlling axillary bud growth in plants,[13,14] essentially regulating the above-ground architecture, biomass accumulation, and agricultural yields of fruit-bearing branches. Genetic mutants exhibiting constitutive and unregulated outgrowth of axillary buds were mapped to genetic loci coding for enzymes possessing amino acid signature domains for CCDs. Interestingly, the newly identified growth regulatory molecule was identified as strigolactone (**5**), the same sesquiterpenoid previously shown to mediate plant associations with mycorrhizal fungi and parasitic plants,[12] yet derived by independent carotenoid cleavage and modification reactions occurring in the roots rather than the aerial portions of the plant.

1.16.2.2 Farnesanes

The farnesane family of sesquiterpenes (**Figure 2**) is derived from FPP (**14**) in one of the two ways. Farnesol (**6**) represents the simplest of the acyclic sesquiterpenes and it can arise from the action of phosphatases cleaving the terminal phosphate substituents.

Biologically, FPP and farnesol are important molecules monitored intracellularly, which are thought to provide for homeostatic control of carbon flux into the MVA pathway.[15,16] Although there are suggestions in the literature of possible FPP-specific phosphatases, none have been confirmed to date. Instead, phosphatases associated with lipid metabolism as well as other nonspecific phosphatases have been shown to use FPP as a substrate.[17] In contrast, successive phosphorylation of farnesol by a CTP-dependent kinase(s) has been described,[18] thus at least in theory providing for the recycling of farnesol back into the active pool of carbon feeding the MVA pathway.

The family of acyclic sesquiterpenes is much larger in terms of family members, and most are initially derived by the action of specific terpene synthases (**Scheme 2**). In common with most terpene synthases, nerolidol[19,20] and farnesene[21] synthases appear to initiate catalysis by cleavage of the diphosphate substituent leaving a delocalized carbocation with positive charge distributed between C1 and C3. Capture of a hydroxyl

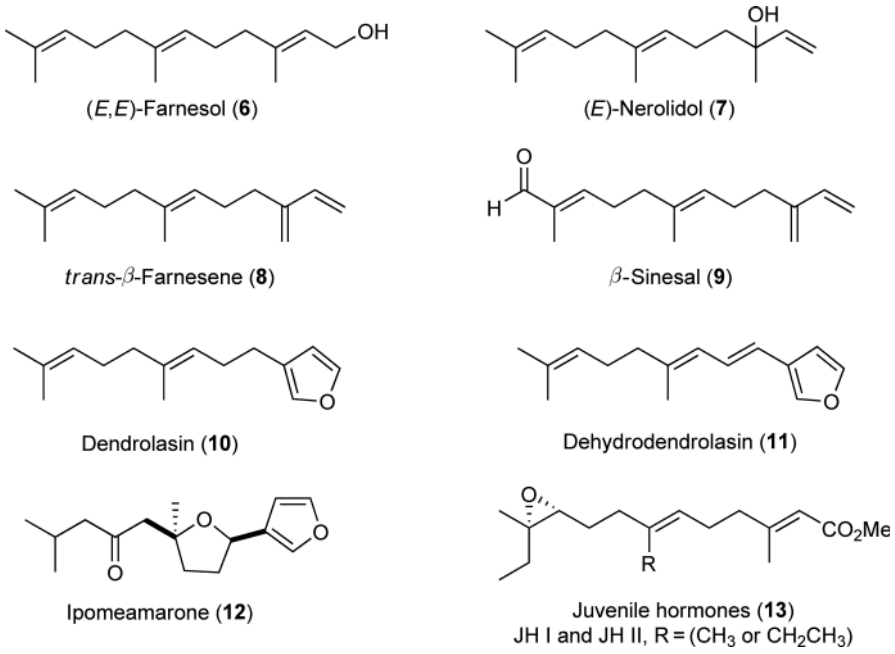
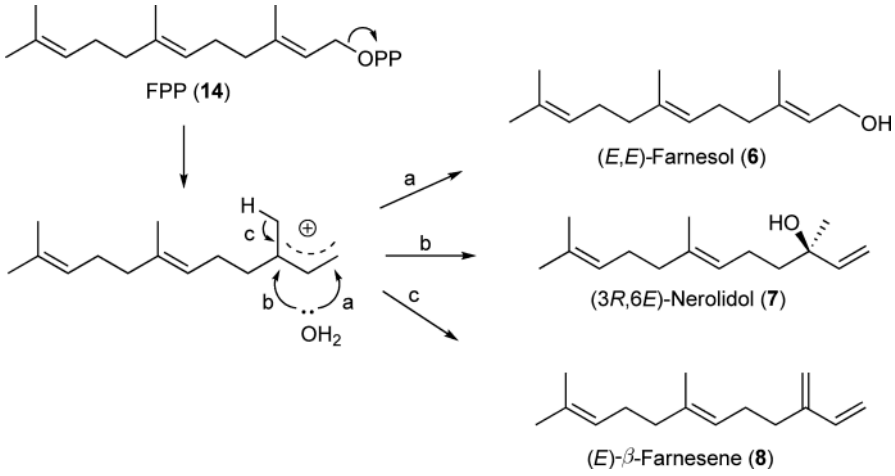


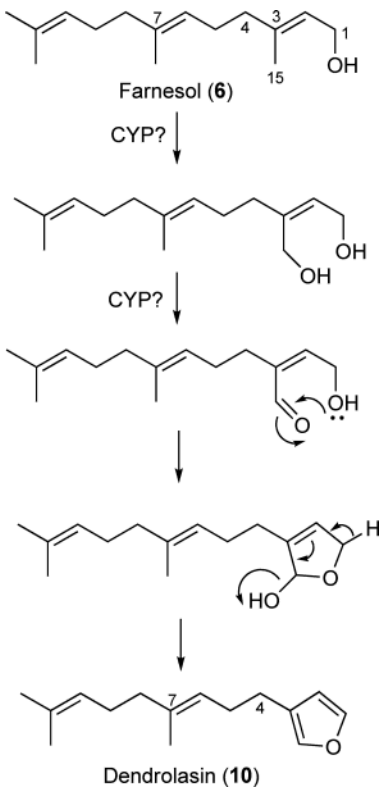
Figure 2 Structures and names of selected farnesane sesquiterpenes.



Scheme 2

group from the bulk solvent at C1 or C3 yields farnesol (6) and nerolidol (7), respectively. Nerolidol is a fragrance emitted by snapdragon flowers, which is thought to attract pollinators.[22] Alternatively, proton abstraction from the C15 methyl group yields β-farnesene (8), an alarmone first described as a volatile emitted from herbivore-damaged plants serving to agitate and disperse aphids.[23] β-Farnesene has been associated with many other activities including its action as an attractant for predator insects to herbivore-damaged plants.[24]

The acyclic sesquiterpenes are subject to further modifications including hydroxylation/oxidations that generate compounds like β-sinesal[25] (9), a flavor component of citrus. Reductions of C═O and C═C double bonds, dehydrogenations, formation of furan rings, epoxidations, and acylations serve to diversify further the straight-chain sesquiterpene backbone. Dendrolasin (10) and dehydrodendrolasin (11) are examples of furan-containing sesquiterpenes characterized as allomones.[26] In marine organisms, these compounds are found in various sponge species where they deter feeding behaviors of fish and other invertebrates,[27] and may play a



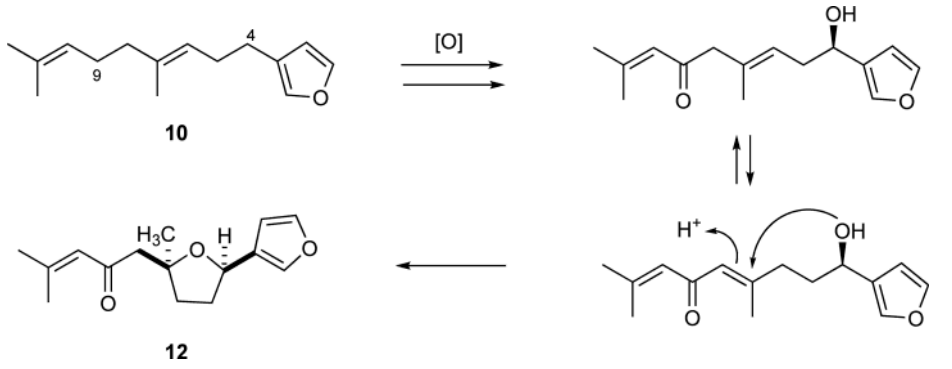
Scheme 3

similar role when emitted from angiosperm trees in response to insect predation.[28] Ipomeamarone (**12**), an antimicrobial sesquiterpene produced by sweet potato in response to pathogen and elicitor challenge,[29] appears to be an interesting derivative of dendrolasin characterized by an internal tetrahydrofuran ring.

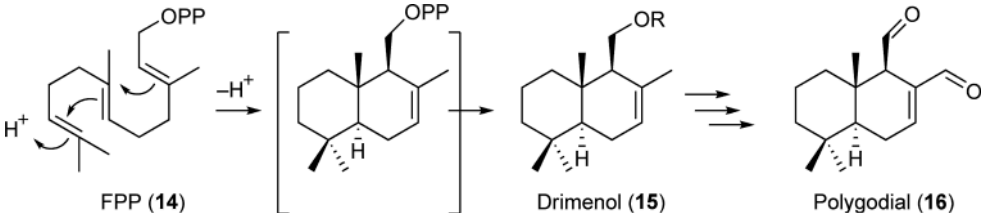
Although furan rings commonly occur within terpene structures, the biochemical mechanism responsible for the introduction of such heterocycles is still not understood. Gaikwad and Madyastha[30] provided evidence for an enzyme-mediated reaction. More specifically, these authors suggested that oxidized substituents could provide specificity for binding in the active site of a cytochrome P-450 enzyme, orienting the substrate such that a methyl group syn to the hydroxyl/carbonyl becomes oxidized. Proximity to, and correct spatial orientation of, the hydroxylmethyl group to the carbonyl group would then allow for spontaneous cyclization to the hemiacetal, which would readily undergo dehydration to yield the furan final product.[30] **Scheme 3** illustrates one possible reaction sequence in which farnesol is successively oxidized at the C15 methyl group to the aldehyde oxidation state, followed by cyclization and dehydration to form the furan ring of dendrolasin (**10**). Introduction of the internal tetrahydrofuran ring of ipomeamarone (**12**) could plausibly occur by the reaction sequence illustrated in **Scheme 4**. A succession of oxidations at C4 and C9, isomerization of the 6,7 double bond into the conjugated 7,8 position, and stereospecific conjugate addition of the C4 hydroxyl onto C7 of the resulting α,β-enone would afford ipomeamarone. The juvenile hormones JH 1(**13**, R = CH_2CH_3) and JH II (**13**, R = CH_3), isolated from male Cecropia silk worms, are homosesquiterpenes involved in maintaining the juvenile stage of development in the life cycle of insects and Crustacea (see also **Figure 10**).

1.16.2.3 Drimanes

Another rather unique group of decalin derivatives comprises the drimane class of sesquiterpenes.[31] These natural products have been found in association with many marine sponges[31] and with primitive and evolutionarily lower plants like liverworts,[32] but not exclusively. Polygodial (**16**) is a common drimane sesquiterpene occurring in various plant species and documented for its rather strong anti-insecticidal activities.[33] The biosynthetic route to the drimanes is not yet fully resolved (**Scheme 5**) and no genes coding for the



Scheme 4

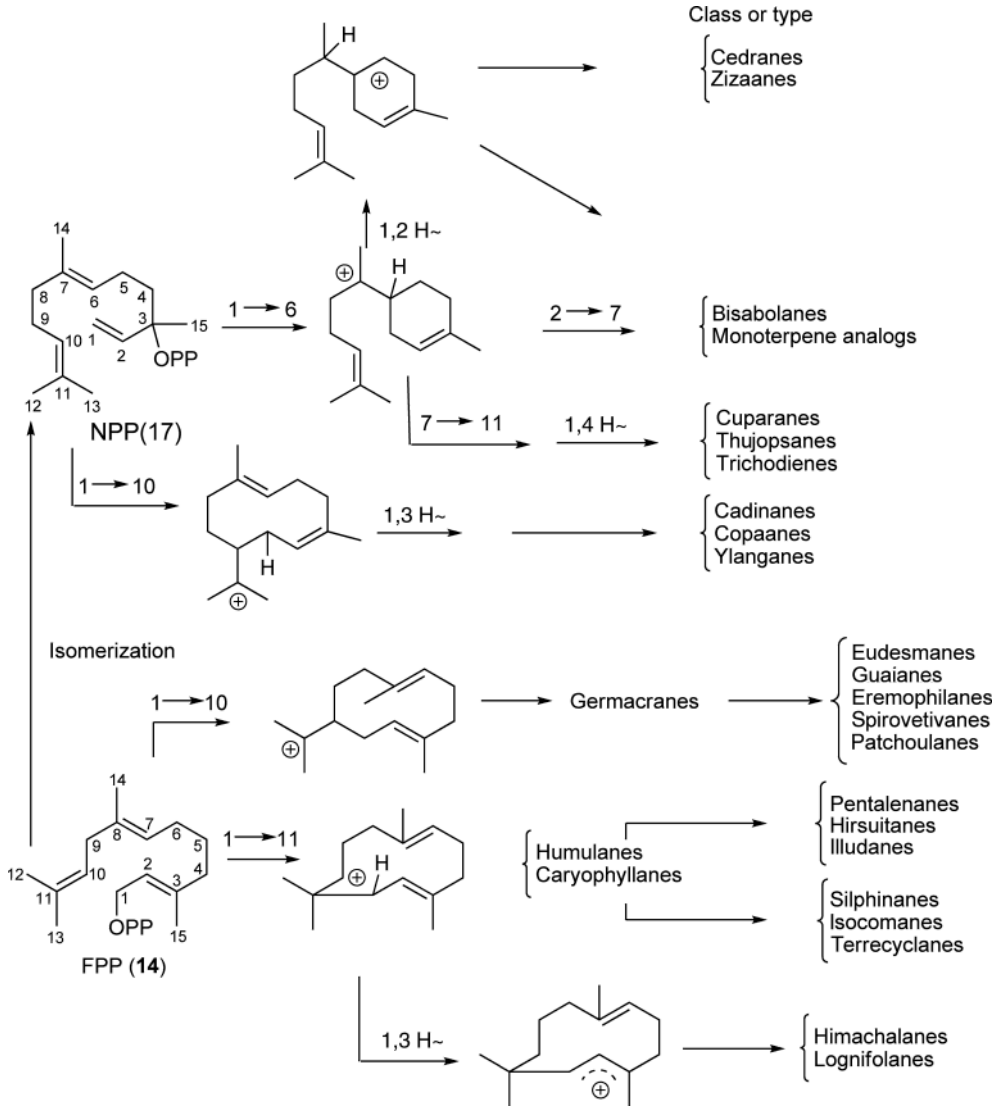


Scheme 5

corresponding synthases and oxidases have yet been reported in the literature. Nonetheless, the mechanism assumed to be involved in the biosynthesis of drimanes has precedent in the labdane diterpene synthases,[34] as well as triterpene cyclases like cycloartenol and β-amyrin synthases.[35] Interestingly, the labdane diterpenes arise from two-step cyclization cascades associated with a single enzyme in fungi[36,37] and two successive enzymes in plants.[34,38] The first step (A activity) consists of a proton-initiated cyclization of geranylgeranyl diphosphate (GGPP) to form copalyl diphosphate (CPP), a diphosphorylated, decalin A,B ring intermediate (or its enantiomer, *ent*-copalyl diphosphate). The second step is initiated by the ionization of the diphosphate substituent, followed by carbocation-mediated cascades to give various tricyclic and tetracyclic diterpenes.[39,40] In contrast, triterpene cyclases initiate catalysis by proton addition to squalene and squalene 2,3-epoxide, nonphosphorylated intermediates, which then undergo multiple cyclizations to form a variety of polycyclic structures with the imposition of strict conformational and stereochemical specificities.[41,42] Drimenol biosynthesis has been demonstrated with a partially purified enzyme preparation from *Polygonum hydropiper*[43] that utilized FPP as the substrate. However, a phosphorylated intermediate was not observed in the case of drimenol (**15**) biosynthesis. The existence of proton-initiated sesquiterpene synthases has speculative significance because this crucial, high-energy step has been suggested by some to represent an evolutionary link between the more primitive ionization-induced mechanism of cyclization and the evolutionarily advanced proton addition mechanism.[31]

1.16.2.4 Cyclic Sesquiterpenes

The vast majority of sesquiterpenes isolated from microbial and plant sources have cyclized structures and, at first glance, a classification scheme based on overall structure seems reasonable. That is, gathering what initially appears to be a disparate array of compounds into classes like monocyclic, bicyclic, tricyclic, and macrocyclic compounds classifies these compounds into groups resembling one another. However, a more cogent classification scheme based on chemical rationalizations for the biogenetic pathways leading to particular classes of compounds is possible, as illustrated in **Scheme 6**. The value of this classification as we hope to illustrate below is that it lends itself to experimental exploration into the origins for this chemical diversity. To avoid confusion, we consistently refer to carbon position numbers for the farnesyl chain marked on structure **14** (**Scheme 6**).



Scheme 6

The same positional numbers will be used in the following schemes to designate individual carbon atoms in intermediate carbocations and the final cyclic sesquiterpene products.

The bewildering diversity of cyclic structures associated with naturally occurring sesquiterpenes can be rationalized by different combinations of relatively few individual reactions of carbocation intermediates. It seems worthwhile at the outset to offer some generalizations about the relative stabilities of carbocations, the basic types of carbocation reactions, and the thermodynamic driving forces in the transformations. Carbocation formation is most commonly initiated by heterolysis of the C–O bond of FPP to generate the allylic farnesyl$^+$/OPP$^-$ ion pair. Less common initiating steps in sesquiterpene biosynthesis are proton transfers to C═C double bonds or epoxides, the former previously presented in the context of drimane biosynthesis. The polycyclic structures are then elaborated by permutations and combinations of the following individual propagating steps: (1) carbocation + C═C double bond cyclizations; (2) 1,2-, 1,3-, and 1,4-hydride shifts; and/or (3) Wagner–Meerwein rearrangements. The cascade of propagating steps is then terminated by 1,2- or 1,3-proton eliminations forming C═C double bonds or cyclopropane rings, or by capture of a water molecule to produce alcohols.

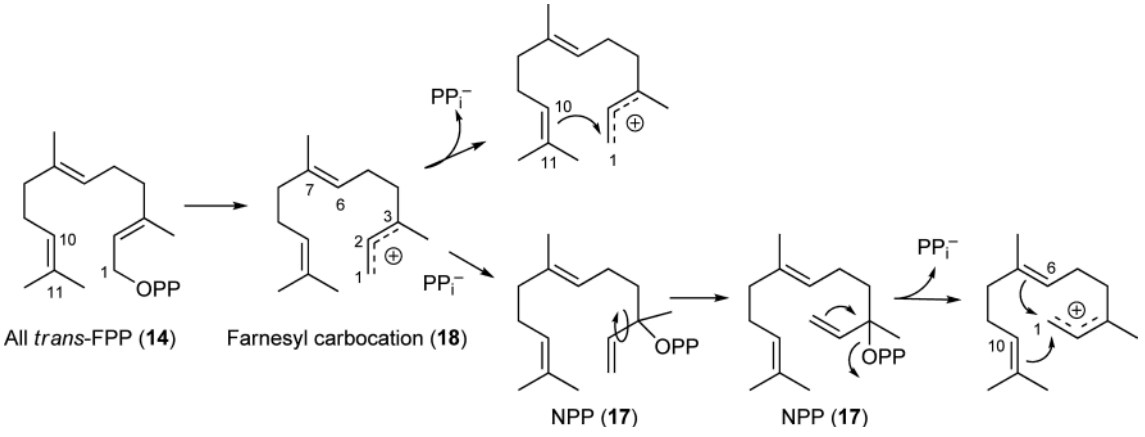
In the special case of the FPP– nerolidol diphosphate (NPP) interconversion to be discussed later, the farnesyl carbocation in the farnesyl$^+$/OPP$^-$ ion pair is captured by the pyrophosphate counterion at C3.

The stabilities of carbocations vary greatly depending upon the substituents attached to the electron-deficient site. Tertiary carbocations are more stable than similar secondary ions by 15–20 kcal mol^{-1}. Allylic carbocations like the farnesyl ion are greatly stabilized by delocalization of the positive charge between the C1 and C3 termini. Cyclizations of carbocations with C═C double bonds are typically exothermic by 20 kcal mol^{-1} that arises from the conversion of the C═C double bond into two C–C single bonds, if the stabilities of the reactant and product carbocations are otherwise similar. The strain energies associated with carbocyclic rings are likely to be significant factors in different types of cyclizations; approximate strain energies of cyclopropane, cyclobutane, cyclopentane, and cyclohexane are approximately 25, 25, 7, and 0 kcal mol^{-1}, respectively.

In general, the multistep schemes in sesquiterpene biosynthesis will proceed to generate more stable carbocations from less stable ones. Relatively high-energy secondary carbocations, usually regarded as key intermediates, may be generated by exothermic carbocation + C═C cyclizations such as the case of the secondary ion intermediate (**24**) formed by anti-Markovnikov cyclization of the bisabolyl ion in trichodiene biosynthesis (see **Scheme 9**). The enzyme-bound carbocation intermediates may be stabilized by specific interactions with peptide residues surrounding the active sites of sesquiterpene synthases, e.g., cation-pi interaction, with aromatic rings of phenylalanine, tyrosine, and tryptophan; dipole–dipole interactions with amide linkages; and ion pairing with carboxylate groups in the side chains of aspartate and glutamate residues. Carbocations are exceptionally strong acids similar in proton-donating ability to concentrated sulfuric acid. Consequently it follows that the common terminating steps of 1,2-proton elimination and nucleophilic capture of water are both highly exothermic transformations. The increased solvation of the liberated pyrophosphate ion and its associated magnesium ion ligands contributes added driving force to the overall cyclization processes. It should be evident that biosyntheses of most sesquiterpenes from FPP are highly exothermic chemical transformations.

The usual starting substrate for all these different sesquiterpene classes is (*E,E*)-FPP (**14**), which undergoes an initial cleavage of the diphosphate substituent leaving a reactive farnesyl carbocation. It is the subsequent cascade of events of this carbocation intermediate that leads to the final reaction products, the hydrocarbon skeleton characterizing a particular class of sesquiterpenes. And, these cascades are dictated by the configurations and conformations of the various reactive intermediates, positioning the reactive carbocation centers into proximity and alignment with other atoms and bonds within the C_{15} scaffold, which then allows for particular steps to proceed while limiting or preventing others. Later in this chapter, we illustrate how the terpene synthases might provide geometric landscapes that essentially create the mold to direct an ionized farnesyl intermediate down a specific catalytic cascade.

As shown in **Scheme 7**, the initial ionization leads to a 2,3-transoid farnesyl carbocation ion pair (**18**) that can either lose the ionized diphosphate counterion, thus permitting bond formation between C1 and C10 or C11 and leading to a series of sesquiterpenes derived from these macrocyclic intermediates, or the ions may

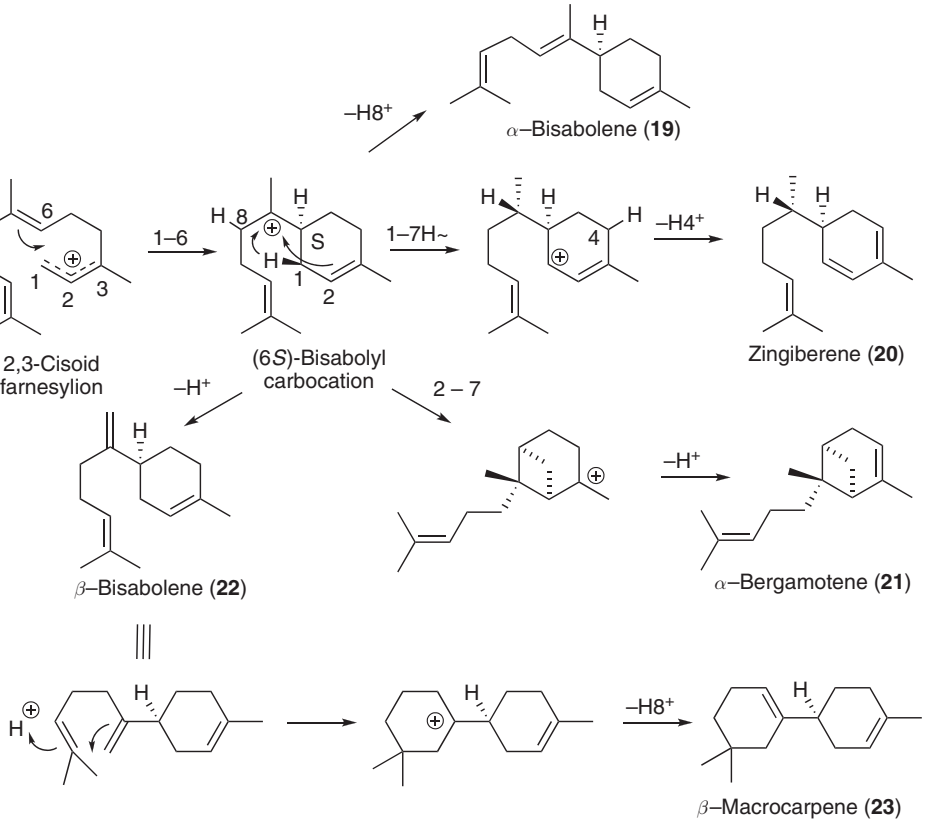


Scheme 7

recombine to form a new intermediate, NPP (**17**).[2] NPP is an important intermediate because it allows for rotation around the newly formed single bond between C2 and C3. A second ionization of the 2,3-cisoid NPP thus creates a new, stereochemically distinct intermediate, positioning the carbocation for ring formation between C1 and C6 or C10, and opening the way to several more classes of sesquiterpenes distinguished by a *cis* configuration, rather than *trans* configuration, about the C2═C3 double bond.

1.16.2.4.1 The bisabolane series

The bisabolane group includes sesquiterpenes found in the fragrances from familiar plants like lavender,[44] as well as many coniferous species.[45,46] This class may also appear to be among the simplest of sesquiterpenes because of their structures, consisting of many, but not exclusively, monocyclic members (**Scheme 8**). They are however representative of the same level of complexity found in all the cyclic sesquiterpenes. Several of the early cDNAs isolated encoding for terpene synthases were for enzymes catalyzing the conversion of (*E*,*E*)-FPP to (*E*)-α-bisabolene (**19**) from gymnosperms,[45] and these have been useful in illustrating the putative reaction mechanism. As noted in **Scheme 7**, formation of the 2,3-cisoid rotomeric form of NPP (**17**) is followed by an ionization step generating a carbocation with charge distributed between C1 and C3, and, if oriented correctly, allowing for ring formation between C1 and C6 in a stereospecific manner. With the carbocation site now centered at C7, proton abstraction at C8 terminates the reaction cascade yielding α-bisabolene (**19**), known sequiterpenes as both the 7E and 7Z isomers. Zingiberene (**20**) biosynthesis proceeds up to the same bisabolyl carbocation intermediate, but is then presumably followed by a 1,3-hydride shift and subsequent proton abstraction to yield the final 1,3-cyclohexadiene product.[47] The biosynthesis of α-bergamotene (**21**) illustrates yet another degree of freedom within these synthase reactions with the formation of a bicyclo[3.1.1]heptane ring system arising from the formation



Scheme 8

of a new C2–C7 bond and recentering of the carbocation at C3. Final proton abstraction from C4, similar to that for zingiberene biosynthesis, yields a new double bond between C3 and C4 and formation of α-bergamotene.[44,48]

The biosynthesis of β-macrocarpene (**23**) illustrates a fourth permutation on the fate of the bisabolyl carbocation intermediate. Abstraction of a proton from C14 yields β-bisabolene (**22**), which is commonly found as a final product in many plant species,[49] but in the case of two highly similar maize enzymes the triene actually undergoes yet another round of catalytic transformations initiated by a proton addition.[50] In this case, the addition of a proton, either from a water molecule bound in the active site or the proton abstracted in an earlier step, to the C10═C11 double bond positions a favorable tertiary carbocation at C11. This carbocation, if brought into close enough proximity to C14 and with the proper stereochemical orientation to the pi orbital of the C7═C14 double bond, allows for a new C–C bond to form, creating a second six-membered ring with gem dimethyl substituents and shifts the carbocation to C7. A second proton abstraction from C8 thus yields β-macrocarpene (**23**).

Support for the chemical rationalization of β-macrocarpene biosynthesis by the maize TPS6/11 enzymes comes from several experimental observations.[50] First, GC-MS profiling of the reaction products identified the predominate product as β-macrocarpene (**23**) with β-bisabolene (**22**) as a significant minor product. Although time course experiments did not demonstrate a precursor–product type relationship between β-bisabolene and β-macrocarpene, the presence of β-bisabolene demonstrates that these two enzymes are capable of biosynthesizing this compound. Second, reinitiation of the second round of reactions via a proton donation was demonstrated by rehydrating lyophilized TPS6 enzyme in D_2O, then observing the incorporation of one extra mass unit into the appropriate MS fragments of the β-macrocarpene reaction product. Third, using a strategy employed by Rising *et al.*[51] for mapping active site residues involved in proton addition and reinitiation of catalysis for 5-epi-aristolochene synthase (EAS), another sesquiterpene synthase discussed below, Köllner *et al.*[50] observed that mutation of a putative active site tyrosine (position 522) to phenylalanine abolished β-macrocarpene biosynthesis without any alteration in the proportion of β-bisabolene accumulating. One interpretation of this result, as concluded by the authors, is that tyrosine 522 is responsible for the reinitiation of catalysis, thus enhancing substrate flux through the enzyme and accumulation of the macrocarpene product. But this conclusion, unlike that of Rising *et al.*,[51] is not consistent with standard biochemical expectations. Knocking out the second series of reactions in these maize enzymes should eliminate β-macrocarpene accumulation, but then β-bisabolene accumulation should be enhanced to a level equal to β-macrocarpene plus β-bisabolene accumulation by the wild-type enzyme. But this was not observed, suggesting that the mutation compromised the enzyme in at least two ways – knocking out the second series of reactions and reducing the flux of substrate through the enzyme, possibly by hindering product release.[52,53]

The bisabolane family of sesquiterpenes is notable because of the different ways in which stereocenters are introduced and regioisomers are created. Hydride shifts and ring closures can lead to the creation of stereoisomers, while the generation of configurational isomers can arise from regiospecific proton abstractions, as illustrated in **Scheme 8**. For example, the maize TPS6 enzyme creates the 6S stereochemistry in the initial 1–6 cyclization, and directs elimination of H8 to yield **23**, rather than an isomer if H14 were abstracted generating a 7,14 double bond.

Köllner *et al.*[54] have also provided compelling evidence that another maize terpene synthase, TPS4, possesses two active site pockets rather than one as observed in the available three-dimensional structures solved for sesquiterpene synthases.[5,55–59] TPS4 catalyzes the biosynthesis of 14 bisabolane type sesquiterpenes, many containing only a single 1,6 ring, and a second family of bicyclic compounds derived after a second 2,6 ring closure, like that in bergamotene biosynthesis. Using the structural coordinates of EAS to build molecular models of TPS4, these investigators noted the possibility of two distinct pockets within the TPS4 enzyme, both sufficiently large to accommodate binding of various proposed reaction intermediates (**Figure 3**). To test if one pocket might direct conversion of FPP to the initial bisabolene intermediate, and the second functions to introduce 2,6 ring closures, they created site-directed mutants in each pocket that were evaluated for the reaction products generated upon incubation with *trans*,*trans*-FPP or *cis*,*trans*-FPP, a possible isomeric intermediate proposed as essential for the 1,6 ring closure. Overall, Köllner *et al.* reported that mutations in pocket I residues prevented formation of reaction products from (*E*,*E*)-FPP, but not (*Z*,*E*)-FPP, and that mutants in pocket II residues lost their capacity for the biosynthesis of bicyclic compounds, but maintained their ability to generate acyclic and monocyclic bisabolane-type compounds. Hence, the results are consistent with reactions

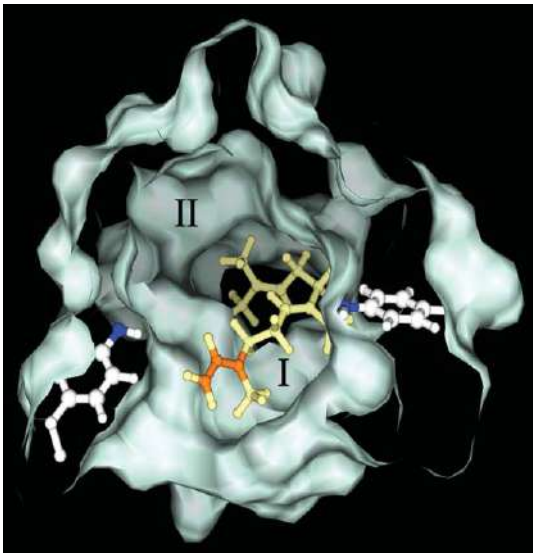


Figure 3 A molecular model of the maize TPS4 enzyme, highlighting the two putative active site pockets, I & II. Reproduced from T. G. Köllner; P. E. O'Maille; N. Gatto; W. Boland; J. Gershenzon; J. Degenhardt, *Arch. Biochem. Biophys.* **2006**, *448* (1–2), 83–92.

leading up to the first ring closure occurring in pocket I, and the diversification of these bisabolene intermediates, including a second ring closure, occurring in pocket II.

More recently, several groups investigating the terpene biosynthetic potential of trichomes (surface appendages that secrete rich arrays of plant secondary compounds) using microarray and expressed sequence tags (EST) DNA sequencing approaches have uncovered new synthases that catalyze the biosynthesis of sesquiterpenes in the bisabolane family.[47,60] The report by Sallaud *et al.*[60] is particularly intriguing; they found that the Sst2 locus in wild tomato, *Solanum habrochaites*, was responsible for the biosynthesis and accumulation of several bisabolene-derived sesquiterpenes, including bergamotene and santalene isomers, and their acidic forms that provide insect resistance to the wild tomato species. Molecular sequencing of the Sst2 locus identified two closely linked terpene synthases, one coding for a prenyltranferase, FPP synthase (FPS), and the other for santalene and bergamotene synthase (SBS), a multifunctional synthase catalyzing the biosynthesis of several isomers each for santalene and bergamotene from FPP. However, what makes this observation especially unique is that both the FPS and the SBS contain plastid-targeting signal sequences at their amino termini, which were shown to mediate vectorial transport of these enzymes into the plastid compartment. Along with the work from Besser *et al.*,[61] a unique pathway for sesquiterpene biosynthesis has been now been localized to the plastid compartment (see **Figure 1**).

Sallaud *et al.*[60] reported two additional extraordinary findings. First, the *S. habrochaites* FPS utilized IPP and DMAPP to synthesize (*cis*,*cis*)-FPP rather than the universally observed all-*trans* FPP, and second, not only did the SBS enzyme readily utilize the (*cis*,*cis*)-FPP to generate stereochemically unique forms of santalene and bergamotene, but based on informatic comparisons, the synthase utilizes a proton addition mechanism to initiate catalysis similar to that found with di- and tri-terpene synthases.[34,35] These observations clearly indicate that a very unique form of coevolution must have been occurring leading up to this novel sesquiterpene biosynthetic capacity in plastids. The evolution of the genes producing (*cis*,*cis*)-FFP synthase and the SBS cyclase that utilizes the geometrically isomeric substrate was explained[60] as a means to circumvent the strong regulatory constraints associated with the biosynthesis of cytosolic (*trans*,*trans*)-FPP.

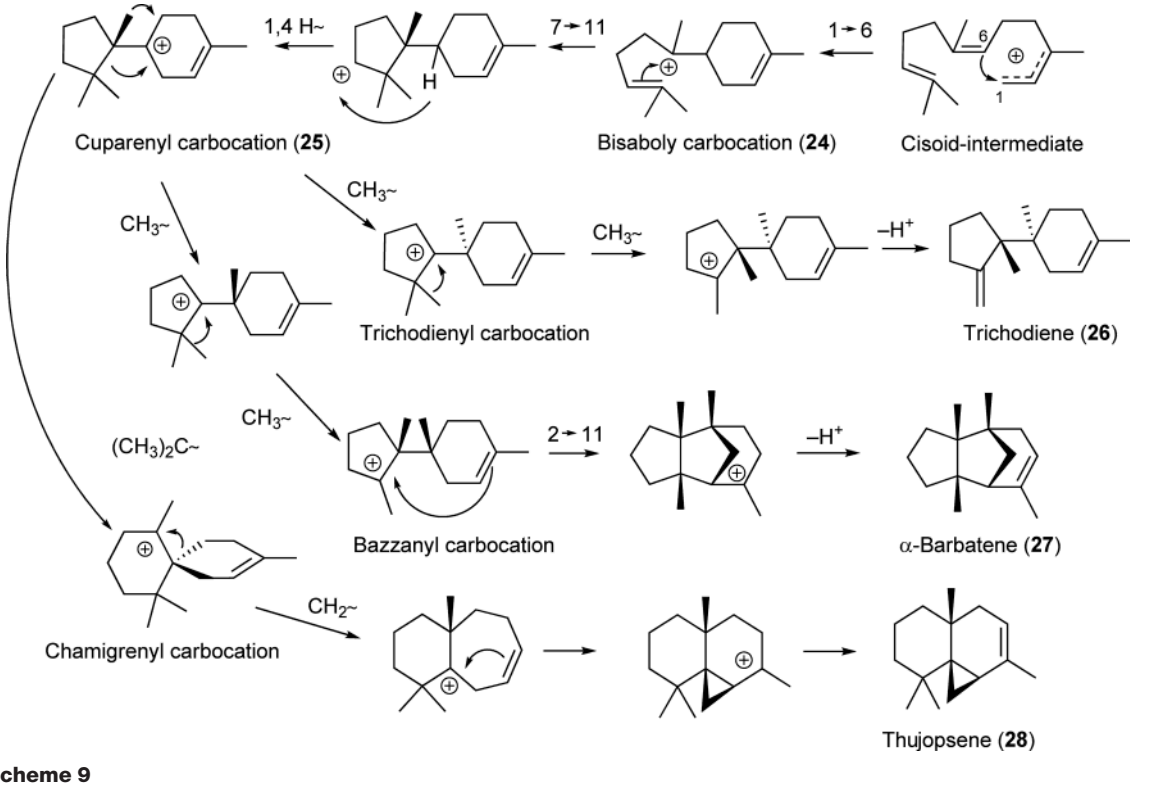
1.16.2.4.2 The cuparane series

The cuparane series of sesquiterpenes is one of the best understood classes of compounds from the perspective of their biological importance and their biosynthesis. Their chemical appreciation arises because some of the most potent and regulated mycotoxins, the trichothecanes, are members of this class.[62] The trichothecanes are derived from the tricyclic hydrocarbon trichodiene (**26**) produced by pathogenic fungi that colonize cereals like

oat, corn, and wheat, and consist of a family of derivatives containing oxidized and epoxide functionalities associated with the biological activity of these mycotoxins.[63] The effects of these compounds on humans and livestock have been studied extensively because of the severity of symptoms caused upon ingestion, which include severe digestive disorders like vomiting and diarrhea, hemolysis, disruption of cognitive function, and death.[64] The permissible level of these mycotoxins in foods is strictly regulated by federal agencies and, while levels below 2 ppm are considered tolerable, there has been movement to lower these allowable levels substantially in the recent past.[65] Investigations have identified protein translation as the primary mode of toxicity[66] for the trichothecanes, and in particular have identified interactions of trichothecanes with the L3 protein of the large ribosomal subunit.[67]

Many of the genes for trichodiene and trichothecane biosynthesis have been isolated and functionally characterized.[62,63] Interestingly, given their fungal origin, the respective genes appear to be clustered, somewhat analogous to the clustering of terpene biosynthetic genes observed in prokaryotic organisms like *Streptomyces*.[68] Clustering is thought to provide some advantages for coordinate expression of the resident genes.

The biosynthesis of trichodiene, the immediate precursor to the entire trichothecane family, has been studied intensively and the results largely support the chemical rationale illustrated in **Scheme 9**. Trichodiene biosynthesis proceeds from a bisabolyl carbocation intermediate (**24**) as discussed above by a ring closure between C7 and C11, followed by a 1,4-hydride shift to yield the cuparenyl intermediate (**25**). Several features of this intermediate distinguish it from other sesquiterpene classes and position it uniquely for the following transformation pathways. First, it consists of a bicyclic ring system, one five-membered and the other six-membered, that are not fused, but rather coupled by simple carbon–carbon linkage. Second, formation of the five-membered ring can result in the methyl substituent at C7 being in one of two stereo configurations. Trichodiene synthases appear to produce exclusively the 6*S*-enantiomer of the trichodienyl carbocation,[3] while the recently characterized α-barbatene synthase from a plant generates the 6*R*-enantiomer shown in **Scheme 9**.[69] The possibility that the (6*R*)-trichodienyl carbocation precursor to α-barbatene actually arises from the 7S stereoisomer of the cuparenyl carbocation (**25**) cannot be excluded at this time. Lastly, generation of the key cuparenyl carbocation intermediate (**25**) requires a unique 1,4-hydride shift, which cannot proceed



Scheme 9

by successive 1,2 shifts, because of the quaternary carbons at C7 and C11. Instead, the 1,4-hydride shift must occur directly between carbons 6 and 10 without the involvement of any intervening carbons, and because C6 and C10 are positioned in close enough proximity to one another, in the proper stereoorientation (the so-called synperi-planar orientation), to allow for the shift to occur. It should also be noted that the 1,4-hydride shift converts a high-energy secondary carbocation to the much more stable tertiary isomer (**25**). Alternatively, quantum mechanical calculations[70] indicate that this overall transformation might occur in two steps by a lower-energy mechanism involving proton transfer from C6 to C10 of (**24**) to form an isomeric bisabolyl carbocation followed by 7–11 cyclization to cuparenyl ion (**25**). Two 1,2 methyl migrations and a regiospecific proton abstraction of the rearranged cuparenyl intermediate thus yields trichodiene (**26**).

Experimental evidence for trichodiene chemical rationalization comes for diverse approaches including inhibitors designed to mimic proposed intermediates,[71–74] site-directed mutagenesis,[75] structural elucidation of the trichodiene synthase with and without ligand binding,[57] and combinations of these efforts.[71,76,77] The three-dimensional structure determined by Rynkiewicz *et al.*,[57] for the *Fusarium sporotrichioides* trichodiene synthase pointed out the importance of three metal cofactors binding near the entrance to the active site, the binding coordination between these metal cofactors and the diphosphate substituent of FPP, and this work identified a significant conformational change that occurred at the entrance into the active site cavity upon ligand binding. These authors were also able to model various intermediates into the enzyme's active site and suggest ways that FPP binding might induce conformation changes within the active site protecting carbocation intermediates, yet positioning active site residues, particularly hydrophobic ones, to coax intermediates down a particular cyclization cascade. Development of substrate analogs to inhibit particular steps in the trichodiene synthase-catalyzed reaction has been described and demonstrated to induce conformation changes in the enzyme much like substrate binding.[71]

However, given the large number of mutants created and critically evaluated, few if any of these mutants have terminated the reaction cascade at a particular step and resulted in the predominate production and release of one intermediate. Instead, mutants like N225D or Y295F exhibited a small increase in the amount of cuparenene (2.4% increased to 10.4%) or bisabolene (0.9–17.5%) released, respectively, while still accumulating the lion's share of product as trichodiene.[76] This type of observation is not particularly surprising once one considers that it is the contour of the active site and not particular active site residue interactions with reaction intermediates that direct catalysis.

The cuparenyl carbocation (**25**, **Scheme 9**) is the likely progenitor of many other important families of sesquiterpenes, including those in the bazzanane, gymnomitrane, and chamigrane families, illustrated by α-barbatene (**27**) (a gymnomitrane member) and thujopsene (**28**) (a chamigrane member). These sesquiterpene families appear to be highly represented in the lower plants, especially liverworts,[78] and until recently were not considered common in higher plants. Two independent efforts recently identified an *Arabidopsis* α-barbatene synthase, one using a reverse genetics approach[79] and the other a novel mechanism for functionally characterizing genes identified in genome sequencing projects.[69] Interestingly, the plant enzyme generates three major products in roughly equal amounts, along with eight or more minor products. This is distinctly different from the *Fusarium* trichodiene synthase that biosynthesizes predominately one reaction product, trichodiene, in greater than 82% yield. However, given the notion from Cane's and Christianson's[5] work that active site geometry dictates reaction product specificity for this synthase, a significant question is whether a single enzyme species can yield multiple products or if there are distinct enzyme conformers responsible for each type of reaction product. The former explanation would suggest that the active site cavity is large enough and flexible enough to accommodate a large geometric variation in the intermediates, but that some dynamic in the protein–ligand structure imparts a particular geometric tunnel (catalytic cascade) to a substrate once it is bound. The latter explanation would suggest that the folding algorithms for a particular amino sequence result in slightly but significantly different protein conformers, and that the relative abundance of a particularly folded conformer would correspond to the relative synthesis rate for a particular reaction product. The current information is insufficient to distinguish between these possibilities. However, given the preponderance of these families of sesquiterpenes in liverworts, it will be very interesting to examine the corresponding synthases for conservation of amino acid sequences and functional conservation in terms of their catalytic and reaction product specificities.

1.16.2.4.3 The cadinane series

Gossypol (**29**) and artemisinin (**30**) (see **Figure 4**) are two of the most prominent members of the cadinane family of sesquiterpenes. Gossypol consists of two highly oxygenated forms of cadinene (hemi-gossypol) joined together by a binaphthyl linkage to form a potent toxin that affords an unparalleled level of insect resistance to cotton plants, *Gossypium barbadense*, and related species.[80] Gossypol accumulates in lysigenous glands found in stems, leaves, and the cotyledons of seeds and knockout mutants of cadinene synthase, the key enzyme dedicating FPP to gossypol biosynthesis, does result in cotton plants much more susceptible to microbe and insect predation, yet the seed meal resulting after oil extraction is now suitable for animal consumption.[81,82] Desoxyhemigossypol and hemigossypol also accumulate in wild-type cotton plants and cells in response to fungal pathogen or elicitor challenge, serving as phytoalexins or defense compounds to microbial infection.[83]

Artemisinin (**30**) is a potent antimalaria drug now being developed from *Artemisia annua*.[84] *A. annua* has been used for hundreds of years in traditional Chinese medicines and teas to treat a variety of ailments including malaria, and artemesinin, a cadinane-type sesquiterpene containing a unique peroxide bridge, was subsequently identified as the active ingredient. Artemisinin appears to be synthesized in leaf/floral trichomes, 4–8 cell surface structures that accumulate and secrete compounds mediating plant–environment interactions. Extensive efforts are currently underway to build high-yielding production platforms for artemisinin using transgenic plants[85] and fermentable organisms;[86] hence considerable information about the genes responsible for the artemisinin biosynthetic pathway and its metabolic engineering is now available.[87]

The biosynthesis of gossypol proceeds from FPP via a δ-cadinene intermediate (**32**), while that for artemisinin utilizes amorph-4,11-diene (**33**, **Scheme 10**). And while these two intermediates share overall structural similarities common to all cadinanes, these sesquiterpene scaffolds now appear to be formed by distinctly different mechanisms. In the case of δ-cadinene, the cisoid farnesyl carbocation derived from (*E,E*)-FPP (**Scheme 7**) undergoes an initial C1–C10 ring closure to generate the macrocyclic germacradienyl intermediate (**31**) having the proper stereochemistry found in the final cadinene compound. A 1,3-hydride shift repositions the carbocation to facilitate a second ring closure, a C1–C6 cyclization, to yield a putative cadinenyl carbocation. A final proton abstraction from C6 releases the final δ-cadinene product (**32**).

The rationale for the *cis*-germacradienyl pathway has been challenged by the possibility of a bisabolyl pathway (**Scheme 10**).[88] In this scenario, the first ring closure is a C1–C6 cyclization, followed by a 1,3-hydride shift and the second C1–C10 ring closure. However, a conceptual difficulty with this pathway is the generation of the carbocation centered at C11 and the necessity for the charge to migrate to C7 in order to allow for the final proton abstraction at C6. Although 1,5-hydride shifts are known to occur readily in long-lived carbocations in solution, the 1,5-hydride shift illustrated in **Scheme 10** would require that the disubstituted cyclohexane ring adopt a high-energy boat conformation with the methyl and isopropyl on the prow positions. Nonetheless, this option must be considered because of additional experimental observations. Although incubation of δ-cadinene synthase with FPP yielded 98% δ-cadinene as the sole reaction product, incubation of this enzyme with NPP (**17**) yielded 62.1% δ-cadinene, 8.4% β-bisabolene, 15.8% α-bisabolene, and 9.8% farnesene.[88] Hence, either this enzyme utilizes a bisabolyl cyclization cascade and incubation with NPP supports the premature release of reaction intermediates, or the enzyme must possess the capacity to catalyze distinct cyclization cascades independently.

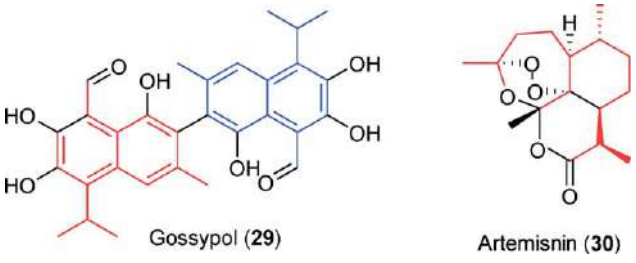
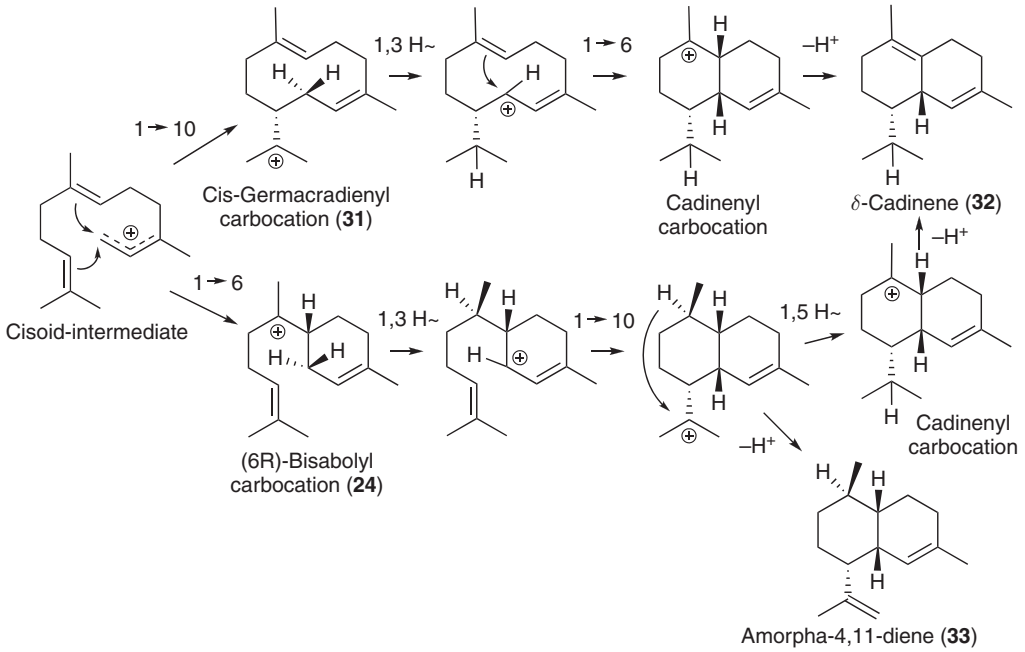


Figure 4 Structures of gossypol and artemisinin.



Scheme 10

In contrast, Yoshikuni *et al.*[89] have provided compelling evidence for the germacradienyl pathway leading to δ-cadinene using a molecular mutagenesis approach. These investigators sought to evoke new catalytic potential from the δ-cadinene synthase by subjecting the gene to a random mutagenesis protocol, then screening the mutant enzymes for their catalytic specificities. Importantly, these investigators observed the development of a germacrene D-4-ol synthase activity within independently derived, mutant cadinene synthase enzymes, but never observed the evolution of a mutant capable of synthesizing bisabolene-type sesquiterpenes. One interpretation of these results is that the authors only generated mutants that terminated catalysis after the initial C1–C10 ring closure and, because of a fortuitously positioned water molecule in the active site, the reaction was terminated upon water capture.

Conversely, the bisabolene pathway seems likely to be the prevailing mechanism operating within the amorphadiene synthase (ADS), and the germacradienyl pathway does not appear to be present in this enzyme. Picaud *et al.*[90] reasoned that because standard incubations of ADS with FPP reliably yielded small amounts of bisabolene-type reaction products, these might be the result of a premature quenching of the reaction cascade via a bisabolene pathway to amorphadiene. In contrast to the observation of Yoshikuni *et al.*[89] with cadinene synthase, Picaud *et al.* also noted the lack of any germacrane-type reaction products in any of their enzyme incubations. These investigators also noted an interesting parallel that incubation of ADS with GPP resulted in several cyclic monoterpene products while EAS, another sesquiterpene synthase previously demonstrated to utilize a germacrenyl intermediate for sesquiterpene biosynthesis,[51] only generated acyclic monoterpenes, consistent with the utilization of an isomerization step by ADS. Following up on these observations, Picaud *et al.* then prepared specific deuterium-labeled FPP and followed the incorporation of the deuterium into the amorphadiene product by NMR analyses. The labeling patterns were fully consistent with the bisabolene cyclization cascade.

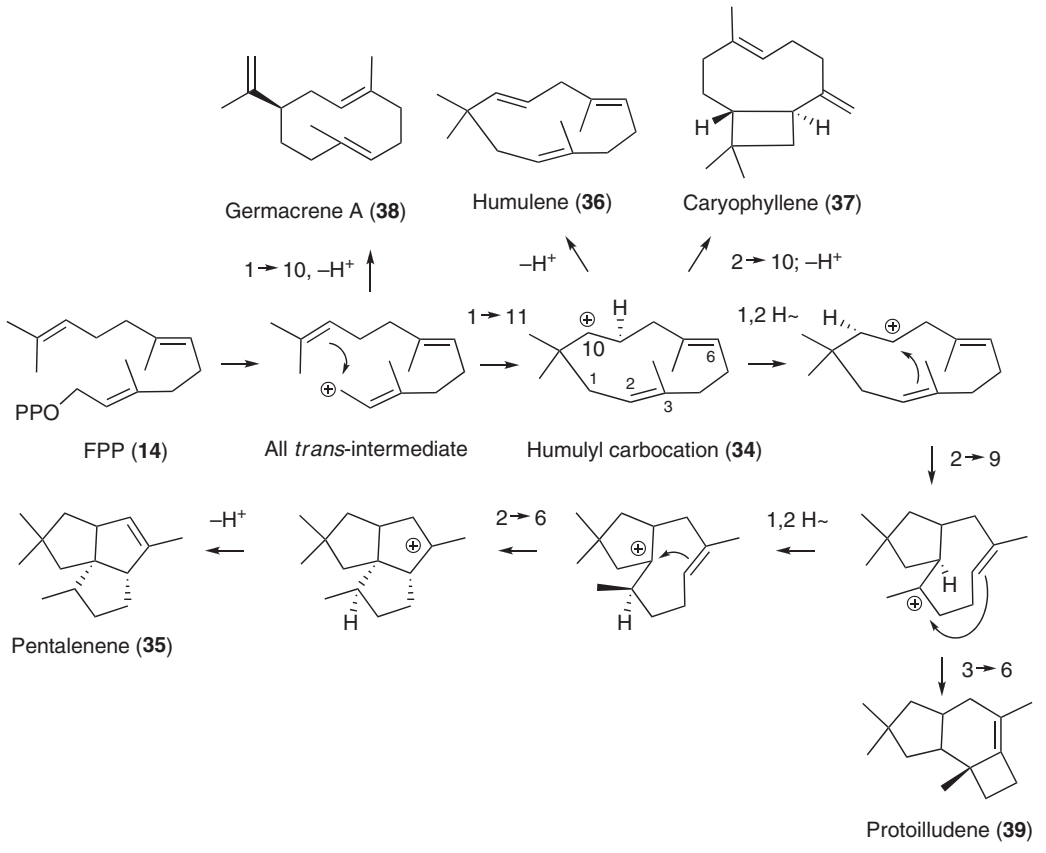
The observation that the cadinane family of sesquiterpenes arises from synthases using different reaction mechanisms shakes our intuitive notions about the origins of this family of compounds. Most researchers in the field naturally assumed that the key synthase enzymes for this class of compounds would rely on a common biosynthetic mechanism, and that somehow this mechanism must have been fixed and selected for over evolutionary time. This is clearly not the case, at least as it pertains to the cadinane family of sesquiterpenes, and may indicate independent evolution of these capacities in plants. Cotton, the species

representing cadinene synthase, is in Malvaceae, which split from Asteraceae, of which *A. annua* is a member, some 113 million years ago.[91] Whether the terpene synthase enzymes found in these extant species arose from a common ancestor or from independent ancestors does not seem to be the key issue here, but rather more fascinating is the development of enzymes producing sesquiterpenes within a common chemical family via distinctly different reaction mechanisms.

1.16.2.4.4 The humulane series

Humulene and caryophyllene are representative of the sesquiterpenes derived from an initial macrocyclic intermediate formed by a C1–C11 ring closure. These sesquiterpenes, like the germacrene series discussed below, are distinguished by being directly derived from the all-*trans* FPP (**14**) substrate and not requiring isomerization to the cisoid conformation or nerolidyl diphosphate intermediate. And in this regard the humulane sesquiterpenes might be considered the simplest of sesquiterpenes. However, their biosynthesis utilizes what appears to the same conserved mechanisms as imposed by all terpene synthases as illustrated by the in-depth studies of pentalenene synthase.

The chemical rationalization put forward by Cane[2] for pentalenene (**35**) formation is depicted in **Scheme 11**. FPP is initially ionized, generating the carbocation positioned in proximity to the pi orbital between C10 and C11, thus allowing for an anti-Markovnikov ring closure between C1 and C11. Proton abstraction from the humulyl carbocation intermediate (**34**) results in formation of the macrocyclic triene humulene (**36**). A second ring closure between C2 and the carbocation centered at C10 prior to a final proton abstraction is also possible to yield the unique sesquiterpene caryophyllene (**37**), having a *trans*-fused bicyclo[7.2.0]undecane nucleus. If the humulyl carbocation is however constrained to a third possible conformation,



Scheme 11

a 1,2-hydride shift is the preferred next step leading to a ring closure between C2 and C9. A second hydride shift, followed by a third ring closure within the cyclooctene ring, and the final proton abstraction yields the pentalenene final product.

Although investigated extensively using very sophisticated approaches of structural biology[59] and site-directed mutagenesis,[92] the catalytic cascade for pentalenene biosynthesis is largely an inferred pathway. For instance, a humulyl intermediate has not been observed. Instead, mutagenesis informed by the three-dimensional structure for pentalenene synthase[92] and parallel studies with other terpene synthases lead to mutant enzymes that yielded significant levels of germacrene A (**38**), sometimes in combination with protoilludene (**39**), another possible premature termination product, with a direct and proportional decline in pentalenene biosynthesis. The biosynthesis of germacrene A by the mutant enzymes is significant because it suggests a flexibility within the active site allowing for the initial ring closure between C1 and either C10 or C11. This flexibility may result as a consequence of flexibility in the spatial orientation of amino acids making up the surface contour of the active site. A similar argument can be made for the biosynthesis of protoilludene by some of the mutants, although in this case a greater conformational constraint would have to be imposed upon the putative bicyclic intermediate by the mutant enzyme in order for the reaction to proceed to the protoilludene product.

Tantillo and coworkers[70,93–95] have attempted to provide additional insight into several terpene synthase reactions, and in particular for the pentalenene synthase reaction, by applying a quantum chemical approach. Suggested intermediates were first energy-minimized, transition states between intermediates were sought using various density functional algorithms, and then predicted catalytic cascades were assembled. Overall, a net exothermic reaction of 41–52 kcal mol^{-1} for converting the humulyl carbocation intermediate (**34**) to pentalenene was calculated with relatively small energetic barriers between transition states. One consequence of the algorithms used in these calculations however was the suggestion of two catalytic paths, neither of which is identical to that shown in **Scheme 11**. One alternative positions a protoilludyl carbocation intermediate into the reaction cascade prior to a final reorganization of the fused six- and four-membered rings into two five-membered rings. The second proposal inserts a hyperconjugated intermediate, a secoilludyl carbocation, which rearranges to another high-energy intermediate positioning a proton between two C═C double bonds, which then leads to the penultimate pentalenyl carbocation. Gutta and Tantillo[95] recognize the speculative nature of these alternative pathways and considered the original pathway as suggested by Seemann *et al.*[92] more likely based on experimental observations with genetic mutants. Regardless of the catalytic cascade, the calculations by Gutta and Tantillo point to the penultimate step as that having the greatest energetic barrier once the diphosphate substituent has been ionized from FPP, but for the most part, transition barriers to be rather modest and readily overcome. These observations, the authors suggest, indicate that the pentalenene synthase active site architecture not only serves to promote and stabilize specific intermediates and transition states, but it must also impose destabilizing forces to disfavor other intermediates/transition states. The possibility that the proposed hydride shifts might occur by equivalent proton transfers suggests a more direct role of the enzymes in catalysis.[70] In the present case (**Scheme 11**), the 1,2-hydride shift might take place by successive proton transfers to and from the synthase active site with participation of an enzyme-bound humulene intermediate.

These notions have been experimentally tested. Previously, Steele *et al.*[96] made the surprising observation that several terpene synthases from grand fir (*Abies grandis*) catalyzed the generation of over 30 reaction products when incubated with FPP. Because of the multitude of reaction products, these enzymes were described as promiscuous, which Yoshikuni *et al.*[97] interpreted as synthases possessing active sites allowing for a wide range of catalytic cascades. Focusing on the *A. grandis* α-humulene synthase as a template, these investigators first identified 19 putative active sites within the α-humulene synthase on the basis of molecular modeling against the EAS crystal structure.[55] They then introduced mutations into all of these sites, not one mutation at a time, but in a combinatorial fashion and evaluated the reaction products of each by GC-MS. Interestingly, while the wild-type enzyme generated approximately 28% α-humulene, 15% sibirene, 11.8% longifolene, and lesser amounts of many other sesquiterpenes like 3.9% for β-bisabolene, single and combinatorial mutants catalyzed reactions dominated by single reaction products (**Figure 5**). Like Gutta and Tantillo,[95] the work of Yoshikuni *et al.* suggests that the geometry of the active site plays a role in defining not only what reactions are allowed, but what reactions are not allowed.

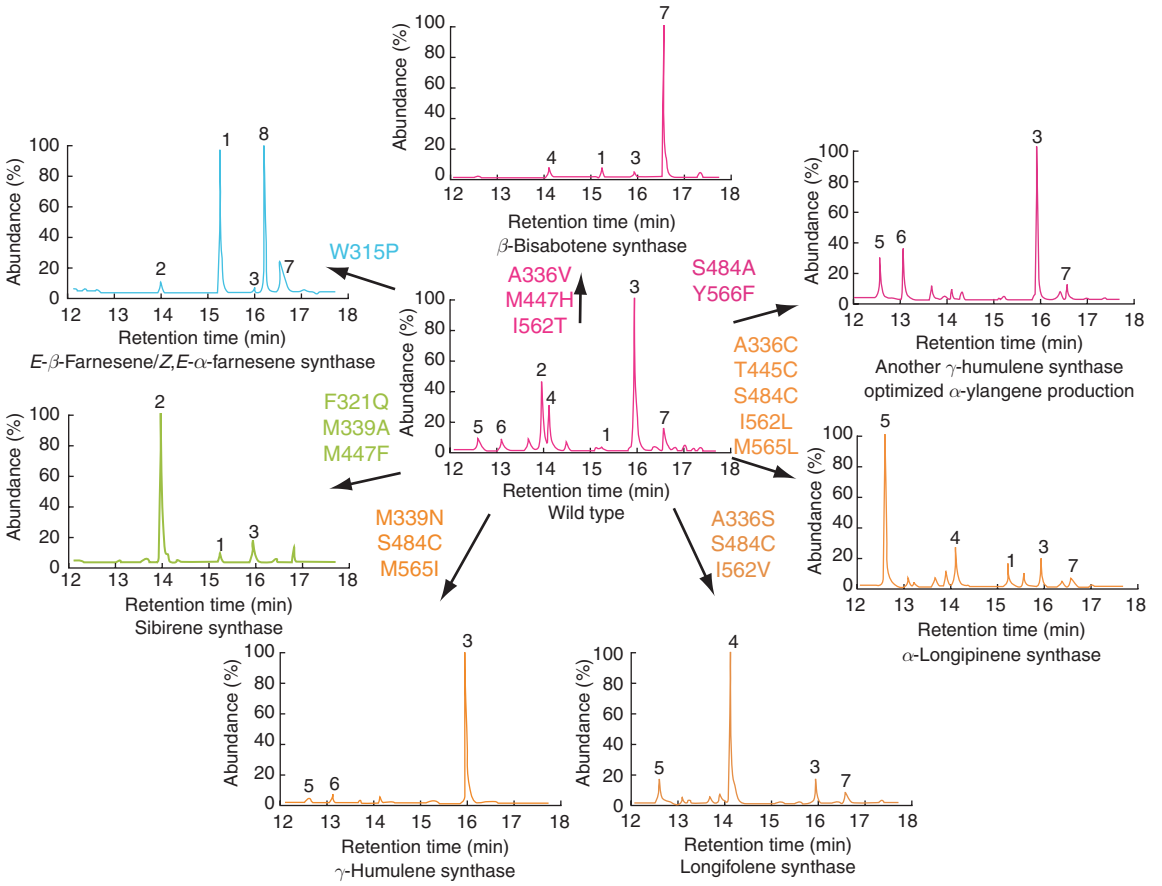
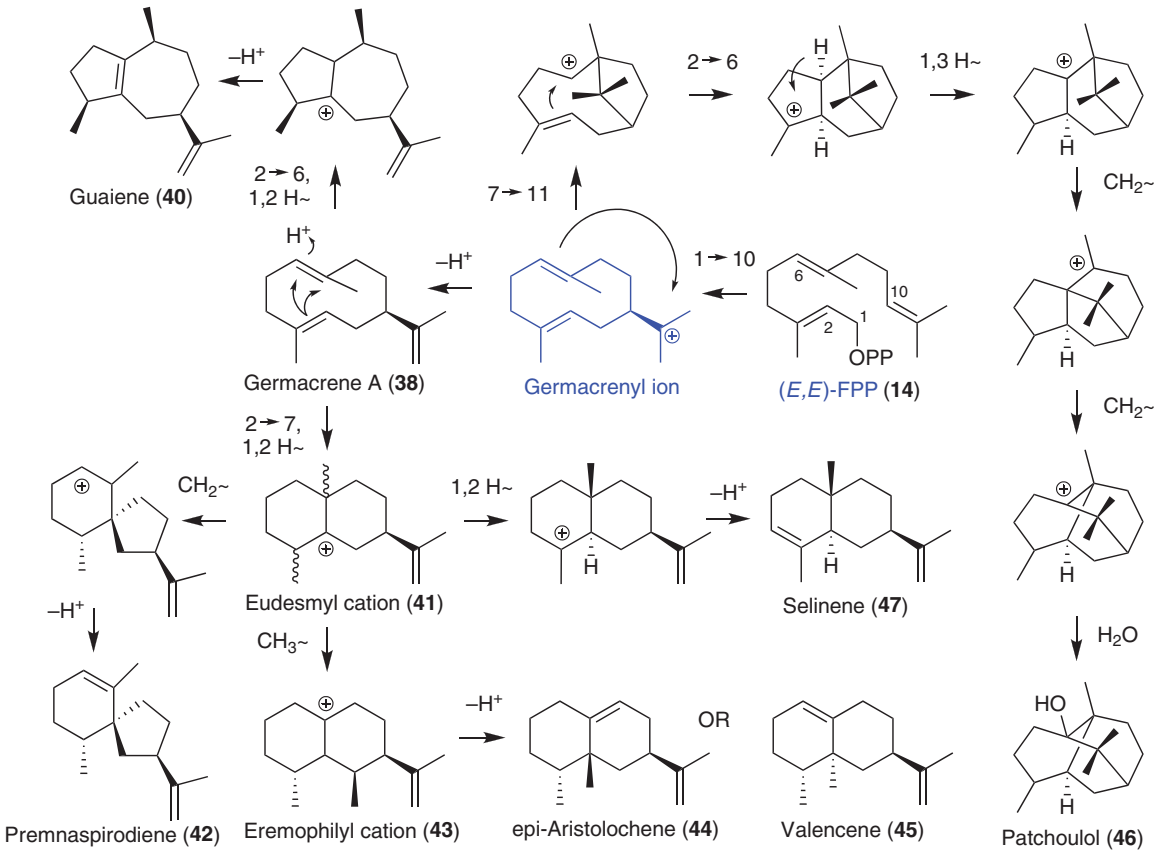


Figure 5 Narrowing down reaction product specificity in mutated forms of α-humulene synthase. Reproduced from Y. Yoshikuni; T. E. Ferrin; J. D. Keasling, *Nature* **2006**, *440* (7087), 1078–1082.

1.16.2.4.5 The germacrane series

Although one may not recognize it, the germacrane-derived sesquiterpenes are some of the most common compounds we interact with in our environment on a daily basis. Patchoulol (**46**) is the key fragrance in many perfumes, colognes, soaps, and all kinds of cleaning products. Nootkatone, the C2 oxidation product of valencene (**45**), is the key citrus flavor in many of our processed foods and soft drinks. Less common, but no less appreciated, are those derivatives like rishitin and capsidiol (**1**), sesquiterpene phytoalexins derived from epi-aristolochene (**44**) by solonaceous plants in response to pathogen challenge and possessing antimicrobial activities.[98] The germacrane sesquiterpenes are also produced by fungi and bacteria. For instance, geosmin (see **Scheme 14**), produced by *Streptomyces coelicolor*, is a major contributor to the odor of soil.[99,100] Aristolochene (**48**, see **Scheme 13**) produced by *Penicillium roqueforti* and *Aspergillus terreus* serves as the precursor to highly oxygenated mycotoxins, like the PR-toxin.[101]

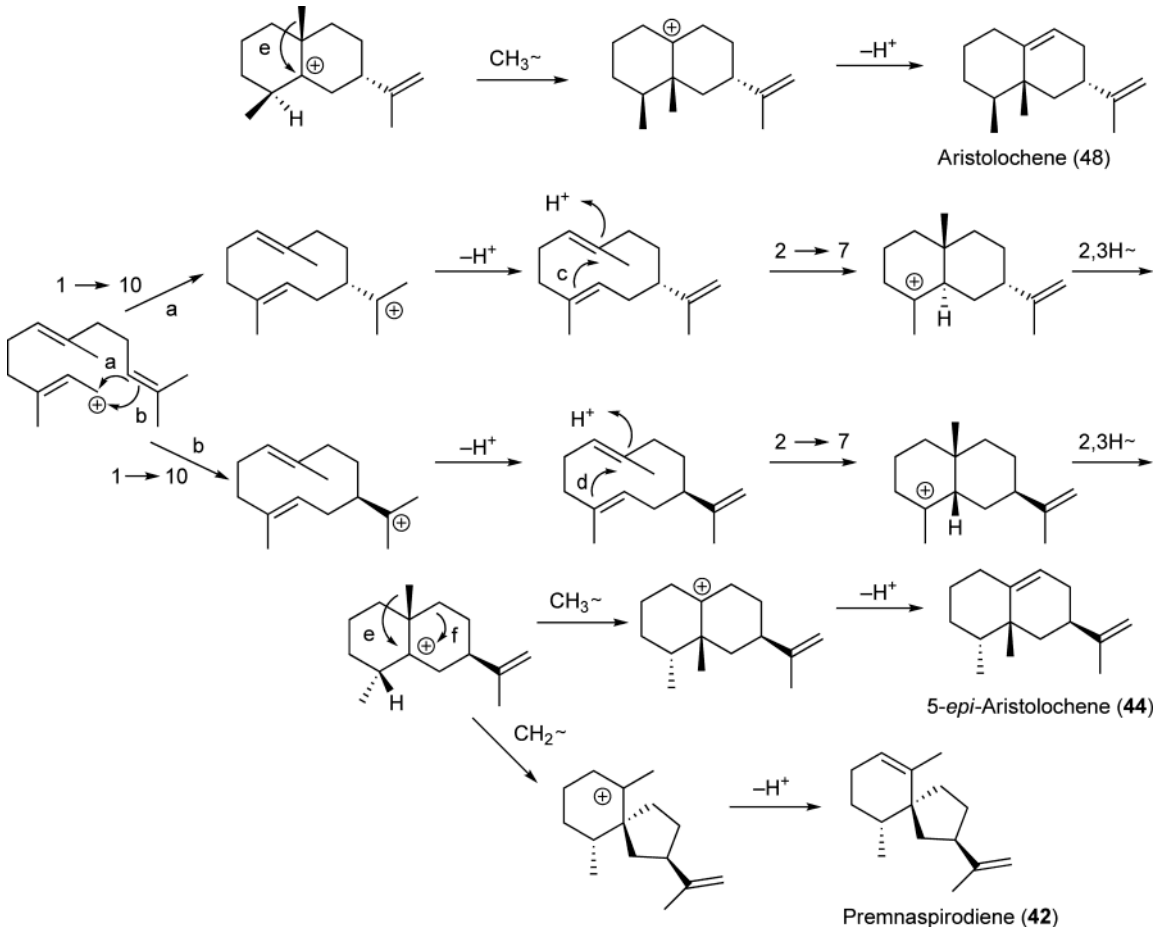
The biosynthesis of the germacrane family of sesquiterpenes has been extensively investigated, accelerated in large part by successes in obtaining and heterologously expressing a wide range of plant and microbial genes coding for the corresponding terpene synthase enzymes.[102–104] These advances have enabled an amazing array of experimental approaches, from presteady-state kinetics[52,53] to solving and rationalizing three-dimensional structures with function.[55,58]. The key chemical entry point into this class of sesquiterpenes is initiated by a C1–C10 ring closure of the all-*trans* farnesyl carbocation. This is an important distinction with other classes of sesquiterpenes, especially those that rely on an isomerization about the C2═C3 double bond prior to the first ring closure reaction. As illustrated in **Scheme 12**, this first ring closure yields an (*E*,*E*,*R*)-germacrenyl carbocation (**39**) that then undergoes a multitude of catalytic pathways to a very diverse range of molecules.



Scheme 12

These include macrocyclic derivatives (i.e., germacranes), bicyclic compounds with fused, bridged, and spiro five-, six-, and seven-membered ring systems (compare guaiene (**40**) and premnaspirodiene (**42**)), bicyclic forms differentiated by how methyl substituents are regio- and stereochemically arrayed about the ring systems (illustrated by the dashed bonds for the eudesmanes and eremophilanes), and compounds consisting of complex bridged ring systems like patchoulol (**46**). The generation of this diversity arises from partial steps or reactions that have already been discussed with regard to other sesquiterpene classes: C–C bond formation mediated by electron flow from double bonds to carbocations; hydride, methyl, and methylene migrations; proton donations and abstractions; and water capture.

The dissection of the reactions leading to aristolochene (**48**), 5-epi-aristolochene (**44**), and premnaspirodiene (**42**) are illustrative of our understanding of the biochemical processes contributing to the germacrane family of sesquiterpenes (**Scheme 13**). Interestingly, some of the first terpene synthases genes to be cloned were those for a fungal aristolochene synthase (AS)[103] and 5-epi-aristolochene synthase (EAS)[105] from a higher plant. Although the two enzymes share very little similarity in primary sequence alignments (<40% similarity), both proteins adopt what has become known as the characteristic 'terpene synthase fold'.[55,56,58] From a chemical perspective, the initial C1–C10 cyclization of the farnesyl carbocation occurs to establish the first chiral center at C10 with the isopropenyl substituent taking on opposite orientations, followed by proton abstractions from the *cis* terminal methyl groups (C13).[105,106] Given the geometric differences between these two reaction intermediates, the enzyme active site surfaces must be very different from one another in order to stabilize one configuration of the germacradienyl intermediate (**39**) versus the other. A similar inference is also suggested to account for the selective stereochemistries created for the methyl substituents at C3 and C7 as a result of the second ring closure reaction.



Scheme 13

Evidence for a germacrenyl intermediate in the initial step of catalysis comes from two different approaches. First, incubation of the fungal AS with 6,7-dihydroFPP, an FPP analog with the C6═C7 double bond saturated, yields 6,7-dihydrogermacrene as the sole reaction product.[107] Similar results were observed using fluorofarnesyl diphosphate analogs with either the plant[108] or fungal[109] enzymes. In both of these cases, a fluorogermacrene A derivative was the only reaction product, an outcome attributed to a premature termination product of catalysis. The second approach relied on molecular modeling according to the crystal structure of EAS, which lead to site-directed mutagenesis of an active site residue positioned to serve as the proton donor initiating the second ring closure steps. When tyrosine 520 was mutated to phenylalanine in EAS, the only reaction product was germacrene A (**38**).[51] An analogous mutation in the *P. roqueforti* AS enzyme, tyrosine 92 to phenylalanine, resulted in germacrene A biosynthesis as well.[110] However, unlike the EAS mutant, the Y92F AS mutant accumulated substantial amounts of aristolochene, prompting further investigations concluding that Y92 was probably not responsible for donating a proton to the germacrene A intermediate and initiating the second series of carbocation-mediated reactions.[111] Instead, acidic residues of the *Pencillium* AS enzyme involved in coordinating and binding the divalent metal cofactors and thought to facilitate binding of the diphosphate substituent of FPP and possibly aligning the allylic tail into the active site, were shown to be crucial for the conversion of germacrene A to aristolochene.[111] Other AS mutants yielded germacrene A in addition to several other unexpected derivatives, providing further evidence for how active site residues might serve to stabilize and promote particular transformations.[112,113]

Abstraction of a proton from the *cis* C13 CH_3 in generating the germacrene A intermediate and the possible recycling of this proton to initiate a second round of catalysis has also been examined. Schenk *et al.*[114] prepared

FPP bearing deuterium and tritium at C13, then, after incubation with EAS or the *Hyoscyamus muticus* premnaspirodiene synthase (HPS), examined the retention of label in epi-aristolochene and premnaspirodiene and its possible redistribution to other positions in the final reaction products by NMR analyses. No recycling for the labels was observed, but a clear secondary kinetic isotope effect on the reaction rate was noted, fully consistent with a concerted cyclization and proton elimination mechanism leading directly from FPP to germacrene A.[114]

The stereoselective incorporation of one deuterium atom into the 6β position of epi-aristolochene following incubation of FPP with EAS in fully deuterated buffer established a *trans* antiperiplanar alignment in the protonation–cyclization of the 6,7 and 2,3 double bonds of the germacrene A intermediate in the conversion to the following *cis*-eudesmyl carbocation intermediate.[114] This work also demonstrated the predominant loss of deuterium after incubation of (*S*)-[8-$^{2}H_{1}$]FPP with EAS, and complementary label retention in the incubation of the enantiomeric labeled FPP, proving that the pseudo-axial allylic hydrogen at C8 is eliminated in forming the ring double bond of epi-aristolochene.

One of the implicit goals driving much of the molecular dissection of terpene biosynthetic enzymes has been that such knowledge might lead to the rational design of enzymes that generate new desired products. While we are still far from such a goal, progress in defining the roles of domains and residues in the catalytic process have been significant.[89,97,115–117] In this regard, the molecular comparisons of EAS, the plant enzyme responsible for epi-aristolochene biosynthesis, to HPS, the enzyme from *H. muticus* responsible for premnaspirodiene production have been particularly useful. As illustrated in **Scheme 13**, the four steps leading up to the *cis*-eudesmyl carbocation intermediate are identical for both these enzymes. They then differ with either a methyl (EAS) or a methylene (HPS) migration, followed by a proton abstraction at C8 (EAS) or C6 (HPS). The EAS and HPS enzymes share greater than 70% primary sequence identity, and very interestingly, separate regions of each protein were shown to mediate the methyl versus methylene migrations in earlier domain-swapping experiments.[115] Based on the crystal structure for the EAS enzyme and molecular modeling of the HPS enzyme upon that scaffold, the active sites cavities for both EAS and HPS were predicted to be identical. This surprising result argued that catalytic specificity must lie outside the active site and most likely with residues that impinge upon the active site residues to alter the active site geometry and architecture.

To discern which residues might be playing a key role in directing methyl versus methylene migrations, Greenhagen *et al.*[116] mapped the active site residues within a van der Waals radius (3.4 Å) of each carbon of a farnesyl analog also resolved within the EAS crystal structure (**Figure 6**). Then the residues outside the active site residues, the so-called second-tier residues, those that were within a van der Waals radius of the first-tier (active site) residues were mapped. Continuation of this contact mapping exercise was extended to a third tier of interacting residues, and this information was then compiled into a comparative algorithm. As shown in a left-hand column of the upper panel of **Figure 6**, the second-tier residues of EAS interacting with the active site residues are arrayed and the specific interactions are denoted by a corresponding solid sphere in the body of the table. The second vertical column in **Figure 6** lists only the second-tier residue differences in HPS generated by molecular model comparisons of HPS and EAS. These differences, highlighted in red, are tracked forward to identify which carbons within FPP are affected. These residues were then targeted for mutagenesis. That is, the identified residues within EAS were mutated to those found in HPS at the corresponding positions, and the converse mutations were effected as well.

When nine of the second- and third-tier residue differences in HPS were introduced into EAS, the M9-EAS mutant enzyme synthesized a reaction profile dominated by premnaspirodiene with very little epi-aristolochene (**Figure 7**). Substituting eight of the same positions within HPS for the corresponding amino acid found in EAS was sufficient to convert the M8-HPS reaction profile to that of EAS (**Figure 7**). Both the M9-EAS and M8-HPS enzymes were comparable to the wild-type enzymes in terms of their catalytic properties, that is, K_m and k_{cat} values.

The notion that specific residues in specific positions within the protein regulate discrete reaction coordinates is of course appealing because it suggests that diversification of the synthase reactions is simply a matter of what amino acids are in these key positions. Unfortunately, this notion is too simplistic. First, what has been reported is that mutations in these particular positions are sufficient to interconvert the reaction specific to the respective enzymes. That does not exclude the possibility that other combinations might be equally or more effective. In fact, O'Maille *et al.*[118] recently tested this possibility by preparing a combinatorial library of tobacco epi-aristolochene synthase (TEAS) mutants, with various combinations of mutations at these nine positions, and demonstrated that more than one combination of mutants was able to interconvert the reaction specificities between TEAS and HPS.

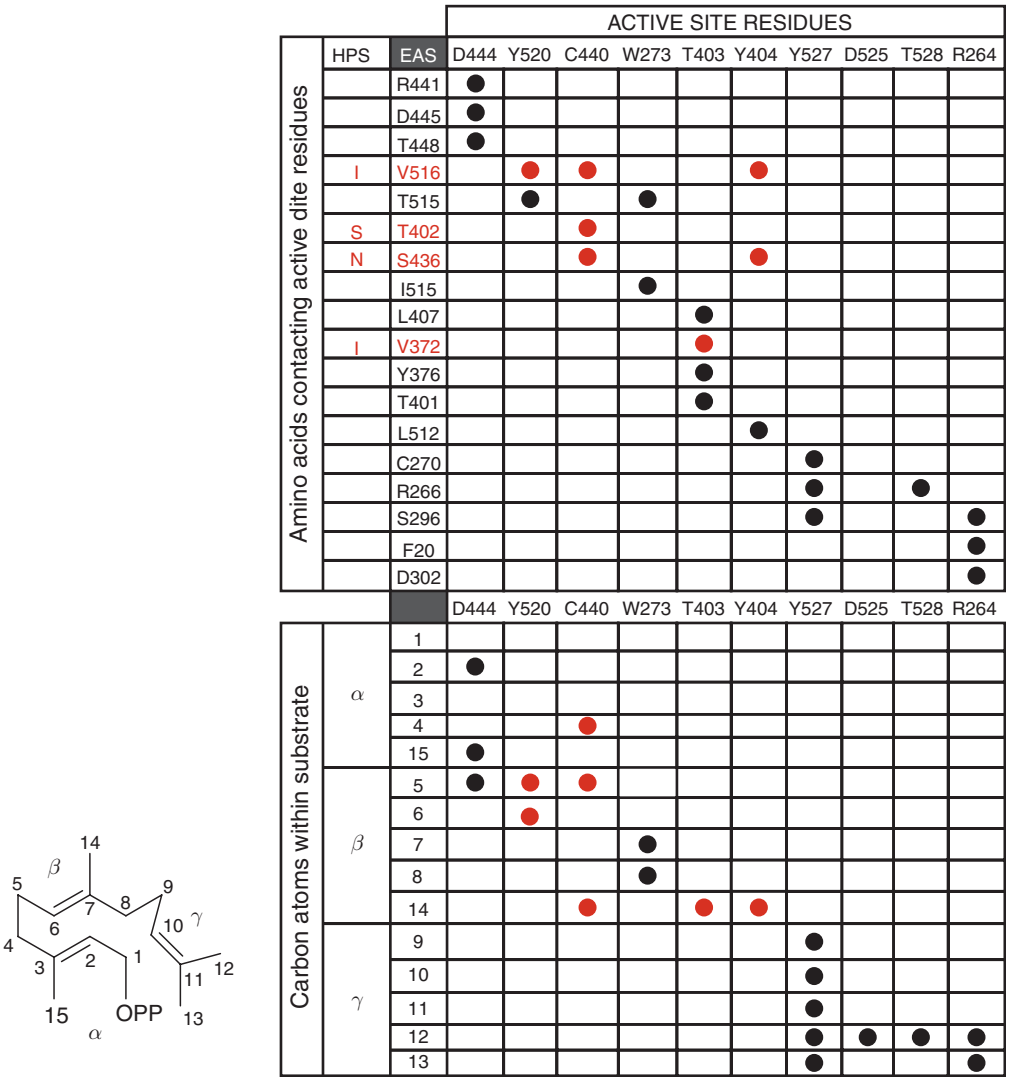
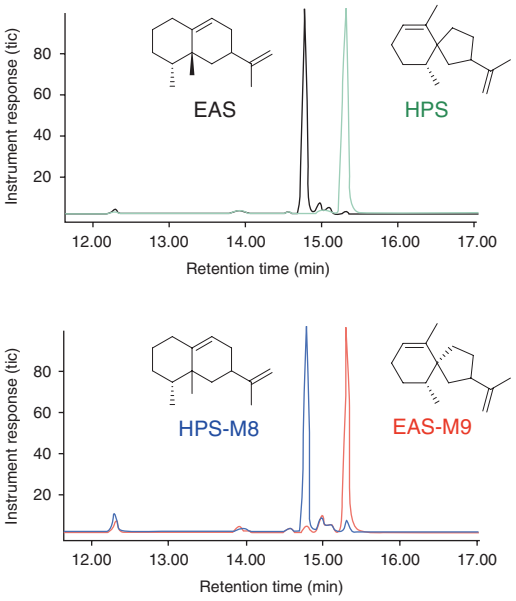


Figure 6 An algorithm developed by Greenhagen *et al.*[116] to map how active site residues of EAS impinge upon specific carbons of FPP (left) bound in the reaction pocket, and then how residues outside the active site, second-tier residues, impinge upon the active site residues.

Equally important, none of the 518 mutants in this collection produced a novel reaction product. That is, the mutants produced products evident in the parental types, but varied in their relative amounts and ratios relative to one another. This would imply that what Greenhagen *et al.*[116] identified were positions within particular synthases important for two particular types of chemical transformations, methyl versus methylene migrations. Whether these positions could play a similar role in synthases catalyzing the biosynthesis of other class of sesquiterpenes, or even how these positions might influence catalysis with substrate analogs modified by regio- and stereochemical variations, remain unknown.

1.16.3 Decorating the Sesquiterpene Scaffolds

Many sesquiterpene hydrocarbons and especially some of those that are synthesized in reactions that terminate the catalytic cascade by water capture forming C_{15} alcohols are potent fragrances. Patchoulol (**46**, **Scheme 12**), already mentioned above, is an extremely popular fragrance in colognes, perfumes, and a wide assortment of



% of the reaction products formed

Synthase/ mutant	epi-Aristolochene	epi-Eremophilene	Premnaspirodiene
EAS	96	<4	<2
EAS-M2	28	70	<2
EAS-M4	16	33	18
EAS-M5	23	24	27
EAS-M7	11	19	43
EAS-M9	3	8	71
HPS	0	<1	98

Figure 7 Identification of the residues controlling reaction product specificity within EAS and HPS. Greenhagen *et al*.[116] identified first-, second-, and third-tier residues serving to create the geometric surface of the active site of two sesquiterpene synthases, EAS and HPS, and then created reciprocal mutants in each. That is, these investigators substituted those contact residues of HPS into EAS and vice versa, creating up to nine mutations (M9), and evaluated each mutant enzyme for its reaction product specificity by GC-MS. Adapted from B. T. Greenhagen; P. E. O'Maille; J. P. Noel; J. Chappell, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (26), 9826–9831.

cleaning reagents. However, in terms of biological activities, some of the most active sesquiterpene scaffolds are further modified, often by at least some form of oxygenation, but also by halogenation,[119] acetylation,[120] and glycosylation.[121] Two notable examples illustrate this point nicely. First, artemisinin (**30**, **Figure 3**) is a highly oxygenated sesquiterpene containing a unique peroxide bridge, derived initially from the cadinene scaffold amorpha-4,11-diene (**33**, **Scheme 10**).[84] Artemisinin is a potent antimalaria drug the use of which is currently limited by its availability. Second, solanaceous plants defend themselves against microbial challenges by producing oxygenated germacrene-derived sesquiterpenes like capsidiol, solavetivone, rishitin, and lubimin.[98] Although progress in defining the biochemical mechanisms responsible for these secondary modifications (decorations) has lagged behind that for the synthases, recent efforts have focused on the identification and characterization the oxygenation reactions. The rationale for this focus is in part because the hydroxyl substituents provide the handles for subsequent modifications, like O-glycosylations and acetylations.

Figure 8 illustrates examples of recently reported oxidation reactions for bacteria, fungi, and plants. Elucidation of the pentalenolactone antibiotic biosynthetic pathway has benefited from detailed bioinformatic inspection of the *Streptomyces avermitilis* genome,[122] and especially the identification of a gene cluster responsible for pentalenolactone biosynthesis. You *et al*.[68] subsequently demonstrated that the ptlH gene within this cluster encodes for a nonheme iron dioxygenase that inserts a hydroxyl function at C11 of only the putative deoxypentalenic acid intermediate, but not pentalenene or a C13 hydroxylated form. Oxidation of sesquiterpenes however is more commonly mediated by cytochrome P-450 oxidases. The albaflavenone antibiotic produced by *S. coelicolor* is initiated by successive hydroxylations of epi-isozizaene at the C4 position to yield the cisoid α,β-unsaturated ketone albaflavenone.[123] The reaction is catalyzed by a typical prokaryotic P-450 enzyme designated CYP170A1, which was found clustered with the corresponding sesquiterpene synthase gene. In contrast to the specificity found in plant P-450 hydroxylases (see 1.16.3), CYP170A1 did not exhibit much specificity for the stereochemistry of the first hydroxylation.

Multifunctional hydroxylases are also found in fungal species. Trichothecene mycotoxins in *Fusarium* species originate from a bisabolene sesquiterpene scaffold, trichodiene (**26**, **Scheme 9** and **Figure 8**) and include hydroxyl groups, oxygen heterocycles, and several O-acetylated substituents.[62,63] The DNA sequence

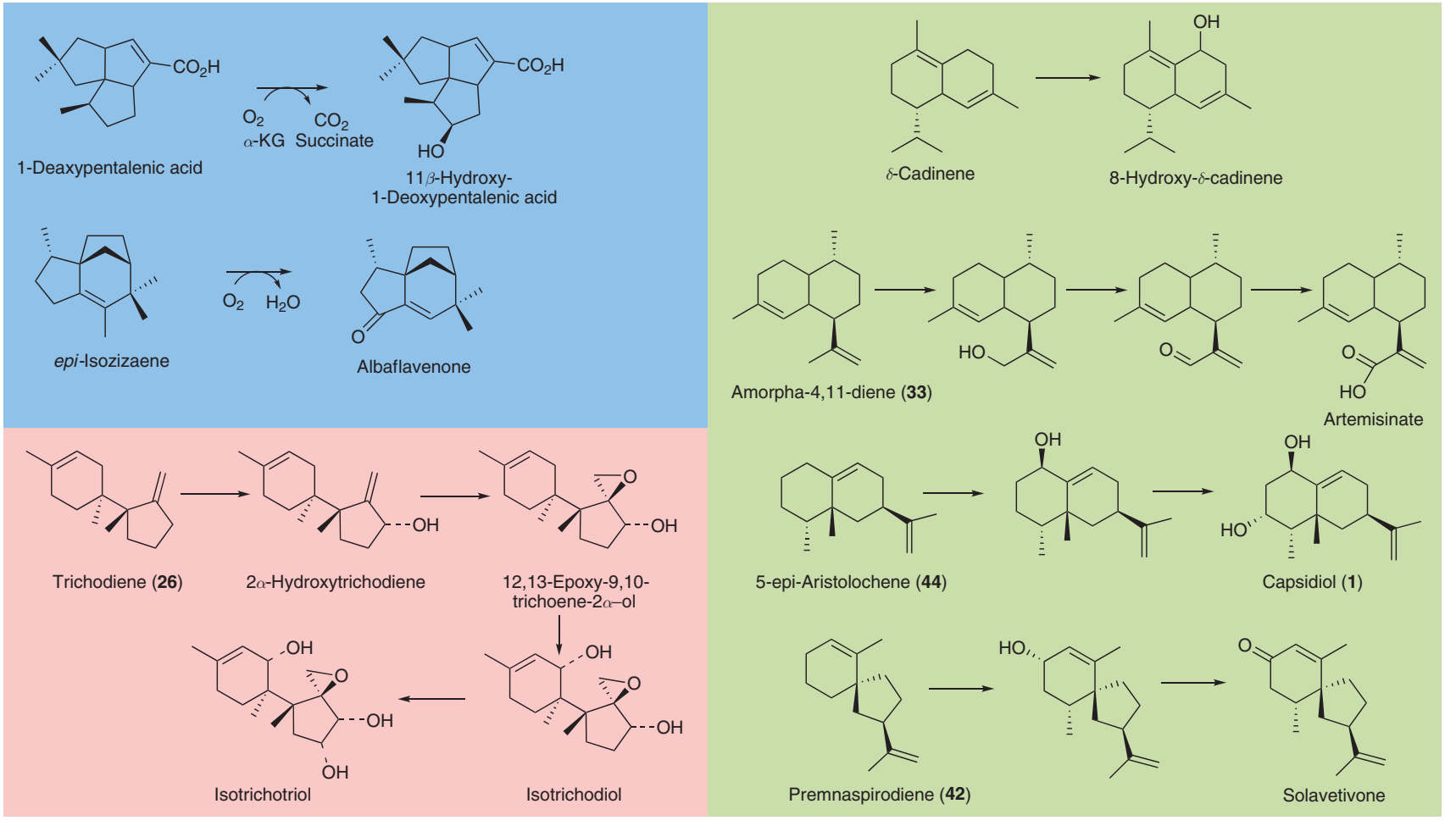


Figure 8 Examples of sesquiterpene oxidation reactions occurring in bacteria (blue panel), fungi (red panel), and plants (green panel). α-KG = α-keto-glutarate.

of the *Fusarium graminearum* genome has been determined, and similar to the sesquiterpene biosynthetic genes in *Streptomyces*, the trichothecene biosynthetic genes also appear to be clustered.[62,63] Interestingly, the Tri4 gene was shown by genetic and biochemical means to catalyze four successive hydroxylations of trichodiene[124,125] (red panel, **Figure 8**). While fungal and plant multifunctional P-450 terpene hydroxylases were recognized earlier,[40,126,127] the Tri4 enzyme is distinctive in catalyzing the introduction of oxygen at very distal positions within the trichodiene molecule and the unusual insertion of oxygen across a double bond to create a spiro-epoxide.

Mono- and multifunctional P-450s acting on sesquiterpene scaffolds are common in plants as well (**Figure 8**, green panel). The hydroxylation of δ-cadinene (**32**) at the C8 position has been reported as possibly the first of several independently regulated biochemical hydroxylations.[128] In comparison, CYP71av1 catalyzes the introduction of a carboxylate group into a terminal methyl of the isopropenyl substituent of amorphadiene (**33**), another cadinene sesquiterpene and intermediate in artemisinin biosynthesis.[86,129]

Hydroxylases/oxygenases responsible for the generation of antimicrobial sesquiterpenes in plants have also been functionally identified. Epi-aristolochene hydroxylase (EAH) has been shown to be responsible for the highly specific successive hydroxylation of 5-epi-aristolochene.[130] In contrast to the multifunction hydroxylase of *S. coelicolor*, the tobacco CYP71D20 enzyme first catalyzes the introduction of a hydroxyl at C1 in the β orientation, followed by stereospecific insertion of a second hydroxyl in the α orientation at C3. In a parallel study, these investigators also demonstrated that premnaspirodiene (**42**), another germacrene-derived scaffold, was successively hydroxylated at C2 by a highly homologous P-450 (to EAH) to yield the spiro α,β-unsaturated ketone, solavetivone (**Figure 8**).[131] Unlike the *Streptomcyes* P-450 responsible for introducing a ketone into epi-zizaene,[123] the *H. muticus* enzyme, hyoscyamus premnaspirodiene oxidase (HPO), exhibited selectivity by introducing the first hydroxyl at C2 in the α orientation.

Using an analogous approach to that of Greenhagen *et al.*[116] for terpene synthases, Takahashi *et al.*[131] began a similar effort for assessing those features of the respective P-450 that might mediate their catalytic specificities. In their initial efforts, molecular models built upon three-dimensional structures for mammalian enzymes were used to direct mutagenesis of putative active site/substrate interacting residues (**Figure 9**). While residues controlling substrate selectivity or the regiospecificity of the hydroxylations were not found, residues dictating kinetic control were. For instance, mutant V482I, A484I improved the k_{cat} of the HPO enzyme for several substrates 10-fold. While these attempts were not successful in mapping residues that control regio- and stereospecificity, they did improve the overall catalytic efficiency of these enzymes, providing new approaches to efforts for improving biological production platforms for modified sesquiterpene products.[132]

Sesquiterpenes also undergo interesting structural modifications by C–C bond cleavage reactions that give rise to norsesquiterpenes, as illustrated by the two examples in **Scheme 14**. Many plants respond to herbivore attack by releasing a mixture of volatile organic compounds from their leaves, which contain high proportions

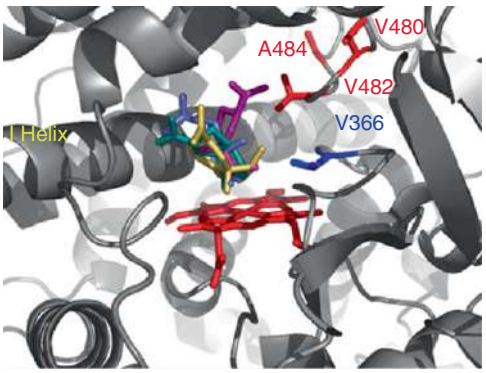
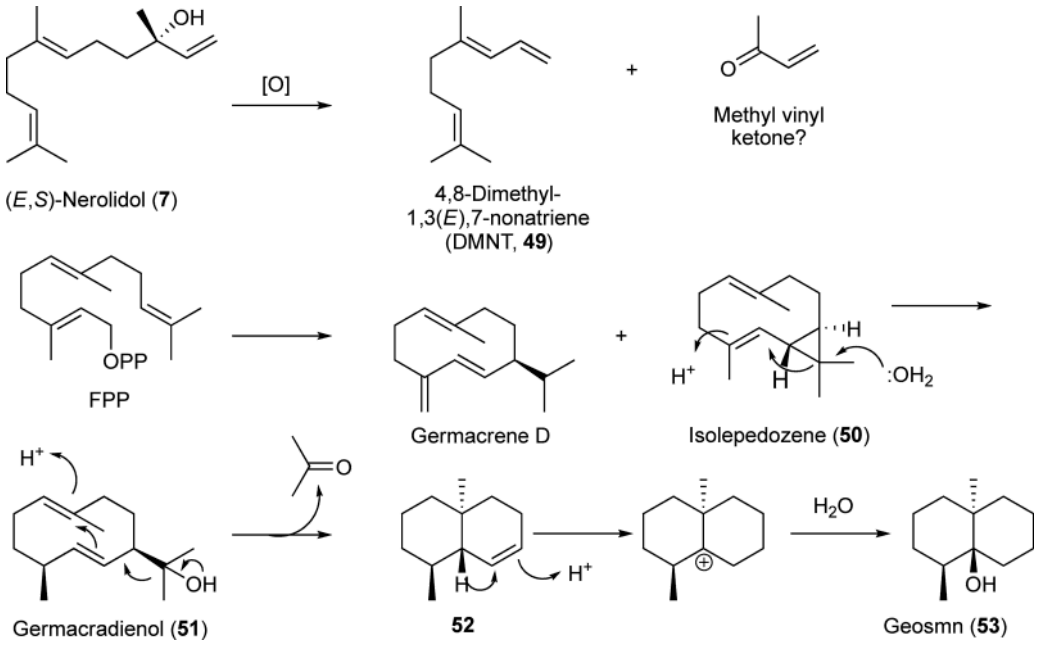


Figure 9 A molecular model of HPO, premnaspirodiene oxidase from *H. muticus*, a cytochrome P-450 enzyme. HPO is responsible for a series of successive hydroxylations at C2 of premnaspirodiene and various orientations of the substrate (cyan, blue, gold, and magenta stick figures) are rendered into the active site, along with the highlighting of residues shown to influence catalysis. Reproduced from S. Takahashi; Y. S. Yeo; Y. X. Zhao; P. E. O'Maille; B. T. Greenhagen; J. P. Noel; R. M. Coates; J. Chappell, *J. Biol. Chem.* **2007**, *282* (43), 31744–31754.



Scheme 14

of the norsesquiterpene (*E*)-4,8-dimethyl-1,3,7-nonatriene (DMNT, **49**) and/or its norditerpene relative (*E,E*)-4,8,12-trimethyl-1,3,7,11-tridecatetraene (TMTT).[133] These polyenes, often referred to inappropriately as 'homoterpenes,' are defensive phytochemicals that attract predators of the herbivore insects. DMNT and TMTT are biosynthesized in plant leaves by stereoselective oxidative cleavage of (*E,S*)-nerolidol (**7**),[134] and presumably the same enantiomer of the diterpene precursor, geranyllinalool. Deuterium-labeling experiments established a high stereoselectivity for removal of the pro S allylic hydrogen at C5 of (*E,S*)-nerolidol and a syn-periplanar alignment of the C–H and C–C bonds cleaved in this fragmentation reaction.[135,136] The expected by-product methyl vinyl ketone has apparently never been detected in the volatile emissions, prompting the suggestion of alternative pathways proceeding through geranylacetone and farnesylacetone intermediates, which then undergo oxidative cleavage to DMNT and TMTT with acetate as the by-product.[135] Although there seems to be no information available about the nature of the novel oxidases that catalyze the C–C cleavage reactions, nerolidol synthases associated with production of DMNT in cucumber (*Cucumis sativus* L.) and lima bean (*Phaseolus lunatus* L.) have been identified,[137] and a gene coding for geranyllinalool synthase in *Arabidopsis thaliana* has been characterized.[138] These enzymes are upregulated in response to herbivore attack and catalyze stereospecific hydrolysis of (*E,E*)-FPP and (*E,E,E*)-GGPP to the respective tertiary alcohols as substrates for the cleavage enzymes.

Geosmin is a rather simple dimethyldecalol (**53**) responsible for the characteristic strong odor of moist soil and associated with unpleasant off-flavors of water, wine, and fish. This trisnorsesquiterpene alcohol is biosynthesized by many microbes, including most *Streptomyces* species. Jiang and Cane[99] recently characterized a single 726-amino acid protein from *S. coelicolor*, which catalyzes the remarkable multistep conversion of (*E,E*)-FPP to geosmin by way of the *trans*-fused bicyclogermacrene, isolepidozene (**50**), (*E,E*)-germacradienol (**51**), and dimethyloctalin (**52**) intermediates (**Scheme 14**). The nonoxidative Prins-type fragmentation of the germacradienol intermediate is triggered by proton-induced cyclization of the (*E,E*)-1,6-cyclodecadiene moiety generating the dimethyloctalin (**52**) and a molecule of acetone. The N-terminal domain of the bifunctional germacradienol-geosmin synthase catalyzes the initial cyclizations of FPP while the C-terminal domain effects the Prins fragmentation step and the proton-induced conversion of **52** to geosmin.

A number of C16 and C17 homosesquiterpenes bearing one or two extra methyl groups that act as hormones or pheromones have been identified (**Figure 10**). The epoxy homofarnesoate esters are designated juvenile hormones I and II (JH I and II) since they, and the corresponding epoxy farnesoate parent (JH III), function by

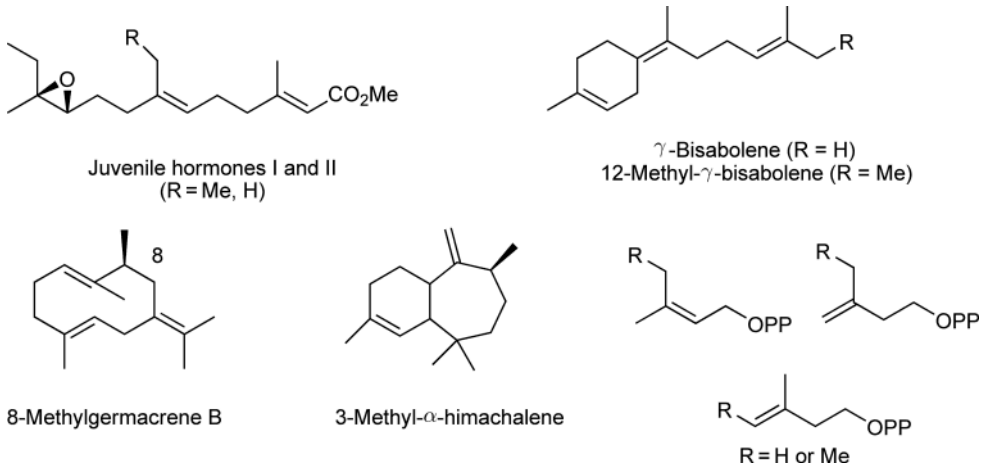


Figure 10 Structures of homosesquiterpenes and homo analogs of dimethylallyl and isopentenyl diphosphates.

maintaining the larval stage of the Cecropia silkworm, *Manduca sexta*.[139] 12-Methyl-γ-bisabolene[140] is a major component of the male sex pheromone for the cereal pest *Eurygaster integriceps Puton*, and the methyl-substituted homologs of germacrene B[141] and α-himachelene[142] are sex pheromones for the Brazilian sandfly *Lutzomyia longipalpis*. JH I and II, and presumably the other known homosesquiterpenes, are biosynthesized by the actions of oxidases and cyclases on the appropriate homologs of FPP, which in turn are produced by prenyl synthases from various combinations of methyl-substituted DMAPPs and IPPs derived from homomevalonate.[139]

1.16.4 Future Challenges

Much has been learned about the chemistry and biochemistry of sesquiterpenes in microbes and plants. The ability to clone and express sesquiterpene biosynthetic genes in heterologous hosts has been immensely helpful, especially for putting the chemical rationalizations together with biochemical functions of these enzymes. Nonetheless, two features of the enzymological conversion of FPP to such a rich diversity of scaffolds standout as unresolved.

First, there are two major divisions within sesquiterpenes that are differentiated by either initial cyclizations of the 2,3-transoid farnesyl carbocation intermediate or initial cyclizations of the 2,3-cisoid farnesyl isomer derived via nerolidyl PP. While many sesquiterpene synthases employ this isomerization activity, many do not. Perhaps adding more confusion to this matter, some synthases appear to utilize both mechanisms. Defining the enzymological features, the amino acids and their respective position within the active site pocket, responsible for the isomerization reaction remain major challenges to the field of sesquiterpene biosynthesis, and this has equally important ramifications for the entire field of isoprenoid biochemistry. This is because sesquiterpenes and isoprenoids in general are known to possess biological activities and these activities are due to the rich stereo- and regiochemistry of these molecules. Elucidating the biochemical mechanisms that impose these features will provide us with tools to better understand, for example, how evolutionary changes came about to give microbes and plants the abilities to occupy ecological niches dependent on their communications with their environment via chemical signals and cues.

A second major challenge is to develop better informatic tools, tools that can identify genes with their true biochemical functions. As is, terpene synthases are readily identified by consensus elements found in the primary sequence of each. The DDXXD motif, a domain associated with metal cofactor binding and coordinating the binding of allylic diphosphate substrates via the negatively charged diphosphate substituent. Sesquiterpene synthases are often differentiated from mono- and di-terpene synthases because they lack an amino terminal domain responsible for targeting the enzymes to the chloroplast compartment. The work of Sallaud *et al.*[60] clearly demonstrates how such a simplistic criterion might fail, because they have documented a

sesquiterpene synthase having just such a targeting signal sequence. Beyond the distinction of mono- versus sesqui- versus di-terpene synthase, what is needed is a tool to predict the geometry of the active site that can then be used to predict specific biochemical transformations. Perhaps coupling physical renderings with quantum chemical calculations, like those of the Tantillo group,[93–95] could be used to identify preferred catalytic cascades within newly discovered sesquiterpene synthase gene sequences.

The remarkable discoveries of Sallaud *et al.*[60] about sesquiterpene biosynthesis in the wild tomato *S. habrochaites* discussed in Section 1.16.2.4.1 contradict many current assumptions, that is, that sesquiterpenes are formed only in the cytosolic compartment by the mevalonate pathway, and that the universal substrate for the cyclases is (*E*,*E*)-FPP. The existence of a (*Z*,*Z*)-FPP synthase raises the question whether (*Z*,*Z*)-, (*E*,*Z*)-, and (*Z*,*E*)-FPP might be involved is sesquiterpene biosynthesis in other plants. Might (*Z*)-specific prenyl transferases and cyclases participate in diterpene biosynthesis? Do the (*Z*)-specific prenyl PPs arise from the mevalonate or MEP pathways? What are the stereochemistry and mechanisms of the coupling reactions catalyzed by (*Z*)-specific prenyl transferases and the cyclizations brought about by (*Z*)-specific cyclases? How do the three-dimensional structures and active site residues of (*E*)- and (*Z*)-specific prenyl transferases and cyclases compare? Do (*E*)- and (*Z*)-specific sesquiterpene synthases coexist in the same plants? How can the evolution of the (*Z*)-specific enzymes be rationalized?

Given the value of sesquiterpenes for agricultural, medical, and industrial applications, there have been numerous efforts to develop robust and sustainable production platforms for these chemicals in microbial and plant systems. For instance, Martin *et al.*[143] described how engineering of the MVA pathway into *Escherichia coli*, which natively has only the MEP pathway, and Wu *et al.*[85] described how engineering novel sesquiterpene branch pathways into the plastids of transgenic plants resulted in high-sesquiterpene-producing platforms. While highly significant, these efforts have not yet fully captured an appreciation for the normal regulatory mechanism imposed upon these pathways. The apparent diurnal (circadian?) rhythms of sesquiterpene metabolism[144,145] and the differential contribution of the MVA and MEP pathways to sesquiterpene biosynthesis by various tissues over development time[144,146] are examples of sesquiterpene physiology not well understood, but which must provide advantages to the organisms reliant upon them. Future investigations are likely to provide insights into the importance of these mechanisms for normal growth and developmental processes, and provide new alternatives for the generation of sesquiterpenes for mankind's use.

Glossary

anti-Markovnikov orientation Reactions of C═C double bonds with electrophiles (e.g., carbocations R^+ or H^+) that generate the less stable carbocation.

carbocation intermediates Reactive trivalent carbon species bearing a positive charge that are intermediates in sesquiterpene biosynthetic mechanisms.

cDNA Copy or complimentary DNA prepared from mRNA as the starting template.

configuration A general term referring to the stereochemistry of molecules.

conformations Refers to the spatial arrangement of the carbon atoms and substituents of a molecules differing by rotations about C–C and/or C–X single bonds.

diterpenes Organic natural products having 20-carbon skeletons formed by combination of four isoprene units.

elicitors Compounds released or produced by microbes that are detected by plants and lead to production of phytoalexins.

enantiomer One of two nonsuperimposable mirror image forms, usually of carbon compounds bearing one or more chiral centers.

first-tier residues The amino acid residues lining the three-dimensional active site of enzymes, which could be in contact with the substrate in the enzyme–substrate complex.

GC–MS An analytical instrument consisting of a gas chromatograph coupled to a mass spectrometer.

heterologous expression A term referring to the expression of a gene taken from one organism and introduced into a completely different host cell.

homosesquiterpenes Sesquiterepenes having a carbon skeleton composed of more than 15 carbon atoms.

knockout mutants Mutations in a specific genetic locus or gene, introduced by either chemical means or the insertion/deletion of DNA fragments into the locus.
lysigenous glands Epidermal and subepidermal reticulated cavities found in select plant species that serve as storage chambers for the accumulation of natural products.
methylerythritol phosphate pathway The biosynthetic pathway operating in the plastids that converts pyruvate and glyceraldehyde-3-phosphate into dimethylallyl and isopentenyl phosphates by way of deoxyxylulose-5-phosphate and 4-*C*-methylerythritol-1-phosphate intermediates.
mevalonate pathway The biosynthetic pathway operating in the cytosolic compartment in plants, which converts three molecules of acetyl-CoA into dimethylallyl and isopentenyl diphosphates by way of mevalonic acid.
monoterpenes Organic natural products having 10-carbon skeletons formed by combination of two isoprene units.
mycotoxins Toxic natural products produced by fungal species, often in association with their colonization of plant material.
norsesquiterpenes Sesquiterepenes having a carbon skeleton composed of less than 15 carbon atoms.
phosphatases Hydrolytic enzymes that can hydrolyze the terminal phosphate substituents from phosphorylated intermediates.
phytoalexins Antimicrobial compounds produced by plants in response to microbial attack.
prenyl transferases Synthases that produce acyclic polyprenyl phosphates by condensation dimethylallyl and isopentenyl phosphates, for example, farnesyl phosphate synthase.
Prins fragmentation Reactions that involve heterolytic cleavage of a C–C bond of an alcohol between the α and β positions with respect to a carbocation or a carbon bearing a leaving group.
second- and third-tier residues The amino acid residues forming concentric layers of residues impinging upon those residues making up the active site of an enzyme.
sesquiterpenes Fifteen-carbon compounds formed by combinations of three isoprene units, which are produced by the isoprenoid biosynthetic pathway.
site-directed mutagenesis A technical means for introducing fragments of nonsense DNA into a specific genetic site.
stereochemistry A term usually referring to the three-dimensional structures of molecules.
terpene hydroxylases Enzymes catalyzing the introduction of oxygen substituents into the hydrocarbon scaffolds of linear and cyclized mono-, sesqui-, and di-terpenes.
terpene synthase fold The α-helical fold that characterizes the three-dimensional structures of terpene synthases in the solid state.
terpene synthases Enzymes catalyzing the initial ionization/cyclization reactions using prenyl diphosphate intermediates. These enzymes are also called ‘‘cyclases’’.
transgenic plants Plants that have been genetically engineered with novel genes.
trichomes Epidermal appendages on plant leaves, stems, petals, and other above-ground plant tissues that synthesize unique arrays of natural products, some of which provide benefit for plant–environment interactions.
triterpenes Organic natural products having 30-carbon skeletons formed by combination of six isoprene units.
Wagner–Meerwein rearrangements Skeletal reorganizations of carbocations occurring by 1,2-migrations of a carbon substituent adjacent to the cationic site, usually generating a more stable carbocation from a less stable one.

References

1. B. M. Fraga, *Nat. Prod. Rep.* **2008**, *25* (6), 1180–1209.
2. D. E. Cane, *Chem. Rev.* **1990**, *90* (7), 1089–1103.
3. D. E. Cane, Sesquiterpene Biosynthesis. In *Comprehensive Natural Product Chemistry*; D. Barton, K. Nakanishi, O. Meth-Cohn, Eds.; Isoprenoids Including Carotenoids and Steroids; Elsevier: Amsterdam, 1999; Vol. 2, pp 155–200.

4. T. M. Hohn, Cloning and Expression of Terpene Synthase Genes. In *Comprehensive Natural Products Chemistry*; D. Barton, K. Nakanishi, O. Meth-Cohn, Eds.; Isoprenoids Including Cartoenoids and Steroids; Elsevier: Amsterdam, 1999; Vol. 2, pp 201–215.
5. D. W. Christianson, *Chem. Rev.* **2006**, *106* (8), 3412–3442.
6. A. Stoessl; C. H. Unwin; E. W. B. Ward, *Phytopathol. Z.* **1972**, *74* (2), 141–152.
7. B. C. Tan; S. H. Schwartz; J. A. D. Zeevaart; D. R. McCarty, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94* (22), 12235–12240.
8. S. H. Schwartz; B. C. Tan; D. A. Gage; J. A. D. Zeevaart; D. R. McCarty, *Science* **1997**, *276* (5320), 1872–1874.
9. M. E. Auldridge; D. R. McCarty; H. J. Klee, *Curr. Opin. Plant Biol.* **2006**, *9* (3), 315–321.
10. J. von Lintig; S. Hessel; A. Isken; C. Kiefer; J. M. Lampert; O. Voolstra; K. Vogt, *Biochim. Biophys. Acta, Mol. Basis Dis.* **2005**, *1740* (2), 122–131.
11. Z. Sun; J. Hans; M. H. Walter; R. Matusova; J. Beekwilder; F. W. A. Verstappen; Z. Ming; E. van Echtelt; D. Strack; T. Bisseling; H. J. Bouwmeester, *Planta* **2008**, *228* (5), 789–801.
12. C. E. Cook; L. P. Whichard; M. E. Wall; G. H. Egley; P. Coggon; P. A. Luhan; A. T. McPhail, *J. Am. Chem. Soc.* **1972**, *94* (17), 6198–6199.
13. V. Gomez-Roldan; S. Fermas; P. B. Brewer; V. Puech-Pages; E. A. Dun; J. P. Pillot; F. Letisse; R. Matusova; S. Danoun; J. C. Portais; H. Bouwmeester; G. Becard; C. A. Beveridge; C. Rameau; S. F. Rochange, *Nature* **2008**, *455* (7210), 189–194.
14. M. Umehara; A. Hanada; S. Yoshida; K. Akiyama; T. Arite; N. Takeda-Kamiya; H. Magome; Y. Kamiya; K. Shirasu; K. Yoneyama; J. Kyozuka; S. Yamaguchi, *Nature* **2008**, *455* (7210), 195–200.
15. R. G. Gardner; H. Shan; S. P. T. Matsuda; R. Y. Hampton, *J. Biol. Chem.* **2001**, *276* (12), 8681–8694.
16. C. C. Correll; L. Ng; P. A. Edwards, *J. Biol. Chem.* **1994**, *269* (26), 17390–17393.
17. L. S. Song, *Appl. Biochem. Biotechnol.* **2006**, *128* (2), 149–157.
18. L. Thai; J. S. Rush; J. E. Maul; T. Devarenne; D. L. Rodgers; J. Chappell; C. J. Waechter, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (23), 13080–13085.
19. A. Aharoni; A. P. Giri; F. W. A. Verstappen; C. M. Bertea; R. Sevenier; Z. K. Sun; M. A. Jongsma; W. Schwab; H. J. Bouwmeester, *Plant Cell* **2004**, *16* (11), 3110–3131.
20. C. Schnee; T. G. Köllner; J. Gershenzon; J. Degenhardt, *Plant Physiol.* **2002**, *130* (4), 2049–2060.
21. J. Crock; M. Wildung; R. Croteau, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94* (24), 12833–12838.
22. D. A. Nagegowda; M. Gutensohn; C. G. Wilkerson; N. Dudareva, *Plant J.* **2008**, *55* (2), 224–239.
23. P. Wohlers, *Z. Angew. Entomol. J. Appl. Entomol* **1981**, *92* (4), 329–336.
24. C. Schnee; T. G. Köllner; M. Held; T. C. J. Turlings; J. Gershenzon; J. Degenhardt, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (4), 1129–1134.
25. S. M. Njoroge; H. Koaze; P. N. Karanja; M. Sawamura, *Flavour Fragr. J.* **2005**, *20* (1), 80–85.
26. A. Fontana; C. Muniain; G. Cimino, *J. Nat. Prod.* **1998**, *61* (8), 1027–1029.
27. Y. Kakou; P. Crews; G. J. Bakus, *J. Nat. Prod.* **1987**, *50* (3), 482–484.
28. D. P. W. Huber; R. Gries; J. H. Borden; H. D. Pierce, *Chemoecology* **2000**, *10* (3), 103–113.
29. K. Oba; K. Oga; I. Uritani, *Phytochemistry* **1982**, *21* (8), 1921–1925.
30. N. W. Gaikwad; K. M. Madyastha, *Biochem. Biophys. Res. Commun.* **2002**, *290* (1), 589–594.
31. B. J. M. Jansen; A. de Groot, *Nat. Prod. Rep.* **2004**, *21* (4), 449–477.
32. T. Hashimoto; M. Tori; Y. Asakawa, *Phytochemistry* **1989**, *28* (12), 3377–3381.
33. Y. Asakawa; G. W. Dawson; D. C. Griffiths; J. Y. Lallemand; S. V. Ley; K. Mori; A. Mudd; M. Pezechkleclaire; J. A. Pickett; H. Watanabe; C. M. Woodcock; Z. N. Zhang, *J. Chem. Ecol.* **1988**, *14* (10), 1845–1855.
34. A. Cyr; P. R. Wilderman; M. Determan; R. J. Peters, *J. Am. Chem. Soc.* **2007**, *129* (21), 6684–6685.
35. I. Abe, *Nat. Prod. Rep.* **2007**, *24* (6), 1311–1331.
36. T. Toyomasu; R. Nuda; H. Kenmoku; Y. Kanno; S. Miura; C. Nakano; Y. Shiono; W. Mitsuhashi; H. Toshima; H. Oikawa; T. Hoshino; T. Dairi; N. Kato; T. Sassa, *Biosci. Biotechnol. Biochem.* **2008**, *72* (4), 1038–1047.
37. R. R. Fall; C. A. West, *J. Biol. Chem.* **1971**, *246* (22), 6913–6928.
38. J. D. Duncan; C. A. West, *Plant Physiol.* **1981**, *68* (5), 1128–1134.
39. J. MacMillan; M. H. Beale, Diterpene Biosynthesis. In *Comprehensive Natural Products Chemistry*; D. Barton, K. Nakanishi, O. Meth-Cohn, Eds.; Elsevier: Amsterdam, 1999; Vol. 2, pp 217–243.
40. C. A. West, Biosynthesis of Diterpenes. In *Biosynthesis of Isoprenoid Compounds*; J. W. Porter, S. L. Spurgeon, Eds.; Wiley: New York, 1981; Vol. 1, pp 375–411.
41. E. J. Corey; S. P. T. Matsuda; B. Bartel, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90* (24), 11628–11632.
42. E. J. Corey; S. P. T. Matsuda; B. Bartel, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91* (6), 2211–2215.
43. D. V. Banthorpe; J. T. Brown; G. S. Morris, *Phytochemistry* **1992**, *31* (10), 3391–3395.
44. C. Landmann; B. Fink; M. Festner; M. Dregus; K. H. Engel; W. Schwab, *Arch. Biochem. Biophys.* **2007**, *465* (2), 417–429.
45. J. Bohlmann; J. Crock; R. Jetter; R. Croteau, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95* (12), 6756–6761.
46. D. M. Martin; J. Fäldt; J. Bohlmann, *Plant Physiol.* **2004**, *135* (4), 1908–1927.
47. Y. Iijima; R. Davidovich-Rikanati; E. Fridman; D. R. Gang; E. Bar; E. Lewinsohn; E. Pichersky, *Plant Physiol.* **2004**, *136* (3), 3724–3736.
48. R. Davidovich-Rikanati; E. Lewinsohn; E. Bar; Y. Iijima; E. Pichersky; Y. Sitrit, *Plant J.* **2008**, *56* (2), 228–238.
49. C. G. Jones; E. L. Ghisalberti; J. A. Plummer; E. Barbour, *Phytochemistry* **2006**, *67* (22), 2463–2468.
50. T. G. Köllner; C. Schnee; S. Li; A. Svatos; B. Schneider; J. Gershenzon; J. Degenhardt, *J. Biol. Chem.* **2008**, *283* (30), 20779–20788.
51. K. A. Rising; C. M. Starks; J. P. Noel; J. Chappell, *J. Am. Chem. Soc.* **2000**, *122* (9), 1861–1866.
52. D. E. Cane; H. T. Chiu; P. H. Liang; K. S. Anderson, *Biochemistry* **1997**, *36* (27), 8332–8339.
53. J. R. Mathis; K. Back; C. Starks; J. Noel; C. D. Poulter; J. Chappell, *Biochemistry* **1997**, *36* (27), 8340–8348.
54. T. G. Köllner; P. E. O'Maille; N. Gatto; W. Boland; J. Gershenzon; J. Degenhardt, *Arch. Biochem. Biophys.* **2006**, *448* (1–2), 83–92.
55. C. M. Starks; K. W. Back; J. Chappell; J. P. Noel, *Science* **1997**, *277* (5333), 1815–1820.

56. E. Y. Shishova; L. Di Costanzo; D. E. Cane; D. W. Christianson, *Biochemistry* **2007**, *46* (7), 1941–1951.
57. M. J. Rynkiewicz; D. E. Cane; D. W. Christianson, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (24), 13543–13548.
58. J. M. Caruthers; I. Kang; M. J. Rynkiewicz; D. E. Cane; D. W. Christianson, *J. Biol. Chem.* **2000**, *275* (33), 25533–25539.
59. C. A. Lesburg; G. Z. Zhai; D. E. Cane; D. W. Christianson, *Science* **1997**, *277* (5333), 1820–1824.
60. C. Sallaud; D. Rontein; S. Onillon; F. Jabés; P. Duffé; C. Giacalone; S. Thoraval; C. Escoffier; G. Herbette; N. Leonhardt; M. Causse; A. Tissier, *Plant Cell* **2009**, *21*, 301–317.
61. K. Besser; A. Harper; N. Welsby; I. Schauvinhold; S. Slocombe; Y. Li; R. A. Dixon; P. Broun, *Plant Physiol.* **2009**, *149* (1), 499–514.
62. A. E. Desjardins; T. M. Hohn; S. P. McCormick, *Microbiol. Rev.* **1993**, *57* (3), 595–604.
63. M. Kimura; T. Tokai; N. Takahashi-Ando; S. Ohsato; M. Fujimura, *Biosci. Biotechnol. Biochem.* **2007**, *71* (9), 2105–2123.
64. R. W. Wannemacher, Jr.; S. T. Wiener, Trichothecene Mycotoxins. In *Medical Aspects of Chemical and Biological Warfare*; F. R. Sidell, E. T. Takafuji, D. R. Franz, Eds.; U.S. Government Fiche – D 104.35:PT.1; 1997; pp 655–676.
65. N. A. Foroud; F. Eudes, *Int. J. Mol. Sci.* **2009**, *10* (1), 147–173.
66. E. Cundliff; M. Cannon; J. Davies, *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71* (1), 30–34.
67. L. J. Harris; S. C. Gleddie, *Physiol. Mol. Plant Pathol.* **2001**, *58* (4), 173–181.
68. Z. You; S. Omura; H. Ikeda; D. E. Cane, *J. Am. Chem. Soc.* **2006**, *128* (20), 6566–6567.
69. S. Wu; M. A. Schoenbeck; B. T. Greenhagen; S. Takahashi; S. B. Lee; R. M. Coates; J. Chappell, *Plant Physiol.* **2005**, *138* (3), 1322–1333.
70. Y. J. Hong; D. J. Tantillo, *Org. Lett.* **2006**, *8* (20), 4601–4604.
71. L. S. Vedula; Y. X. Zhao; R. M. Coates; T. Koyama; D. E. Cane; D. W. Christianson, *Arch. Biochem. Biophys.* **2007**, *466* (2), 260–266.
72. D. E. Cane; T. E. Bowser, *Bioorgan. Med. Chem. Lett.* **1999**, *9* (8), 1127–1132.
73. D. E. Cane; G. H. Yang, *J. Org. Chem.* **1994**, *59* (19), 5794–5798.
74. D. E. Cane; G. H. Yang; R. M. Coates; H. J. Pyun; T. M. Hohn, *J. Org. Chem.* **1992**, *57* (12), 3454–3462.
75. L. S. Vedula; D. E. Cane; D. W. Christianson, *Biochemistry* **2005**, *44* (38), 12719–12727.
76. L. S. Vedula; J. Y. Jiang; T. Zakharian; D. E. Cane; D. W. Christianson, *Arch. Biochem. Biophys.* **2008**, *469* (2), 184–194.
77. M. J. Rynkiewicz; D. E. Cane; D. W. Christianson, *Biochemistry* **2002**, *41* (6), 1732–1741.
78. Y. Asakawa, *Phytochemistry* **2004**, *65* (6), 623–669.
79. D. Tholl; F. Chen; J. Petri; J. Gershenzon; E. Pichersky, *Plant J.* **2005**, *42* (5), 757–771.
80. G. T. Bottger; E. T. Sheehan; M. J. Lukefahr, *J. Econ. Entomol.* **1964**, *57* (2), 283–285.
81. G. Sunilkumar; L. M. Campbell; L. Puckhaber; R. D. Stipanovic; K. S. Rathore, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (48), 18054–18059.
82. B. J. Townsend; A. Poole; C. J. Blake; D. J. Llewellyn, *Plant Physiol.* **2005**, *138* (1), 516–528.
83. K. J. Abraham; M. L. Pierce; M. Essenberg, *Phytochemistry* **1999**, *52* (5), 829–836.
84. P. S. Covello, *Phytochemistry* **2008**, *69* (17), 2881–2885.
85. S. Q. Wu; M. Schalk; A. Clark; R. B. Miles; R. M. Coates; J. Chappell, *Nat. Biotechnol.* **2006**, *24* (11), 1441–1447.
86. D. K. Ro; E. M. Paradise; M. Ouellet; K. J. Fisher; K. L. Newman; J. M. Ndungu; K. A. Ho; R. A. Eachus; T. S. Ham; J. Kirby; M. C. Y. Chang; S. T. Withers; Y. Shiba; R. Sarpong; J. D. Keasling, *Nature* **2006**, *440* (7086), 940–943.
87. P. S. Covello; K. H. Teoh; D. R. Polichuk; D. W. Reed; G. Nowak, *Phytochemistry* **2007**, *68* (14), 1864–1871.
88. C. R. Benedict; J. L. Lu; D. W. Pettigrew; J. G. Liu; R. D. Stipanovic; H. J. Williams, *Plant Physiol.* **2001**, *125* (4), 1754–1765.
89. Y. Yoshikuni; V. J. J. Martin; T. E. Ferrin; J. D. Keasling, *Chem. Biol.* **2006**, *13* (1), 91–98.
90. S. Picaud; P. Mercke; X. F. He; O. Sterner; M. Brodelius; D. E. Cane; P. E. Brodelius, *Arch. Biochem. Biophys.* **2006**, *448* (1–2), 150–155.
91. C. L. Anderson; K. Bremer; E. M. Friis, *Am. J. Bot.* **2005**, *92* (10), 1737–1748.
92. M. Seemann; G. Z. Zhai; J. W. de Kraker; C. M. Paschall; D. W. Christianson; D. E. Cane, *J. Am. Chem. Soc.* **2002**, *124* (26), 7681–7689.
93. M. W. Lodewyk; P. Gutta; D. J. Tantillo, *J. Org. Chem.* **2008**, *73* (17), 6570–6579.
94. D. J. Tantillo, *J. Phys. Org. Chem.* **2008**, *21* (7–8), 561–570.
95. P. Gutta; D. J. Tantillo, *J. Am. Chem. Soc.* **2006**, *128* (18), 6172–6179.
96. C. L. Steele; J. Crock; J. Bohlmann; R. Croteau, *J. Biol. Chem.* **1998**, *273* (4), 2078–2089.
97. Y. Yoshikuni; T. E. Ferrin; J. D. Keasling, *Nature* **2006**, *440* (7087), 1078–1082.
98. A. Stoessl; J. B. Stothers; E. W. B. Ward, *Phytochemistry* **1976**, *15* (6), 855–872.
99. J. Y. Jiang; D. E. Cane, *J. Am. Chem. Soc.* **2008**, *130* (2), 428–429.
100. J. Y. Jiang; X. F. He; D. E. Cane, *J. Am. Chem. Soc.* **2006**, *128* (25), 8128–8129.
101. H. H. Jelen, *J. Agric. Food Chem.* **2002**, *50* (22), 6569–6574.
102. D. E. Cane; I. Kang, *Arch. Biochem. Biophys.* **2000**, *376* (2), 354–364.
103. R. H. Proctor; T. M. Hohn, *J. Biol. Chem.* **1993**, *268* (6), 4543–4548.
104. K. Back; J. Chappell, *J. Biol. Chem.* **1995**, *270* (13), 7375–7381.
105. P. J. Facchini; J. Chappell, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89* (22), 11088–11092.
106. D. E. Cane; P. C. Prabhakaran; J. S. Oliver; D. B. McIlwaine, *J. Am. Chem. Soc.* **1990**, *112* (8), 3209–3210.
107. D. E. Cane; Y. S. Tsantrizos, *J. Am. Chem. Soc.* **1996**, *118* (42), 10037–10040.
108. J. A. Faraldos; Y. Zhao; P. E. O'Maille; J. P. Noel; R. M. Coates, *ChemBioChem* **2007**, *8* (15), 1826–1833.
109. D. J. Miller; F. L. Yu; R. K. Allemann, *ChemBioChem* **2007**, *8* (15), 1819–1825.
110. M. J. Calvert; P. R. Ashton; R. K. Allemann, *J. Am. Chem. Soc.* **2002**, *124* (39), 11636–11641.
111. B. Felicetti; D. E. Cane, *J. Am. Chem. Soc.* **2004**, *126* (23), 7212–7221.
112. A. Deligeorgopoulou; S. E. Taylor; S. Forcat; R. K. Allemann, *Chem. Commun.* **2003**, (17), 2162–2163.
113. S. Forcat; R. K. Allemann, *Chem. Commun.* **2004**, (18), 2094–2095.
114. D. J. Schenk; C. M. Starks; K. R. Manna; J. Chappell; J. P. Noel; R. M. Coates, *Arch. Biochem. Biophys.* **2006**, *448* (1–2), 31–44.
115. K. W. Back; J. Chappell, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93* (13), 6841–6845.

116. B. T. Greenhagen; P. E. O'Maille; J. P. Noel; J. Chappell, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (26), 9826–9831.
117. D. B. Little; R. B. Croteau, *Arch. Biochem. Biophys.* **2002**, *402* (1), 120–135.
118. P. E. O'Maille; A. Malone; N. Dellas; B. A. Hess; L. Smentek; I. Sheehan; B. T. Greenhagen; J. Chappell; G. Manning; J. P. Noel, *Nat. Chem. Biol.* **2008**, *4* (10), 617–623.
119. C. S. Vairappan; M. Suzuki; T. Ishii; T. Okino; T. Abe; M. Masuda, *Phytochemistry* **2008**, *69* (13), 2490–2494.
120. G. S. Garvey; S. P. McCormick; I. Rayment, *J. Biol. Chem.* **2008**, *283* (3), 1660–1669.
121. C. Karamenderes; E. Bedir; R. Pawar; S. Baykan; K. A. Khan, *Phytochemistry* **2007**, *68* (5), 609–615.
122. H. Ikeda; J. Ishikawa; A. Hanamoto; M. Shinose; H. Kikuchi; T. Shiba; Y. Sakaki; M. Hattori; S. Omura, *Nat. Biotechnol.* **2003**, *21* (5), 526–531.
123. B. Zhao; X. Lin; L. Lei; D. C. Lamb; S. L. Kelly; M. R. Waterman; D. E. Cane, *J. Biol. Chem.* **2008**, *283* (13), 8183–8189.
124. T. Tokai; H. Koshino; N. Takahashi-Ando; M. Sato; M. Fujimura; M. Kimura, *Biochem. Biophys. Res. Commun.* **2007**, *353* (2), 412–417.
125. T. M. Hohn; A. E. Desjardins; S. P. McCormick, *Mol. Gen. Genet.* **1995**, *248* (1), 95–102.
126. B. Tudzynski; M. C. Rojas; P. Gaskin; P. Hedden, *J. Biol. Chem.* **2002**, *277* (24), 21246–21253.
127. C. A. Helliwell; A. Poole; W. J. Peacock; E. S. Dennis, *Plant Physiol.* **1999**, *119* (2), 507–510.
128. P. Luo; Y. H. Wang; G. D. Wang; M. Essenberg; X. Y. Chen, *Plant J.* **2001**, *28* (1), 95–104.
129. K. H. Teoh; D. R. Polichuk; D. W. Reed; G. Nowak; P. S. Covello, *FEBS Lett.* **2006**, *580* (5), 1411–1416.
130. S. Takahashi; Y. Zhao; P. E. O'Maille; B. T. Greenhagen; J. P. Noel; R. M. Coates; J. Chappell, *J. Biol. Chem.* **2005**, *280* (5), 3686–3696.
131. S. Takahashi; Y. S. Yeo; Y. Zhao; P. E. O'Maille; B. T. Greenhagen; J. P. Noel; R. M. Coates; J. Chappell, *J. Biol. Chem.* **2007**, *282* (43), 31744–31754.
132. S. Takahashi; Y. Yeo; B. T. Greenhagen; T. McMullin; L. Song; J. Maurina-Brunker; R. Rosson; J. P. Noel; J. Chappell, *Biotechnol. Bioeng.* **2007**, *97* (1), 170–181.
133. W. Boland; Z. Feng; J. Donath; A. Gabler, *Naturwissenschaften* **1992**, *79* (8), 368–371.
134. J. Donath; W. Boland, *Phytochemistry* **1995**, *39* (4), 785–790.
135. W. Boland; A. Gabler; M. Gilbert; Z. F. Feng, *Tetrahedron* **1998**, *54* (49), 14725–14736.
136. A. Gabler; W. Boland; U. Preiss; H. Simon, *Helv. Chim. Acta* **1991**, *74* (8), 1773–1789.
137. H. J. Bouwmeester; F. W. A. Verstappen; M. A. Posthumus; M. Dicke, *Plant Physiol.* **1999**, *121* (1), 173–180.
138. M. Herde; K. Gartner; T. G. Köllner; B. Fode; W. Boland; J. Gershenzon; C. Gatz; D. Tholl, *Plant Cell* **2008**, *20* (4), 1152–1168.
139. D. A. Schooley; K. J. Judy; B. J. Bergot; M. S. Hall; J. B. Siddall, *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70* (10), 2921–2925.
140. B. W. Staddon; A. Abdollahi; J. Parry; M. Rossiter; D. W. Knight, *J. Chem. Ecol.* **1994**, *20* (10), 2721–2731.
141. J. G. C. Hamilton; A. M. Hooper; H. C. Ibbotson; S. Kurosawa; K. Mori; S. E. Muto; J. A. Pickett, *Chem. Commun.* **1999**, (23), 2335–2336.
142. J. G. C. Hamilton; A. M. Hooper; K. Mori; J. A. Pickett; S. Sano, *Chem. Commun.* **1999**, (4), 355–356.
143. V. J. J. Martin; D. J. Pitera; S. T. Withers; J. D. Newman; J. D. Keasling, *Nat. Biotechnol.* **2003**, *21* (7), 796–802.
144. N. Dudareva; S. Andersson; I. Orlova; N. Gatto; M. Reichelt; D. Rhodes; W. Boland; J. Gershenzon, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (3), 933–938.
145. N. Dudareva; D. Martin; C. M. Kish; N. Kolosova; N. Gorenstein; J. Fäldt; B. Miller; J. Bohlmann, *Plant Cell* **2003**, *15* (5), 1227–1241.
146. K. P. Adam; R. Thiel; J. Zapp, *Arch. Biochem. Biophys.* **1999**, *369* (1), 127–132.

Biographical Sketches

Joseph Chappell has been on the faculty at the University of Kentucky since April 1985, where he has developed an internationally recognized research program pioneering the molecular genetics and biochemistry of natural products in plants. Dr. Chappell earned his B.A. degree in biology from the University of California, San Diego in 1977, his Ph.D. in biology in 1981 from the University of California, Santa Cruz, and pursued postdoctoral studies at the University of Freiburg, Germany and the Max Planck Institute, Cologne,

Germany, where he worked on the isolation and characterization of the first genes cloned for plant biosynthetic pathways. At the University of Kentucky, Dr. Chappell's research has focused on the mechanisms plants use to defend themselves against microbial pathogens and especially the biosynthesis of antimicrobial terpene-type compounds.

Robert M. Coates has been emeritus professor of chemistry at the University of Illinois at Urbana-Champaign (UIUC) since his retirement in 2008. He received his BS in chemistry from Yale University in 1960 and his Ph.D. from the University of California, Berkeley in 1964. After his postdoctoral research at Stanford University, he joined the faculty at the UIUC in 1965. His major research interests lie in synthetic and mechanistic organic chemistry, natural products, and biosynthesis, especially the chemistry and biochemistry of isoprenoid compounds. His research publications are primarily concerned with the synthesis and investigation of novel carbocyclic structures, carbocation rearrangements, isotope labeling for elucidation of stereochemistry, and design and synthesis of inhibitors of key enzymes on the isoprenoid biosynthesis pathway. He has been a member of the editorial board for the journal *Steroids* since 2001, and served as an editor for the *Encyclopedia of Reagents for Organic Synthesis* during 1992–98. He has been on the board of directors of Organic Syntheses, Inc. since 1984 and president since 2002.

1.17 Diterpenes

Tomonobu Toyomasu and Takeshi Sassa, Yamagata University, Tsuruoka, Japan

© 2010 Elsevier Ltd. All rights reserved.

1.17.1	**Introduction**	643
1.17.2	**Labdane-Type Diterpenes**	644
1.17.2.1	Formation of Four Copalyl Diphosphate Stereoisomers	644
1.17.2.2	From Copalyl Diphosphate	646
1.17.2.2.1	Diterpene resin acids in conifers	646
1.17.2.2.2	Phyllocladane-related diterpenes in a fungus	647
1.17.2.3	From *ent*-Copalyl Diphosphate	650
1.17.2.3.1	*ent*-Labdane-related diterpenes in rice	650
1.17.2.3.2	*ent*-Kaurane-related diterpenes in moss	652
1.17.2.3.3	*ent*-Pimarane-related diterpenes in eubacteria	652
1.17.2.3.4	Gibberellins	653
1.17.2.3.5	Steviol glycosides	656
1.17.2.4	From *syn*-CPP	656
1.17.2.4.1	*syn*-Labdane-related diterpenes in rice	656
1.17.2.4.2	Aphidicolane-related diterpenes in a fungus	659
1.17.3	**Clerodane- and Halimane-Type Diterpenes**	660
1.17.3.1	Clerodane	660
1.17.3.2	Halimane	661
1.17.4	**Other types of Diterpenes**	662
1.17.4.1	Taxane and Phomactane	662
1.17.4.2	Casbene and Cembratriene	663
1.17.4.3	Fusicoccane	665
1.17.5	**Summary and Future Prospects**	668
References		669

1.17.1 Introduction

A variety of cyclic diterpenes are derived from geranylgeranyl diphosphate (GGPP) through the cyclization of GGPP and the chemical modification of carbon skeletons, such as oxidation, reduction, acetylation, methylation, and glycosylation. Cyclization is an important and interesting process, from the perspective of the formation of a variety of carbon skeletons, and is a major branch of the biosynthesis of cyclic diterpenes. In 1994, the casbene synthase gene from caster bean (*Ricinus communis*) was first reported as a diterpene cyclase (see Section 1.17.4.2),[1] as well as the *ent*-copalyl diphosphate (*ent*-CPP) synthase (see Section 1.17.2.1) from *Arabidopsis thaliana*, which functions in gibberellin biosynthesis.[2] As for labdane-related diterpene cyclases, cDNAs encoding the *ent*-kaurene synthase (see Section 1.17.2.3) in pumpkin (*Cucurbita maxima*) and the abietadiene synthase (see Section 1.17.2.2.1) in grand fir (*Abies grandis*), which function in gibberellin biosynthesis and abietic acid biosynthesis, respectively, were isolated in 1996.[3,4] The taxadiene synthase gene (see Section 1.17.4.1), which functions in taxol biosynthesis, was isolated from Pacific yew (*Taxus brevifolia*) in 1996.[5] MacMillan and Beale,[6] in their introduction to diterpene cyclases, described cyclizations initiated by ionization of the diphosphate of GGPP as type A and cyclizations initiated by protonation at the 14,15-double bond of GGPP as type B, based mainly on information of labdane-related diterpene cyclases. According to their classification, casbene synthase and *ent*-kaurene synthase are of type A, *ent*-CPP synthase is of type B, abietadiene synthase is of type B-A, and taxadiene synthase is of type A-B (**Figure 1**). Abietadiene synthase catalyzes the conversion of GGPP into (+)-CPP (type B) and successively converts (+)-CPP into abietadiene (type A). The cyclization of GGPP by taxadiene synthase is initiated by ionization of the diphosphate of GGDP

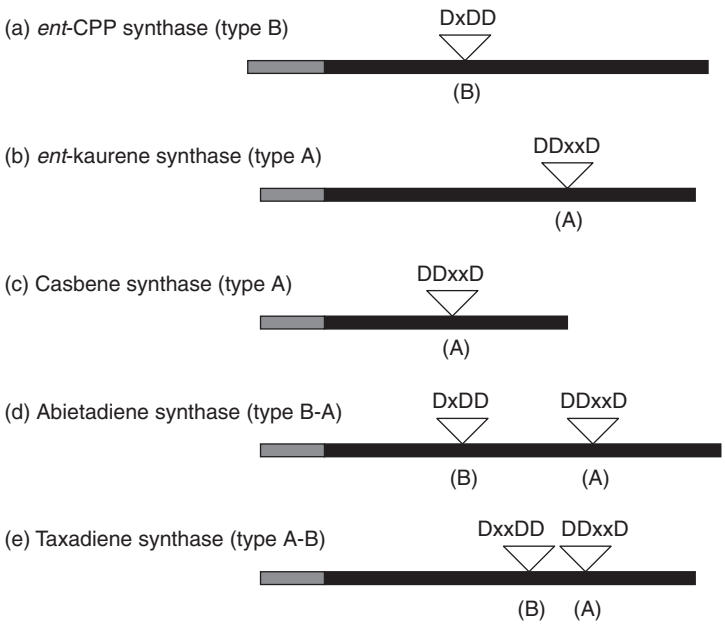


Figure 1 Schematic diagrams of the primary structures of representative diterpene cyclases in plants: (a) *ent*-CPP synthase, (b) *ent*-kaurene synthase, (c) casbene synthase, (d) abietadiene synthase, and (e) taxadiene synthase. Black bars and gray bars indicate putative pseudomature enzymes and transit peptides for plastid targeting, respectively. Type A and type B cyclizations are initiated by ionization of allylic diphosphate group and by protonation of double bond, respectively. Reverse triangles indicate each motif. The type B motif in taxadiene synthase is a modified one (DxxDD).

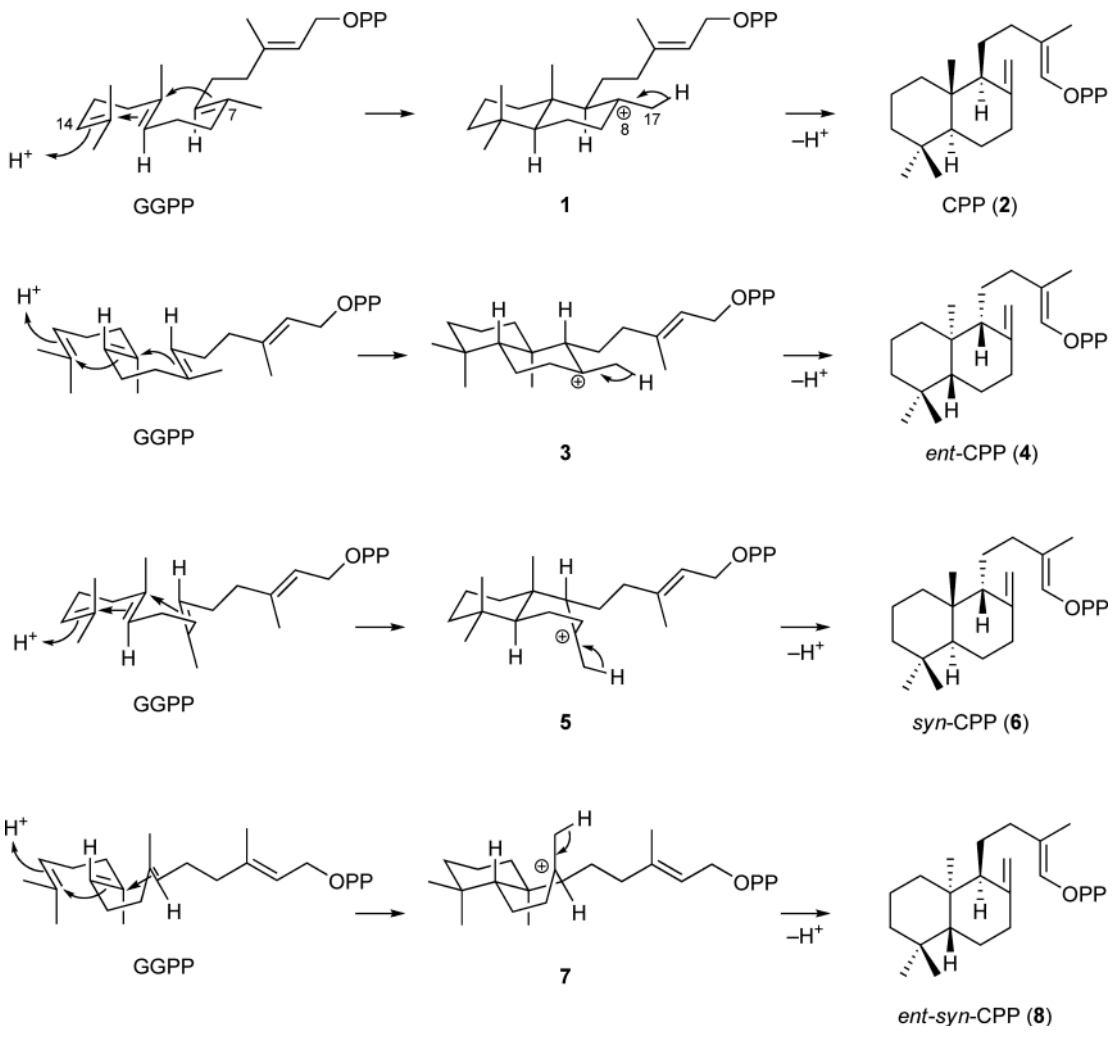
(type A), and further cyclization of the cyclic hydrocarbon intermediate is initiated by protonation at the 7,8-double bond (type B). The details of these reactions are described later. It has been shown that both types of reactions require the D/E-rich motif, as illustrated in **Figure 1**. All diterpene cyclases isolated from higher plants possess transit peptide-like sequences at their N-termini (**Figure 1**), suggesting that the cyclization of GGPP occurs at the plastid in plant cells.

To date, a variety of diterpene cyclase genes have been isolated from higher plants and other taxa, including bryophytes, fungi, and bacteria. An unusual diterpene cyclase gene fused with the GGPP synthase gene was recently isolated from a fungus (see Section 1.17.4.3). This chapter describes the reactions catalyzed by diterpene cyclases that have been isolated so far. In particular, the labdane-related diterpene cyclases are described in more detail. Modification enzymes, which function in the biosynthesis of some diterpenes, such as gibberellins, steviol, and taxol, are also described.

1.17.2 Labdane-Type Diterpenes

1.17.2.1 Formation of Four Copalyl Diphosphate Stereoisomers

The first step in the biosynthesis of labdane-related diterpene hydrocarbons from GGPP is the formation of CPP stereoisomers. The reaction is initiated by protonation of the 14,15-double bond of GGPP, followed by an attack of C-10 on C-15, then C-6 on C-11, affording four stereoisomers of 8-carbonium ions **1**, **3**, **5**, and **7** (**Scheme 1**). The chair–chair conformation of GGPP gives **1** and the chair–boat conformation gives **5**, and their antipodal conformations give **3** and **7**, respectively. The successive deprotonation of 17-H in **1**, **3**, **5**, and **7** gives CPP (**2**), *ent*-CPP (**4**), *syn*-CPP (**6**), and *ent-syn*-CPP (**8**), respectively (**Scheme 1**). The *ent*-CPP synthase gene, which is responsible for gibberellin biosynthesis, was first isolated from *Arabidopsis* (*AtCPS* encoding 802 amino acid residues)[2] and its orthologs have been isolated from several plant species. In higher plants, *ent*-kaurene, a biosynthetic hydrocarbon intermediate of gibberellin, has been indicated to be converted from GGPP by two distinct cyclases, *ent*-CPP synthase and *ent*-kaurene synthase (see Section 1.17.2.3).[7] On the other hand,



Scheme 1

ent-kaurene is converted from GGPP by a single cyclase in moss[8] (see Section 1.17.2.3.2) and in gibberellin-producing fungi (see Section 1.17.2.3.4).[9,10] These cyclases possess type B and type A motifs at their N- and C-termini, respectively, similar to abietadiene synthase (**Figure 1(d)**). Abietadiene synthase converts GGPP into abietadiene via CPP (**2**).[4] Mutational studies have indicated that the type B motifs in these bifunctional cyclases are responsible for the conversion of GGPP into *ent*-CPP (**4**) or CPP (**2**).[8,11,12] Moreover, an *ent*-CPP synthase gene was also identified from the eubacterium *Streptomyces* sp. strain KO-3988 (ORF2).[13,14] The primary structure of the eubacterial *ent*-CPP synthase ORF2 (511 amino acids) is more similar to the primary structure of terpentedienol diphosphate synthase (see Section 1.17.3.1) in *Streptomyces griseolosporeus* MF730-N6 than to the primary structure of eukaryotic *ent*-CPP synthases.

The rice (*Oryza sativa* L. cv. Nipponbare) genome includes two *ent*-CPP synthase genes. *OsCPS1* is responsible for gibberellin biosynthesis, whereas *OsCPS2* functions in the biosynthesis of oryzalexins A–F and phytocassanes A–E, both of which are classes of rice phytoalexins (see Section 1.17.2.3.1).[15] Loss-of-function mutants for *OsCPS1* displayed severe dwarf phenotypes.[16] *OsCPS2* transcription was drastically upregulated upon ultraviolet (UV) treatment in rice leaves.[15] AMO-1618 is a quaternary ammonium compound that not only inhibits the activities of plant *ent*-CPP synthases for gibberellin biosynthesis, but also the type B activity of fungal *ent*-kaurene synthase.[11] Interestingly, OsCPS2 activity was not suppressed in the presence of

Table 1 Labdane-related diterpene cyclase genes in rice cv. Nipponbare

Gene name	*a*	*b*	*Main product*	*References*
OsCPS1	Os02g17780	Os02g0278700	*ent*-CPP	15, 16
OsCPS2	Os02g36210	Os02g0571100	*ent*-CPP	15
OsCPS4	Os04g09900	Os04g0178300	*syn*-CPP	15
OsKS1	Os04g52230	Os04g0611800	*ent*-kaur-16-ene	16, 18
OsKSL4	Os04g10060	Os04g0179700	*syn*-pimara-7,15-diene	19
OsKSL5	Os02g36220	Os02g0571300	*ent*-pimara-8(14),15-diene	20
OsKSL6	Os02g36264	Os02g0571800	*ent*-kaur-15-ene	20
OsKSL7	Os02g36140	Os02g0570400	*ent*-cassa,12,15-diene	21
OsKSL8	Os11g28530	Os11g0474800	stemar-13-ene	18, 22
OsKSL10	Os12g30824	Os12g0491800	*ent*-sandaracopimaradiene	19

Possible pseudogenes are omitted. Loci in Nipponbare by (a) Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/) or (b) Rice Annotation Project Database (http://rapdb.dna.affrc.go.jp/) are shown.

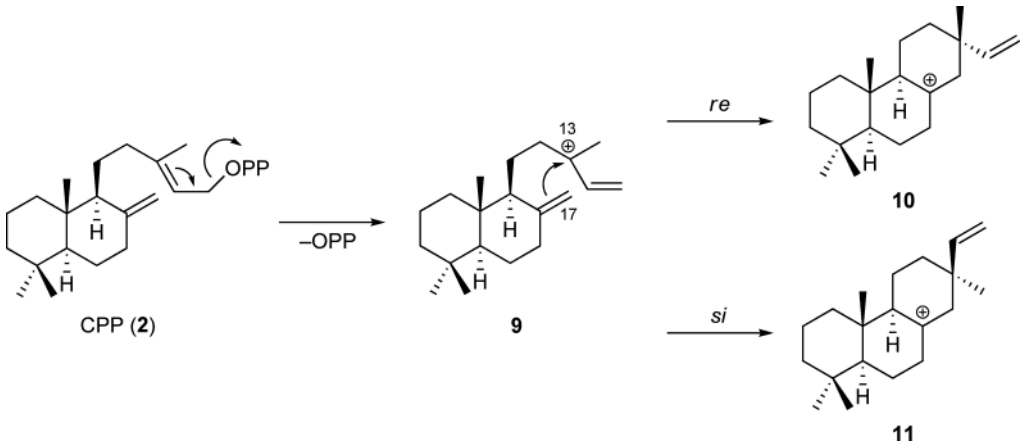
1×10^{-4} mol l^{-1} AMO-1618, under which condition OsCPS1 activity was suppressed.[17] From rice, the *syn*-CPP synthase gene, *OsCPS4*, was also identified as a phytoalexin biosynthetic gene,[15] which functions in the biosynthesis of momilactones A and B and oryzalexin S (see Section 1.17.2.4.1). The primary structure of OsCPS4 (767 amino acids) is similar to those of eukaryotic *ent*-CPP synthases, all of which possess the type B motif, as shown in **Figure 1(a)**. The type A and type B labdane-related cyclase genes in rice are summarized in **Table 1**.[23,24] The *OsCPS1*, *OsCPS2*, and *OsCPS4* genes were also identified in the *indica* rice cultivar IR24.[25,26] Type A rice cyclases are discussed in detail in Sections 1.17.2.3.1 and 1.17.2.4.1.

The cDNAs encoding CPP synthase and *ent-syn*-CPP synthase have not been identified in any organism. The recombinant protein encoded by *PaDC2* (984 amino acids) from the fungus *Phomopsis amygdali* converted GGPP into CPP (**2**) as an end product.[27] However, this reaction may have been due to a naturally occurring mutation at the C-terminal region of the fungal bifunctional diterpene cyclase (see Section 1.17.2.2.2), which may eliminate the activity for conversion of CPP (**2**) into cyclic hydrocarbons. Artificial mutants for the bifunctional abietadiene synthase, in which ^{621}D in the type A motif was substituted with alanine, exhibited a loss of type A activity and also converted GGPP into CPP (**2**).[28]

1.17.2.2 From Copalyl Diphosphate

1.17.2.2.1 Diterpene resin acids in conifers

Abietadiene synthase (868 amino acids) from grand fir (AgAS) produces abietadiene (**16**), levopimaradiene (**17**), and neoabietadiene (**18**) as major products, as well as plaustradiene (**19**), sandaracopimaradiene (**12**), and pimara-8(14),15-diene (**22**) as minor products; these compounds are produced from GGPP via CPP (**2**; see **Schemes 2–4**).[4,29] Abietadiene (**16**) is a biosynthetic intermediate of abietic acid (**20**), a compound that confers resistance against herbivory and pathogen attacks. The P-450 monooxygenase responsible for abietic acid (**20**) biosynthesis (PtAO, CYP720B1) was identified.[30] Cyclization mechanisms of these products are as follows: after ionization of the diphosphate of CPP (**2**), carbocation **9** is cyclized by the attack of C-17 at the *re*- or *si*-face of C-13, giving carbocation **10** or **11** (**Scheme 2**). The deprotonation of 14-H in carbocation **10** gives **12**, successive protonation of the 15,16-double bond of **12** gives carbocation **14**, and a 1,2-shift of the methyl from C-13 to C-15 gives carbocation **15** (**Scheme 3**). The deprotonation of 7-H, 12-H, 15-H, and 9-H in **15** gives compounds **16**, **17**, **18**, and **19**, respectively (**Scheme 3**), whereas the deprotonation of 14-H in carbocation **11** gives **22** (**Scheme 4**). As described in Section 1.17.1, AgAS possesses the type A and type B motifs (**Figure 1(d)**), and the type A motif DDxxD is required for the above reaction.[28] Other required amino acid residues were identified via mutational studies.[31] Notably, the two cyclization sites of AgAS cannot be dissected into catalytically distinct domains,[32] unlike the fungal *ent*-kaurene synthase (see Section 1.17.2.3.4). From Norway spruce (*Picea abies*) and loblolly pine (*Pinus taeda*), the cDNAs *PaLAS* and *PtLAS* for levopimaradiene/abietadiene synthases were cloned, respectively,[33,34] and a cDNA encoding levopimaradiene synthase



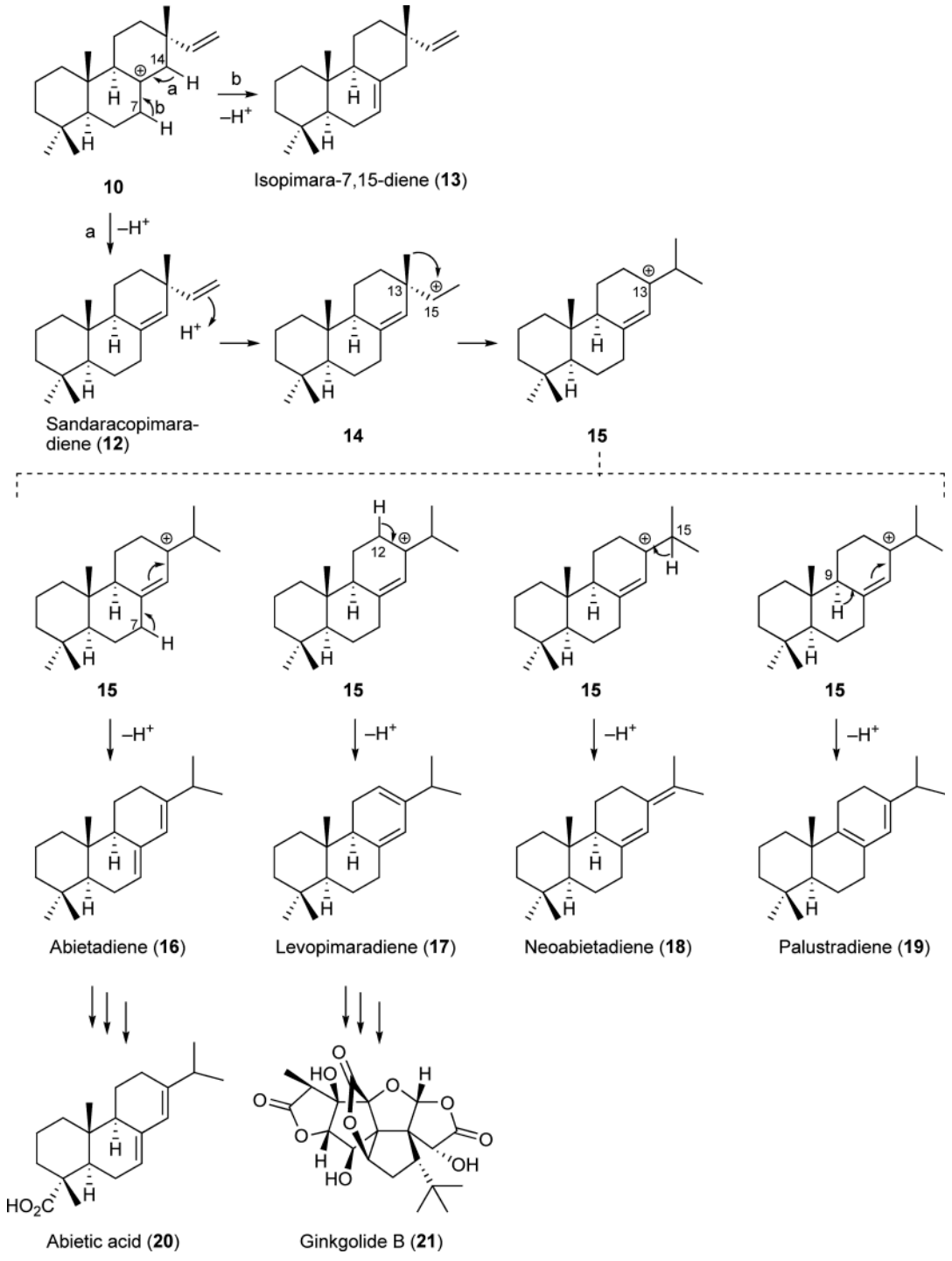
Scheme 2

(GbLS, 873 amino acids) was isolated from *Ginkgo biloba*.[35] Levopimaradiene (**17**) is a possible biosynthetic precursor of ginkgolides (e.g., **21**), which are potent platelet-activating factor (PAF) antagonists isolated from *G. biloba*. Another cyclase gene was isolated from spruce.[33] Isopimaradiene synthase in spruce (PaIso) produced one product, isopimara-7,15-diene (**13**), via the deprotonation of 7-H in carbocation **10**. The details of the biosynthesis of diterpene resin acids in conifers were previously reviewed by Keeling and Bohlmann.[36]

Conifer diterpene cyclases are highly homologous. Domain swapping and site-directed mutagenesis were carried out to identify amino acid substitutions that lead to alternative products. Mutant AgAS, in which ^{723}A was substituted with serine, produced mainly isopimara-7,15-diene (**13**) and sandaracopimaradiene (**12**).[37] The substitution of ^{713}A of PaLAS with serine (PaLAS A713S) similarly changed the products.[38] On the other hand, the PaIso S721A mutant produced mainly **13** and a small amount of **12**; the quadruple mutant, PaIso L687W/H694Y/S721A/L725V, produced abietadiene (**16**)/levopimaradiene (**17**).[38] The catalytic activities of these mutants did not differ significantly compared to the wild type. The results of substituting the alanine with serine suggest that the introduced hydroxyl group stabilizes carbocation **10** long enough for deprotonation to occur. Similar speculation on the rice type A cyclases is presented in Section 1.17.2.3.1.

1.17.2.2.2 Phyllocladane-related diterpenes in a fungus

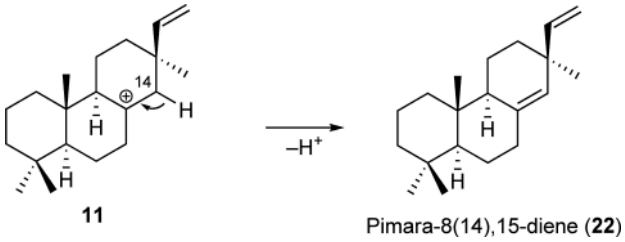
Phomopsis amygdali is a plant-pathogenic fungus that produces fusicoccins (see Section 1.17.4.3). In addition to fusicoccane-related hydrocarbons, a series of phyllocladane-related hydrocarbons were identified from the mycelia of *P. amygdali* F6 strain.[39] Two labdane-related cyclase genes, *PaDC1* and *PaDC2*, were isolated from the F6 strain. The recombinant PaDC1 converted GGPP to mainly phyllocladan-16α-ol (**26**; **Scheme 5**) via CPP.[27] Minor products included sandaracopimaradiene (**12**), isopimara-8,15-diene (**24**), phyllocladene (**27**), pimara-8(14),15-diene (**22**), pimara-8,15-diene (**30**), and kaurene (**32**). During the rearrangement of **10**, carbocation **25** is formed via **23**, and the deprotonation of 17-H in **25** gives phyllocladene (**27**), while stereoselective attack by a water molecule generates phyllocladan-16α-ol (**26**; **Scheme 5**). Compounds **12** and **24** are produced by the deprotonation of 14-H and 9-H in **10**, respectively (**Scheme 5**). Sequential rearrangement of **11** via **29** gives carbocation **31**, and the deprotonation of 17-H in **31** yields **32** (**Scheme 6**). The deprotonation of 14-H and 9-H in compound **11** yields compounds **22** and **30**, respectively (**Scheme 6**). PaDC1 (1017 amino acids) possesses the type B motif DxDD (positions 321–324) and the type A motif DExxE (positions 667–671), similar to AgAS (**Figure 1(d)**). The primary structure of PaDC1 is homologous to those of other fungal labdane-related cyclases (see Sections 1.17.2.3.4 and 1.17.2.4.2). As described in Section 1.17.2.1, recombinant PaDC2 converted GGPP into CPP, but did not catalyze the further cyclization of CPP, which may have been due to a naturally occurring mutation in the type A domain. With the presence of both a type B



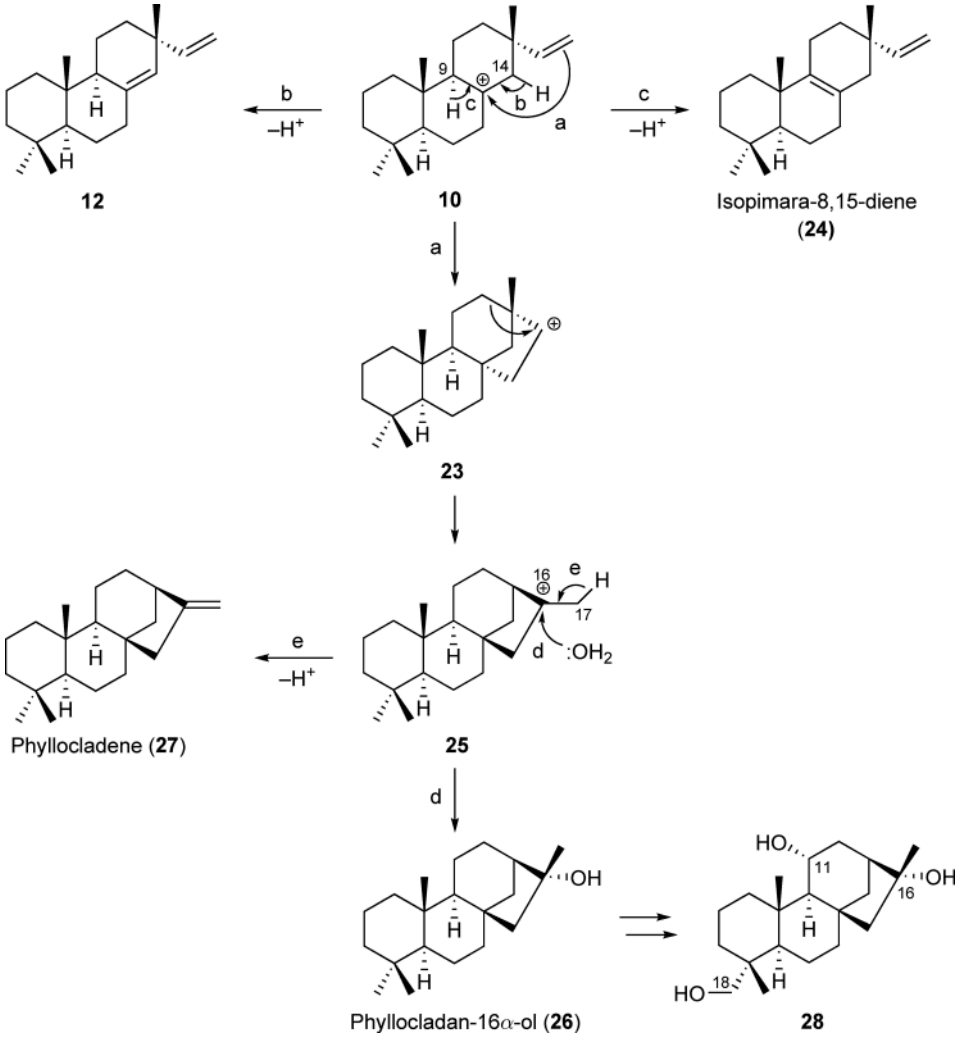
Scheme 3

motif (positions 318–321) and a type A motif (positions 656–660) in PaDC2, a C-terminal region other than that of the type A motif may be mutated, causing a loss of the type A function.[27]

Since the discovery of the gibberellin biosynthetic gene cluster in *Gibberella fujikuroi* (see Section 1.17.2.3.4), it has been shown that diterpene biosynthetic genes are clustered within the fungal genome. Five P-450

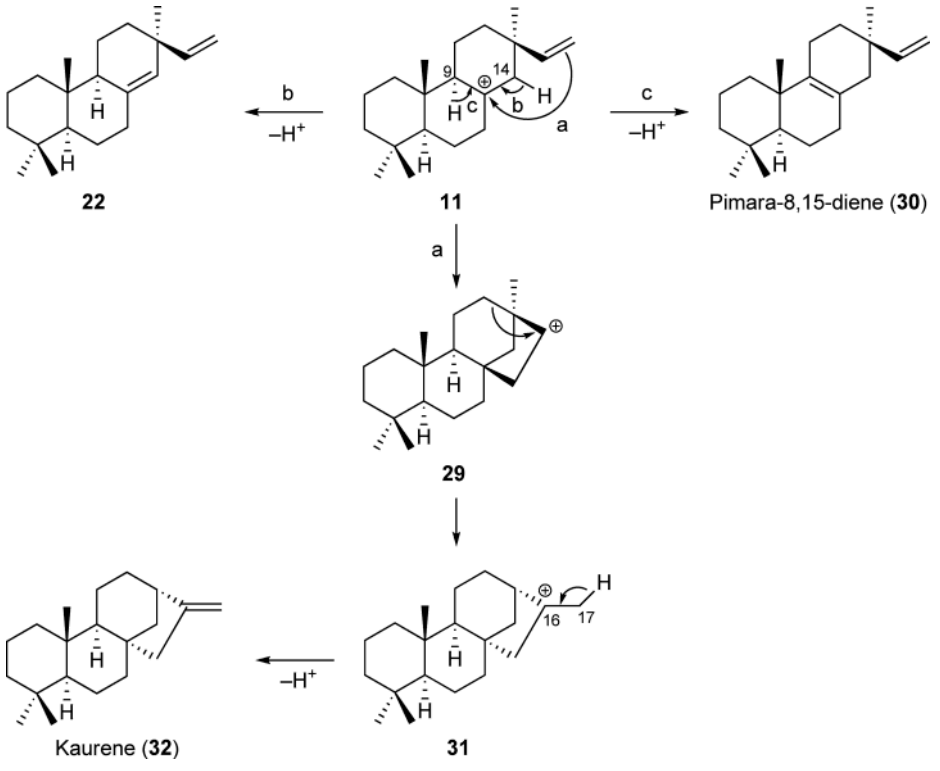


Scheme 4



Scheme 5

monooxygenase-like genes, one dehydrogenase-like gene, and one dioxygenase-like gene are located around *PaDC1* in the *P. amygdali* genome.[27] From the F6 mycelia, phyllocladan-11α,16α,18-triol (**28**) was identified as a possible hydroxylated metabolite of **26**.[27] Two P-450-like genes, among the five genes around *PaDC1*, may be responsible for the hydroxylation of C-11α and C-18 in **26**. Highly oxygenated phyllocladanes, which have not been identified from the *P. amygdali* mycelia, may be generated by other oxygenases.

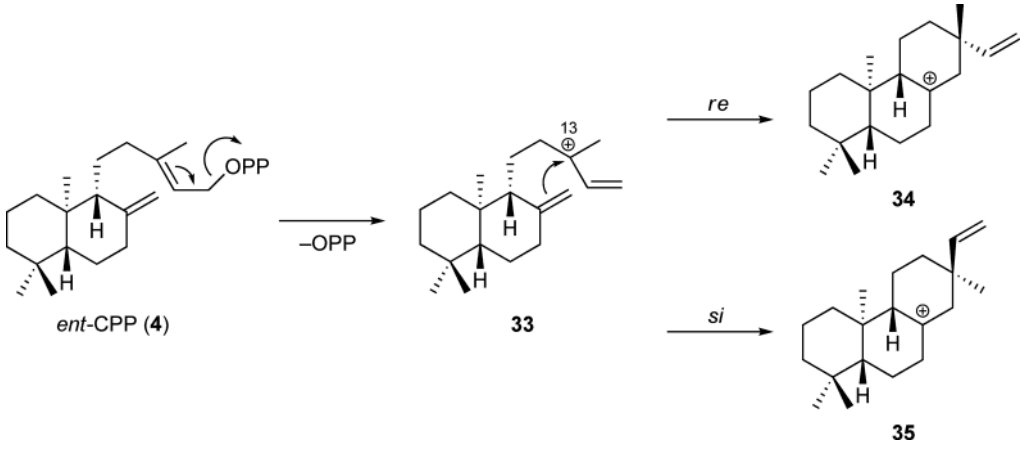


Scheme 6

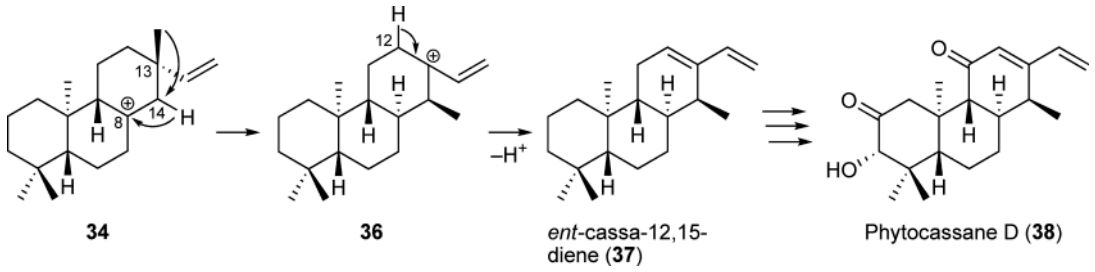
1.17.2.3 From *ent*-Copalyl Diphosphate

1.17.2.3.1 ent-*Labdane-related diterpenes in rice*

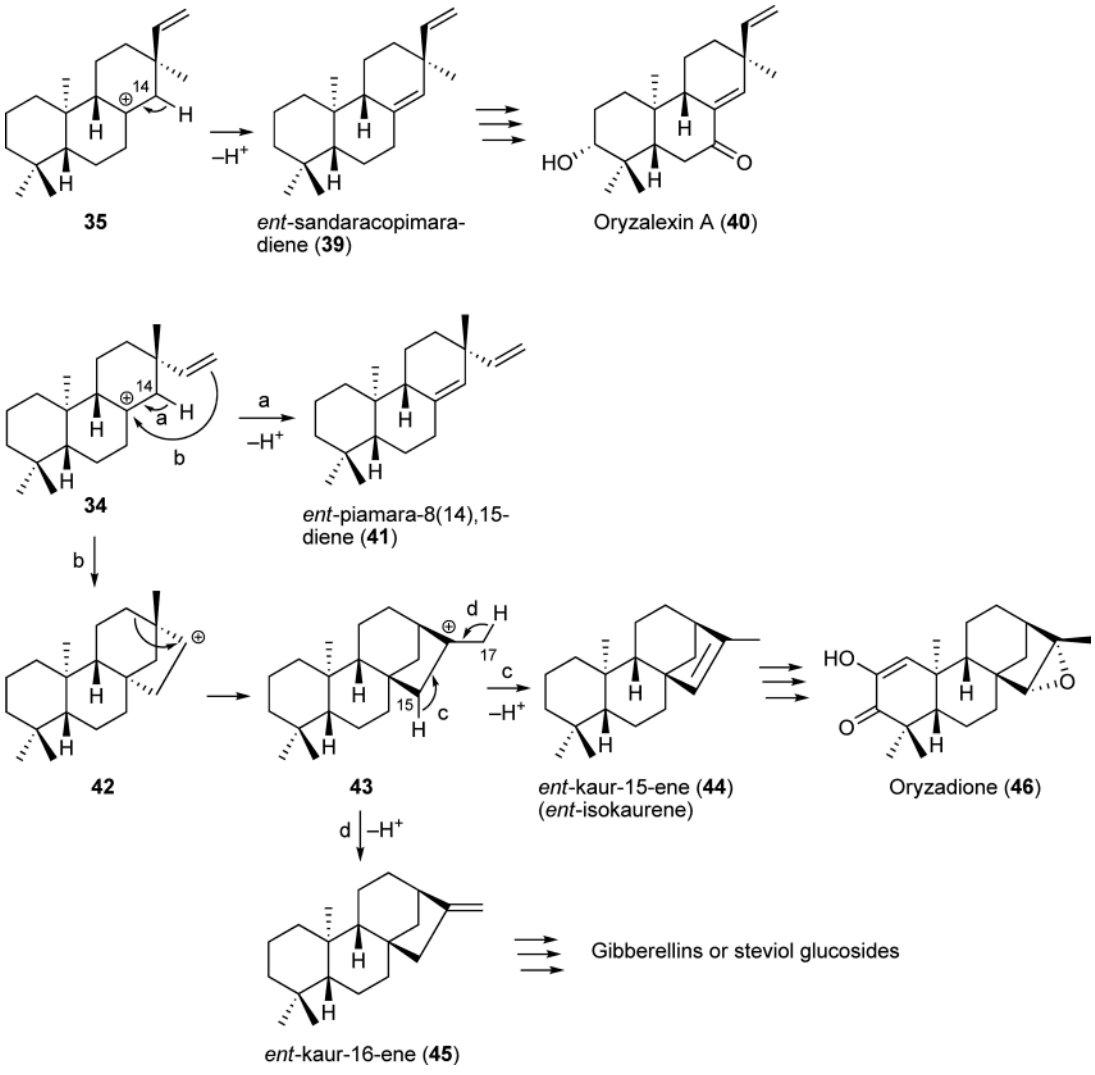
As type A *ent*-labdane-related diterpene cyclase genes, *OsKS1*, *OsKSL5*, *OsKSL6*, *OsKSL7*, and *OsKSL10* were identified in both *japonica* (Nipponbare) and *indica* (IR24) rice cultivars. These genes are homologs of the *ent*-kaurene synthase possessing a type A motif (**Figure 1(b)**). *ent*-Cassa-12,15-diene (**37**) and *ent*-sandaracopimaradiene (**39**), possible biosynthetic intermediate cyclic hydrocarbons of phytocassanes A–E (e.g., **38**) and oryzalexins A–F (e.g., **40**), are converted from *ent*-CPP (**4**) by OsKSL7 and OsKSL10, respectively (**Schemes 7–9** and **Table 1**).[18,19,21] As in the case of CPP (**2**; **Scheme 2**), the C-13 carbocation **33**, derived from *ent*-CPP



Scheme 7



Scheme 8



Scheme 9

(**4**) after the ionization of diphosphate, gives carbocation **34** or **35** via the attack of C-17 at the *re*- or *si*-face of C-13, respectively (**Scheme 7**). The 1,2-hydride shift from C-14 to C-8 and successive 1,2-shift of a methyl group from C-13 to C-14 in carbocation **34** gives carbocation **36**. The deprotonation of 12-H in **36** gives *ent*-cassa-12,15-diene (**37**), whereas the deprotonation of 14-H in **35** gives *ent*-sandaracopimaradiene (**39**);

ent-pimara-8(14),15-diene (**41**) was also identified as a minor product of *ent*-CPP (**4**), catalyzed via OsKSL10.[19] Compound **41** is generated by the deprotonation of 14-H in **34**. The transcription of *OsKSL7* and *OsKSL10*, as well as *OsCPS2* and *OsCPS4*, was induced by UV irradiation in rice leaves or by chitin-elicitor treatment in cell suspensions.[18,19,21]

OsKSL5 and OsKSL6 in Nipponbare converted *ent*-CPP (**4**) into *ent*-pimara-8(14),15-diene (**41**) and *ent*-kaura-15-ene (*ent*-isokaurene, **44**), respectively (**Schemes 7 and 9** and **Table 1**).[20] Compound **44** is a possible precursor of oryzalide-related compounds (e.g., oryzadione **46**), which serve as antifungal agents.[40] Sequential rearrangements of **34** via **42** yields carbocation **43**, and the deprotonation of 15-H in **43** gives **44** (**Scheme 9**). Interestingly, both OsKSL5 and OsKSL6 in IR24 converted **4** into **44** (**Scheme 9**).[18] The comparison of the OsKSL5 amino acid sequence in IR24 with that in Nipponbare led to the identification of specific residues responsible for this difference in compounds produced.[41] The product of catalysis by the I664T mutant of OsKSL5 in IR24 was **41**, and conversely, the T661I mutant of OsKSL5 in Nipponbare produced mainly **44**. These results suggest that the introduced hydroxyl group stabilized carbocation **34**, similar to conifer diterpene cyclases (see Section 1.17.2.2.1). The hypothesis is consistent with the case of OsKSL6, OsKS1, and AtKS.[41]

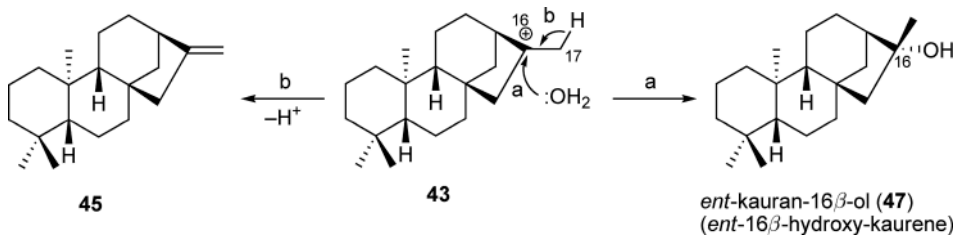
Loss-of-function mutants for *OsKS1* exhibited a severe dwarf phenotype, similar to *OsCPS1* mutants.[16] The recombinant OsKS1 protein converted *ent*-CPP (**4**) into *ent*-kaur-16-ene (**45**; **Table 1**).[18] Compound **45** is produced via the deprotonation of 17-H in carbocation **43** (**Scheme 9**). The orthologs of *ent*-kaurene synthase that function in gibberellin biosynthesis have been isolated from several higher plants, including pumpkin and *Arabidopsis*.[3,7] Compound **45** is metabolized into steviol and its glycosides in *Stevia rebaudiana* (see Section 1.17.2.3.5).[42]

1.17.2.3.2 ent-*Kaurane-related diterpenes in moss*

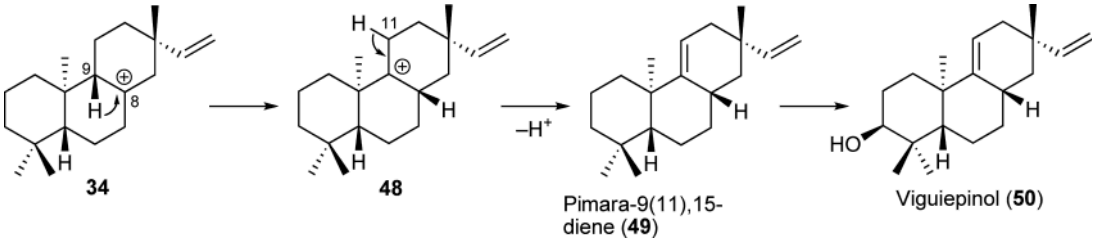
A cDNA encoding a bifunctional type B-A diterpene cyclase (PpCPS/KS, 881 amino acids) has been isolated from moss (*Physcomitrella patens*). PpCPS/KS converted GGPP into mainly *ent*-kauran-16β-ol (*ent*-16β-hydroxy-kaurene, **47**) via *ent*-CPP (**4**).[8] Stereoselective attack of a water molecule on the C-16 cation in **43** gives **47** (**Scheme 10**). As a by-product, *ent*-kaur-16-ene (**45**) was produced by the deprotonation of 17-H in **43**. The type B and type A motifs in PpCPS/KS are required to convert GGPP into *ent*-CPP (**4**), and **4** into **47** and **45**,[8] respectively, similar to AgAS (see Section 1.17.2.2.1) and fungal *ent*-kaurene synthase (see Section 1.17.2.3.4). The metabolism of gibberellins from **45** in bryophytes remains unexplained.

1.17.2.3.3 ent-*Pimarane-related diterpenes in eubacteria*

ORF1, ORF3, and ORF4 were identified near ORF2 of the *ent*-CPP synthase gene (see Section 1.17.2.1) in the genome of eubacteria (*Streptomyces* sp. strain KO-3988).[13] ORF1 encodes a P-450-like protein, whereas ORF4 encodes a GGPP synthase. The recombinant protein from ORF3 (295 amino acids) converted *ent*-CPP (**4**) into pimara-9(11),15-diene (**49**, **Scheme 11**).[43] The 1,2-hydride shift from C-9 to C-8 in carbocation **34** and successive deprotonation of 11-H in **48** gives **49** (**Scheme 11**). The primary structure of ORF3 shows 37% similarity to Cyc2, a terpentetriene synthase (see Section 1.17.3.1). Both enzymes possess a type A motif. The *Streptomyces lividans* TK23 strain, in which the genomic DNA including ORFs 1–4 was introduced, produced viguiepinol (**50**).[14] This result indicates that ORFs 1–4 are clustered for the biosynthesis of **50**, and that ORF1 may be responsible for the 3β-hydroxylation of **49**.



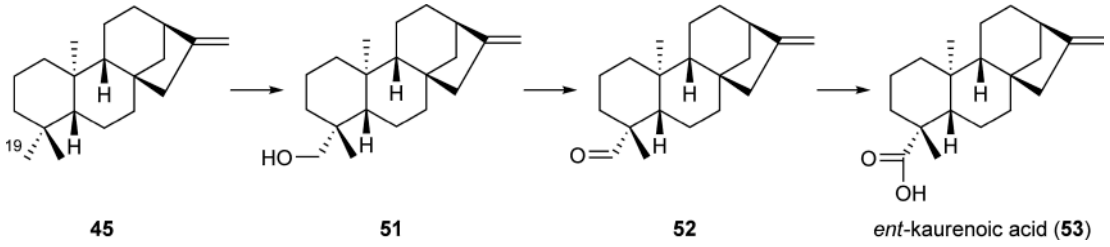
Scheme 10



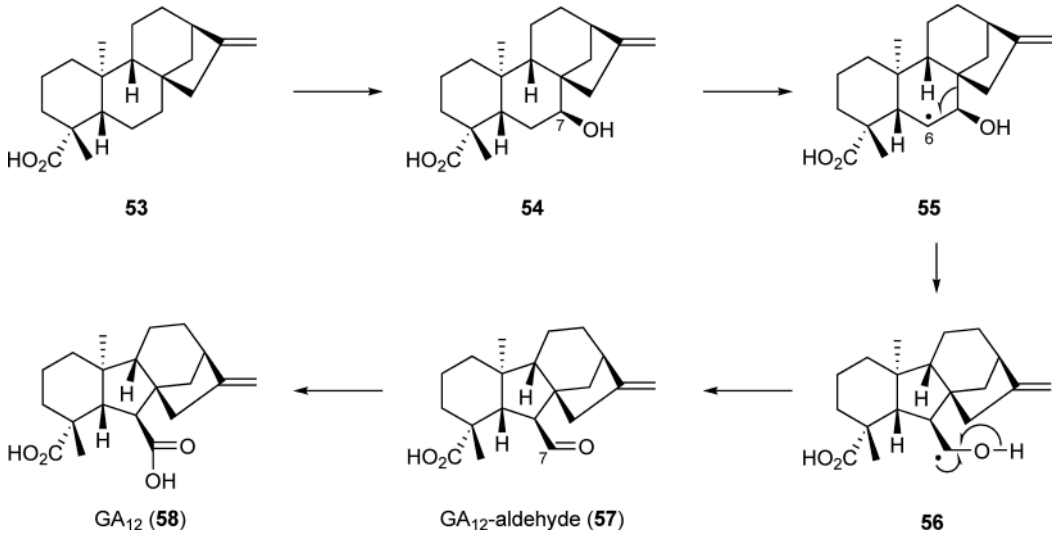
Scheme 11

1.17.2.3.4 Gibberellins

Gibberellins are phytohormones that regulate various aspects of plant growth.[7] Gibberellin-deficient mutants have demonstrated the importance of these compounds in plant growth regulation. Gibberellin metabolism genes have been identified in various higher plants and fungi.[7] As described above, *ent*-kaur-16-ene (**45**) is an intermediate of gibberellins and is produced by two distinct plastid-localized cyclases, *ent*-CPP synthase (see Section 1.17.2.1) and *ent*-kaurene synthase (see Section 1.17.2.3.1) in higher plants. A loss-of-function mutant of the *ent*-kaurene synthase gene in *Arabidopsis* resulted in severe dwarf phenotypes,[44] similar to rice (see Section 1.17.2.3.1). C-19 in **45** is successively oxidized to produce *ent*-kaurenoic acid (**53**) via *ent*-kaurenol (**51**) and *ent*-kaurenal (**52**; **Scheme 12**) by *ent*-kaurene oxidase (KO), which belongs to the P-450 monooxygenase family. The cDNA encoding KO was first cloned from *Arabidopsis* and characterized (CYP701A5).[45,46] Furthermore, **53** is successively oxidized to GA_{12} (**Scheme 13**) via *ent*-7α-hydroxy-kaurenoic acid (**54**) and GA_{12}-aldehyde (**57**)



Scheme 12



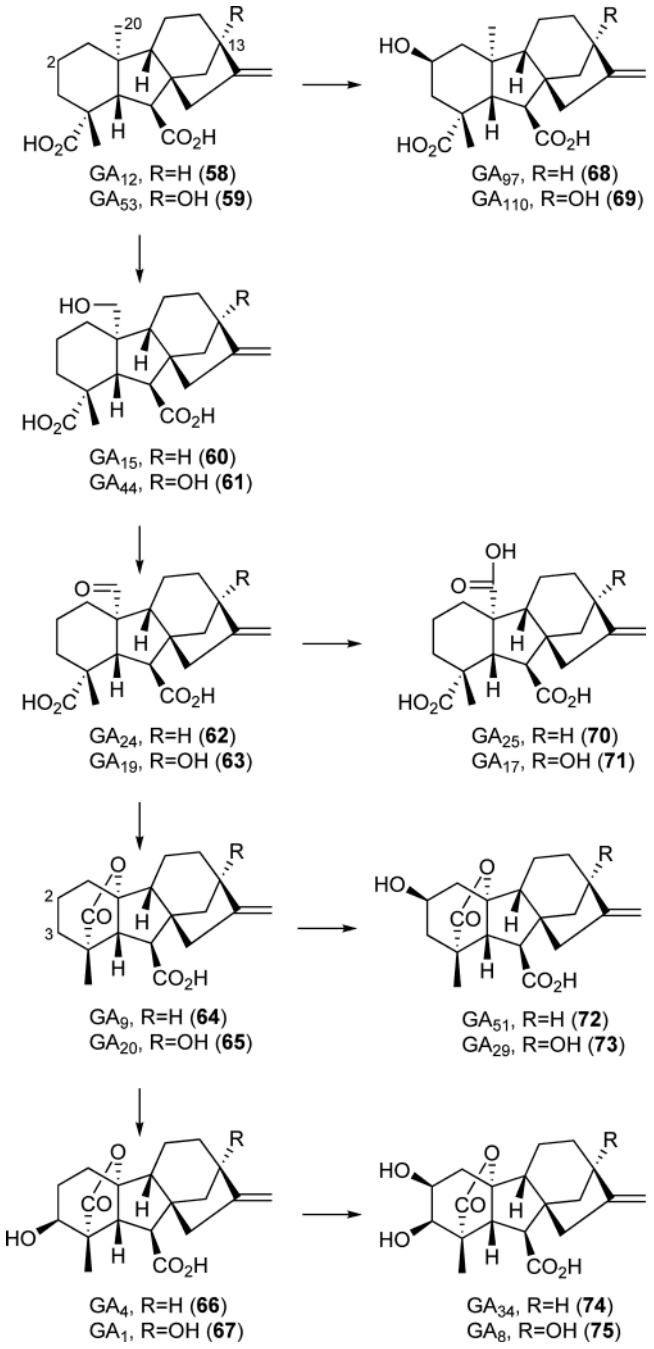
Scheme 13

by *ent*-kaurenoic acid oxidase (KAO), which is also a member of the P-450 monooxygenase family. KAO was first identified in barley (*Hordeum vulgare*) and *Arabidopsis* (CYP88A3 and CYP88A4).[47] The radical (**55**) produced by the extraction of 6β-H undergoes rearrangement, whereby the 7–8 bond migrates to C-6 to form GA_{12}-aldehyde (**57**) via **56**. In pumpkin, gibberellin 7-oxidase (GA7ox), which is a member of the 2-oxoglutarate-dependent soluble dioxygenase family, was identified.[48] GA7ox converts GA_{12}-aldehyde (**57**) to GA_{12} (**58**). Orthologs of the *GA7ox* gene have not been identified in *Arabidopsis* or rice.

Physiologically active gibberellins are biosynthesized from GA_{12} (**58**) through two primary parallel pathways in higher plants, the early-nonhydroxylation pathway and early-13-hydroxylation pathway, which lead to GA_4 (**66**) and GA_1 (**67**), respectively (**Scheme 14**). These steps are catalyzed by 2-oxoglutarate-dependent dioxygenases, although gibberellin 13-hydroxylase has not been identified so far in higher plants. Gibberellin 20-oxidase (GA20ox) converts GA_{12} (**58**) and GA_{53} (**59**) into GA_9 (**64**) and GA_{20} (**65**), respectively, by the oxidation of C-20. The γ-lactone formation type GA20ox, which leads to bioactive gibberellins, was first identified in *Arabidopsis*.[49] Pumpkin GA20ox converted GA_{12} (**58**) and GA_{53} (**59**) into mainly GA_{25} (**70**) and GA_{17} (**71**), tricarboxyl-type gibberellins, respectively.[50] Finally, 3β-hydroxylation of GA_9 (**64**) and GA_{20} (**65**) produces physiologically active gibberellins GA_4 (**66**) and GA_1 (**67**), respectively. Bioactive GA_4 (**66**) and GA_1 (**67**) are deactivated through 2β-hydroxylation, which produces GA_{34} (**74**) and GA_8 (**75**), respectively. These steps are catalyzed by gibberellin 3-oxidase (GA3ox)[51,52] and gibberellin 2-oxidase (GA2ox),[53] respectively. Loss-of-function mutants for *GA3ox* exhibited a dwarf phenotype, and those of *GA2ox* displayed an elongated phenotype.[7] These results indicate that GA_4 (**66**) and GA_1 (**67**) are physiologically active gibberellins in higher plants. GA2ox enzymes also convert GA_9 (**64**) and GA_{20} (**65**), direct precursors of GA_4 (**66**) and GA_1 (**67**), into GA_{51} (**72**) and GA_{29} (**73**). A novel class of GA2ox, which prefers C_{20}-gibberellins to γ-lactone-type C_{19}-gibberellins as substrates, was identified in *Arabidopsis*, and converts GA_{12} (**58**) and GA_{53} (**59**) into GA_{97} (**68**) and GA_{110} (**69**).[54]

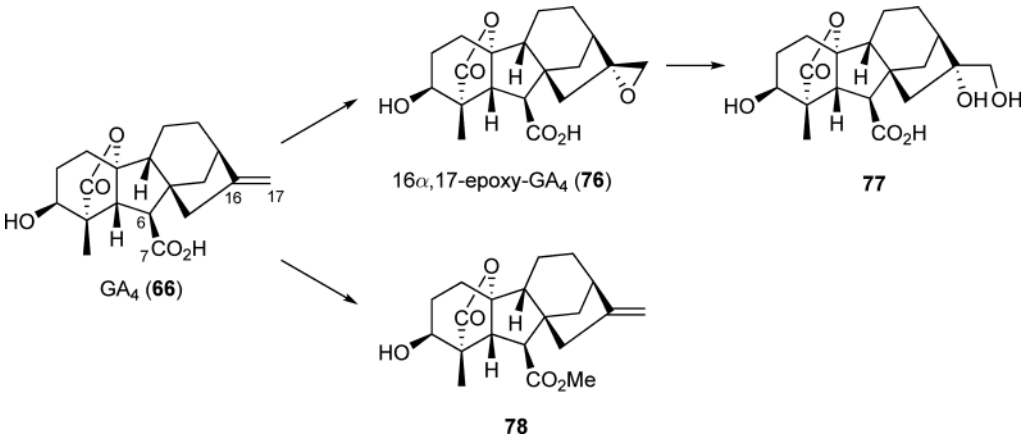
Recently, other types of gibberellin metabolic enzymes were identified in rice and *Arabidopsis*. Rice *eui* mutants possessing an impaired P-450 monooxygenase gene exhibited an elongated phenotype. The P-450 enzyme (CYP714D1) converts non13-hydroxylated C_{20}- and C_{19}-gibberellins (e.g., **66**) into their 16,17-epoxide types (e.g., **76**; **Scheme 15**).[55] The 16α,17-dihydrodiols (e.g., **77**) may be nonenzymatically produced by hydration of the epoxides (e.g., **76**). Compounds **76** and **77** are not physiologically active. Taken together, these results suggest that CYP714D1 serves in gibberellin deactivation. Methyltransferases, which prefer γ-lactone-type gibberellins (e.g., **66**) as substrates to produce methylated gibberellins (e.g., **78**; **Scheme 15**), were identified in *Arabidopsis* (GAMT1 and GAMT2).[56] Transgenic *Arabidopsis*, in which GAMT1 or GAMT2 was overexpressed, exhibited severe dwarf phenotypes, suggesting that methylation is a deactivation step for gibberellins, although methylated gibberellins have not been identified in higher plants.

Gibberellins were originally isolated as secondary metabolites of a fungus, *G. fujikuroi*. The biosynthetic pathway of gibberellins from GA_{12} (**58**) to bioactive GA_3 (**81**) in *G. fujikuroi* is shown in **Scheme 16** by bold arrows. On the other hand, *Phaeosphaeria* sp. L487 produces GA_1 (**67**), as indicated by normal arrows in **Scheme 16**. Several steps from GGPP to GA_{12} (**58**) via *ent*-kaurene (**45**) are common to both fungi. As described, *ent*-kaurene (**45**) is converted from GGPP via *ent*-CPP (**4**) by a single enzyme (PhCPS/KS or GfCPS/KS) in both fungi.[9,10] The fungal bifunctional cyclases possess both type A and type B motifs, similar to AgAS (**Figure 1(d)**). Site-directed mutagenesis of PhCPS/KS demonstrated that the type B motif is required for the conversion of GGPP into *ent*-CPP (**4**), and that the type A motif is required to convert **4** into *ent*-kaurene (**45**).[11] Furthermore, experiments using truncated mutants for PhCPS/KS revealed that the N- and C-terminal domains, respectively, possess *ent*-CPP synthase and *ent*-kaurene synthase activity.[11] The genes for the enzymes responsible for the steps from *ent*-kaurene (**45**) to GA_3 (**81**) were identified in *G. fujikuroi* by gene walking from the GfCPS/KS gene. One GGPP synthase-like gene and four P-450-like genes (P-450-1, P-450-2, P-450-3, and P-450-4) were found in the flanking region of genomic DNA near the *GfCPS/KS* gene.[57] P-450-4 (CYP503)[58] and P-450-1 (CYP68A)[59] are responsible for the conversion of *ent*-kaurene (**45**) to *ent*-kaurenoic acid (**53**) and **53** to GA_{14} (3β-hydroxylated GA_{12}, **79**), respectively. The primary structures of these two fungal P-450 enzymes are not highly homologous to the corresponding P-450 enzymes in higher plants. Interestingly, fungal KAO not only converts *ent*-kaurenoic acid (**53**) into GA_{12} (**58**), but also catalyzes the 3β-hydroxylation of GA_{12}. This is because the early-3-hydroxylation



Scheme 14

gibberellin biosynthetic pathway operates in *G. fujikuroi*. Unlike higher plants, fungal GA20ox (P-450-2) in *G. fujikuroi* is a member of the P-450 family (CYP68B).[60] Finally, GA_4 desaturase, which was found in subsequent gene walking studies, converts GA_4 (**66**) to GA_7 (**80**), and P-450-3 (CYP69A1) hydroxylates C-13 in GA_7 (**80**) to produce GA_3 (**81**).[61] This information indicates that gibberellin biosynthetic genes are clustered within the *G. fujikuroi* genome (**Figure 2**). A possible gibberellin biosynthetic gene cluster was also suggested in *Phaeosphaeria*.[62]



Scheme 15

1.17.2.3.5 Steviol glycosides

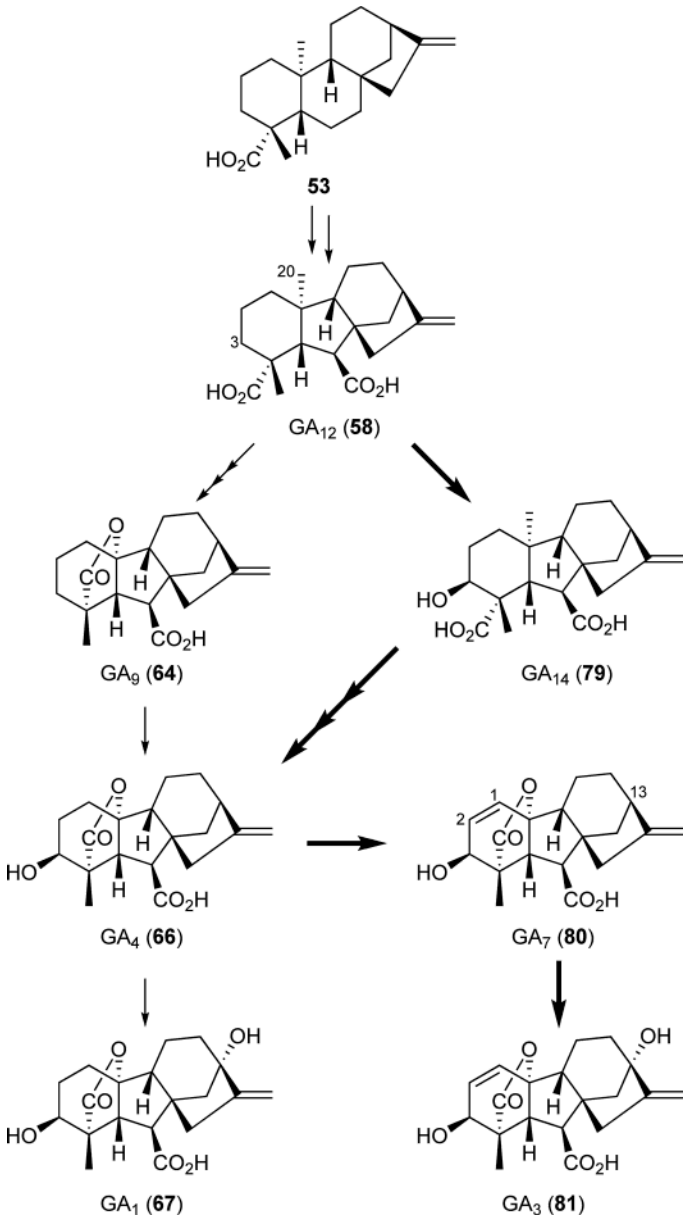
Stevia rebaudiana produces a series of steviol glycosides in its leaves (e.g., **85** in **Scheme 17**). These compounds have an intensely sweet flavor, and their biosynthesis was reviewed by Brandle and Telmer.[63] The aglycon of the steviol glycosides is steviol (**82**), which is derived from GGPP via *ent*-CPP (**4**), *ent*-kaurene (**45**), and *ent*-kaurenoic acid (**53**; **Schemes 1**, **7**, **9**, and **12**). From *S. rebaudiana*, one *ent*-CPP synthase gene and two *ent*-kaurene synthase genes were cloned and characterized.[64] KO in *S. rebaudiana* catalyzed the conversion of **45** into **53**,[65] whereas 13-hydroxylase, which converts **53** into **82**, has not been identified. The UDP-glucosyltransferases (UGTs) in *S. rebaudiana* have been characterized.[66] UGT85C2 transfers glucose to C-13 in steviol (**82**) to produce 13-*O*-β-glucosylated steviol (steviolmonoside, **83**). UGT74G1 and UGT76G1 are responsible for the addition of glucose to the hydroxyl of the C-4 carboxyl group in steviolbioside (**84**) to produce stevioside (**85**) and for the transfer of glucose to C-3′ of the 13-*O*-glucosyl moiety in **85** to produce rebaudioside (**86**), respectively. The UGT, which transfers glucose to C-2′ of the 13-*O*-glucosyl moiety in **83** to produce **84**, has not been identified.

1.17.2.4 From *syn*-CPP

1.17.2.4.1 syn-*Labdane-related diterpenes in rice*

In rice, *syn*-pimara-7,15-diene (**90**) and stemar-13-ene (**98**) are the possible biosynthetic intermediate hydrocarbons of momilactones A (**92**) and B (**93**), and oryzalexin S (**99**), respectively, and are derived from *syn*-CPP (**6**), as shown in **Schemes 18–20**. Compounds **92**, **93**, and **99** serve as phytoalexins in rice. OsKSL4 and OsKSL8, type A cyclases in *japonica* Nipponbare rice, catalyze the conversion of **6** into mainly **90** and **98** (**Table 1**).[18,19,22] The *re*- and *si*-face attacks of C-17 on C-13 in **87** yields the C-8 carbocations **88** and **89**, respectively (**Scheme 18**), whereas the deprotonation of 7-H in **89** yields **90** (**Scheme 19**). The OsKSL4 in *indica* IR24 rice also catalyzes the same reaction.[67] The 1,2-hydride shift from C-9 to C-8 in **88** gives the C-9 carbocation **94**, and the successive attack of C-16 on C-9 gives the C-15 carbocation **95** (**Scheme 20**). The rearrangement (pathway a in **Scheme 20**) and successive deprotonation of 14-H in **95** gives **98** via **96**. An alternative rearrangement (pathway b in **Scheme 20**) of **95** gives the C-13 carbocation **97**. The deprotonation of 17-H and 12-H in **97** gives stemod-13(17)-ene (**100**) and stemod-12-ene (**101**), respectively. The identified OsKSL8-catalyzed products were **98** (major) and **101** (minor).[18,22] The recombinant protein from *OsKSL11*, a possible allele of *OsKSL8* in IR24, produced mainly **100** and trace amounts of **98** and **101**.[68]

The genomic region flanking rice *OsKSL4* encodes not only *OsCPS4*, a *syn*-CPP synthase (see Section 1.17.2.1), but also two P-450-like genes (*CYP99A2* and *CYP99A3*) and one dehydrogenase-like gene (AK103462). The transcription of the three modification enzyme genes, as well as two cyclase genes, was drastically upregulated in rice suspension cells after elicitor treatment. The RNA interference (RNAi)-mediated



Scheme 16

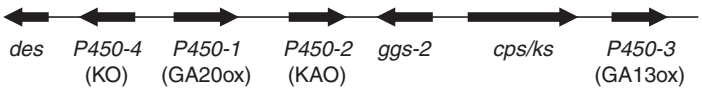
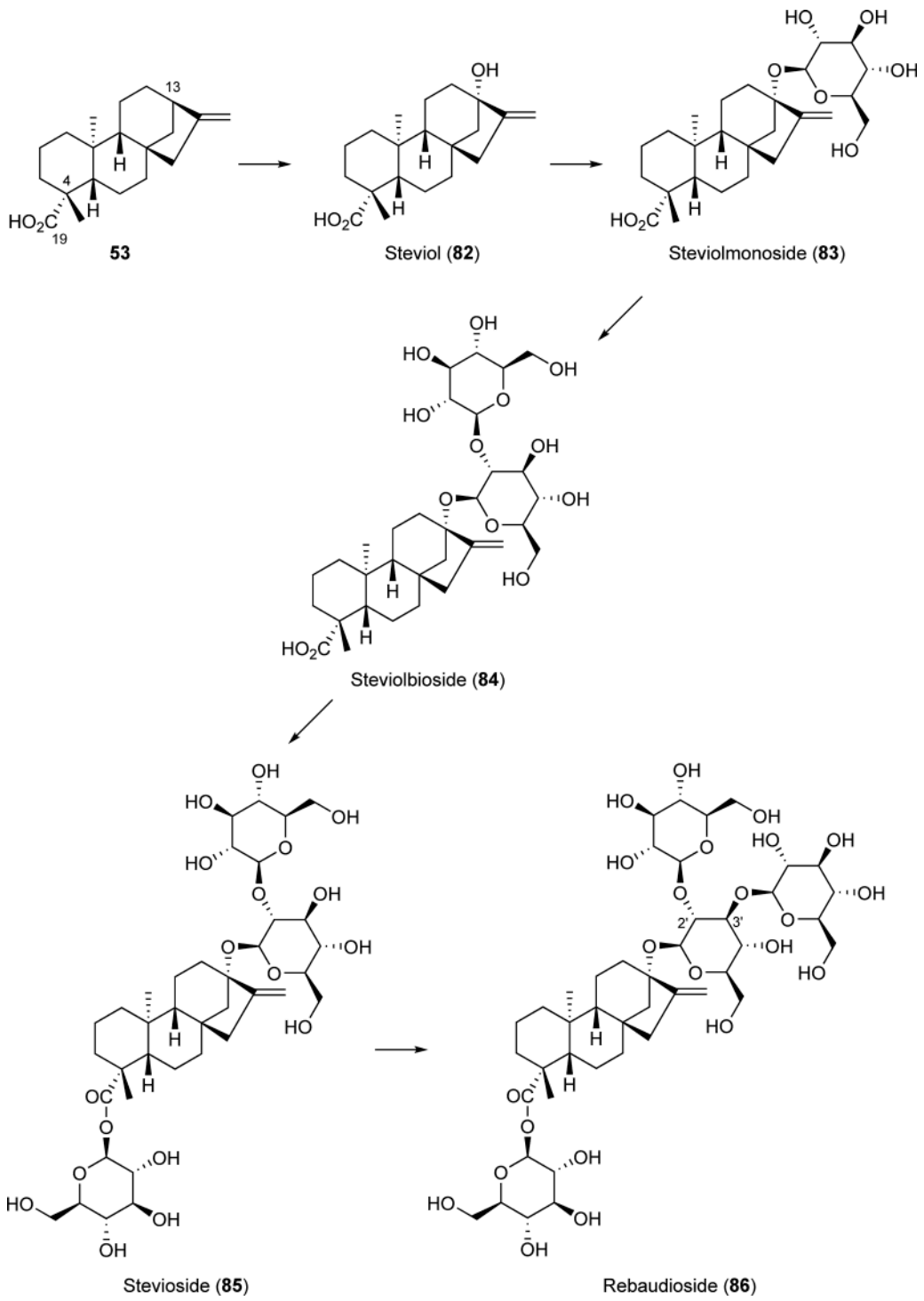


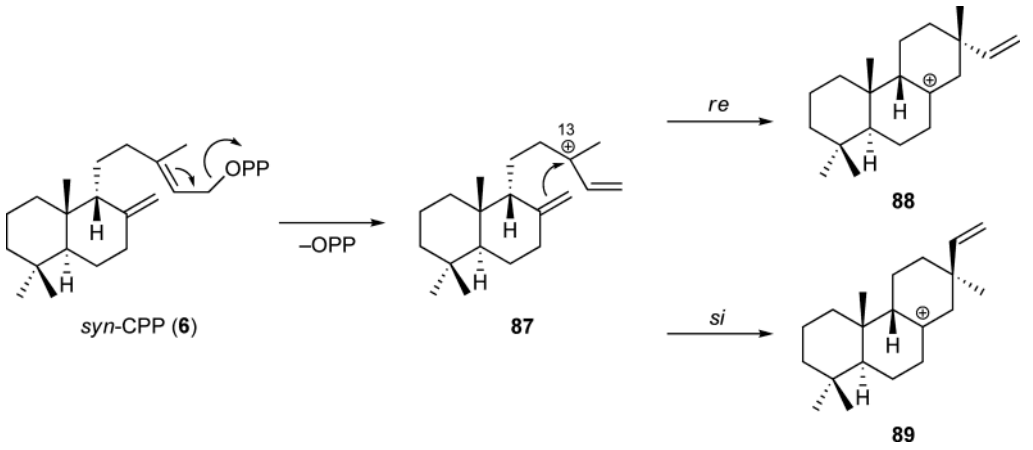
Figure 2 Gibberellin biosynthetic gene cluster in *Gibberella fujikuroi*. GGPP synthase, *ent*-CPP/*ent*-kaurene synthase, *ent*-kaurene oxidase, *ent*-kaurenoic acid oxidase, gibberellin 20-oxidase, gibberellin 13-hydroxylase, and GA_4 desaturase are encoded by *ggs-2*, *cps/ks*, *P-450-4*, *P-450-2*, *P-450-1*, *P-450-3*, and *des*, respectively.

downregulation of both *CYP99A2* and *CYP99A3*, which possess similar nucleotide sequences, resulted in the reduced production of **92** and **93** in rice cells, suggesting that *CYP99A2* and *CYP99A3* genes are responsible for the biosynthesis of momilactones. Furthermore, recombinant AK103462 catalyzed the conversion of

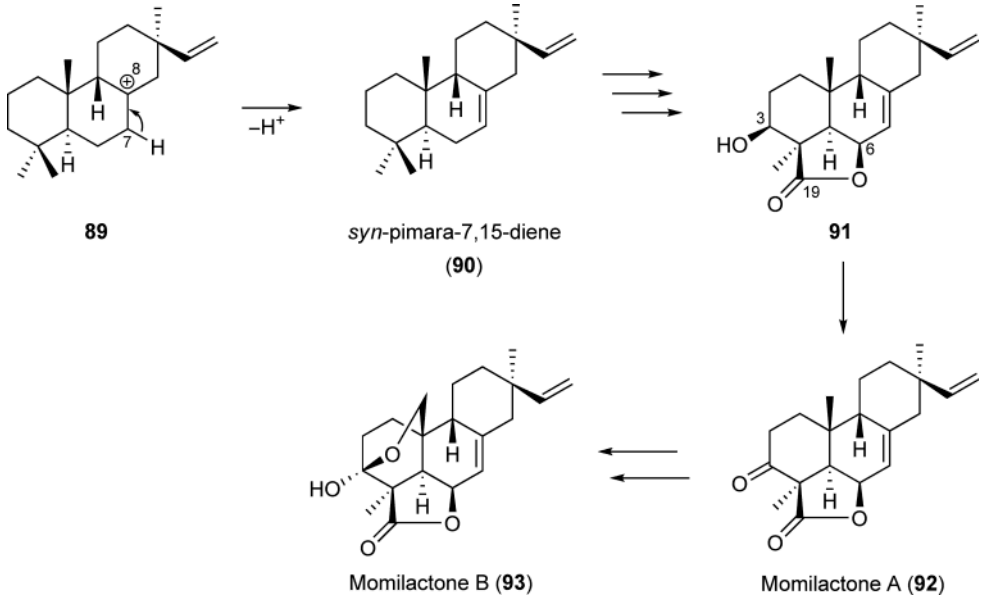


Scheme 17

3β-hydroxy-9β*H*-pimara-7,15-dien-19,6β-olide (**91**) into **92**, indicating that AK103462 encodes momilactone A synthase. These results strongly suggest that the biosynthetic genes of momilactones are clustered within the rice genome,[69] as well as in bacteria and fungi.



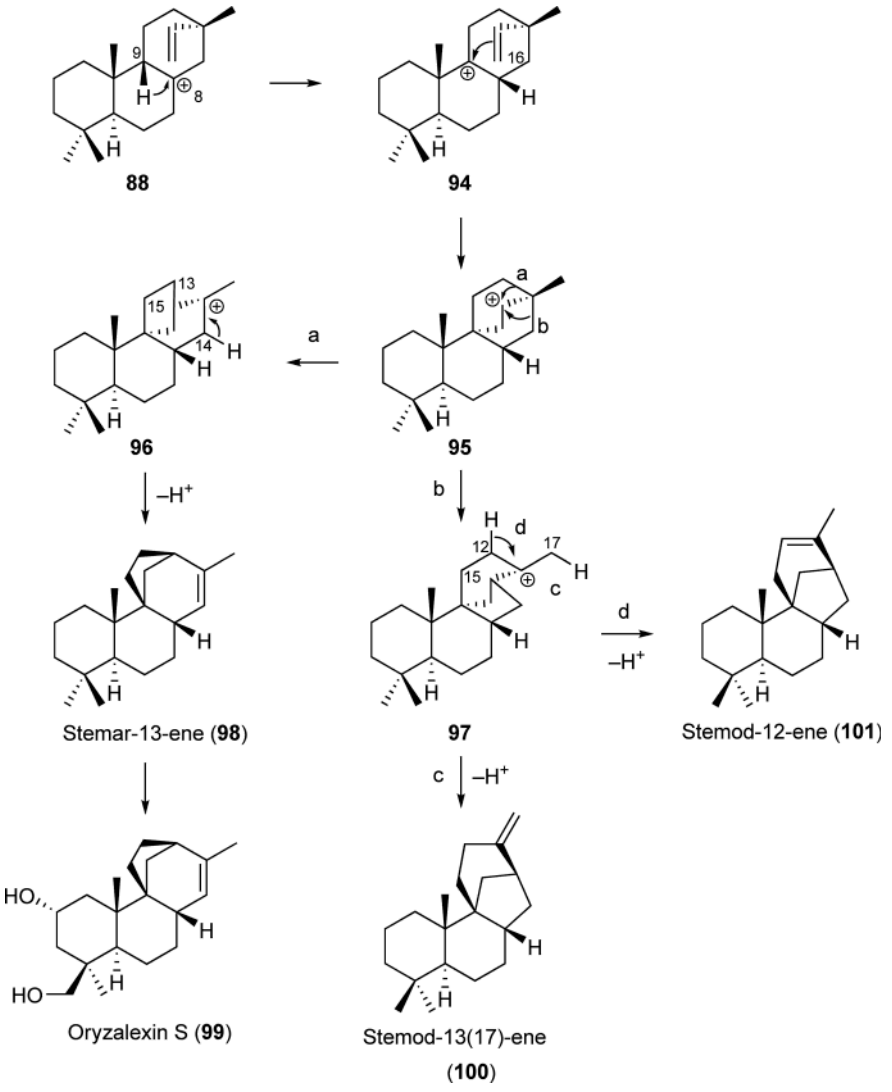
Scheme 18



Scheme 19

1.17.2.4.2 Aphidicolane-related diterpenes in a fungus

The plant-pathogenic fungus *Phoma betae* produces aphidicolin (**107** in **Scheme 21**), derived from *syn*-CPP (**6**). In physiological studies, aphidicolin (**107**) is used to induce cell cycle arrest via the specific inhibition of DNA polymerase α. A cDNA (*PbACS*) encoding the diterpene cyclase responsible for aphidicolin biosynthesis was isolated from the mycelia of *P. betae*.[70] PbACS (944 amino acids) converted GGPP into mainly aphidicolan-16β-ol (**104**) via *syn*-CPP (**6**). Aphidicol-16-ene (**105**) and aphidicol-15-ene (**106**) were also identified as minor by-products. The 1,2-hydride shift from C-9 to C-8 and successive attack of C-16 on C-9 in the C-8 carbocation **89**, derived from *syn*-CPP as shown in **Scheme 18**, yields the C-15 carbocation **102** (**Scheme 21**). The rearrangement of **102** produces the C-13 carbocation **103**. Quenching through the stereoselective capture of water (pathway a in **Scheme 21**) in **103** gives **104**. The deprotonation of 17-H (pathway b) and 12-H (pathway c) in **103** gives **105** and **106**, respectively (**Scheme 21**). These results suggest that pathway



Scheme 20

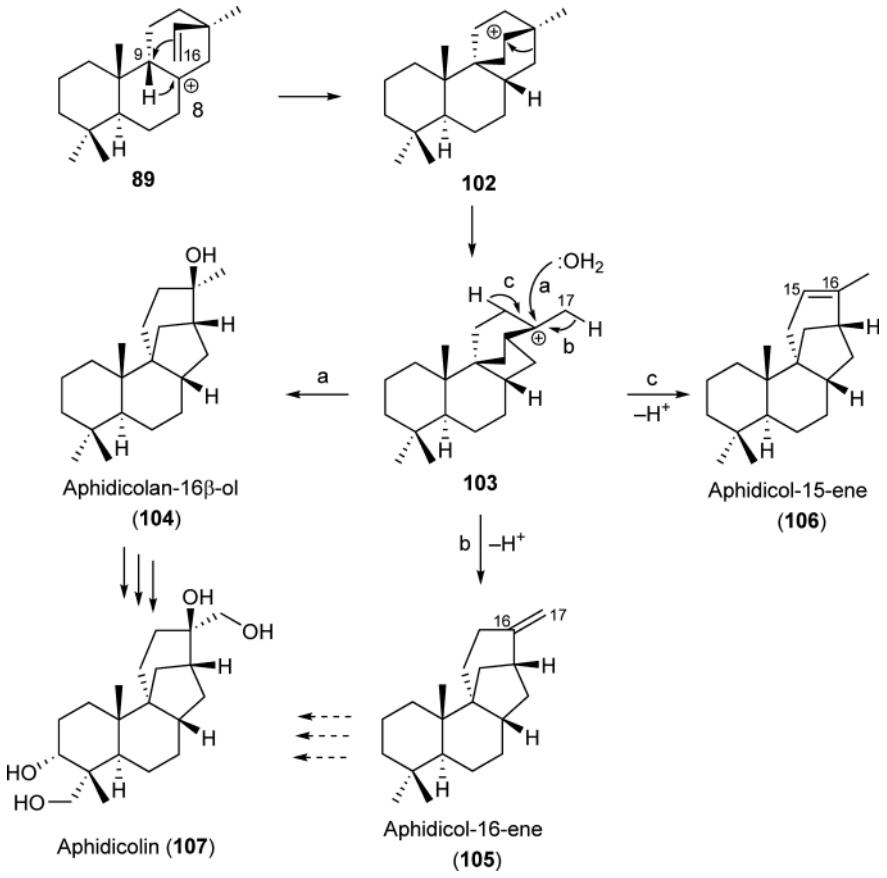
a is the major and pathway b is the minor pathway in the biosynthesis of **107**. The cyclization mechanism of PbACS was proposed based on experiments on biomimetic cyclization and *ab initio* calculation.[71]

Gene walking was used to identify two P-450-like genes and one GGPP synthase-like gene near the *PbACS* gene,[72] suggesting that the genes involved in aphidicolin biosynthesis are clustered within the *P. betae* genome, although the functions of these additional genes have not yet been determined.

1.17.3 Clerodane- and Halimane-Type Diterpenes

1.17.3.1 Clerodane

Terpentecin (**113**), a diterpene antibiotic, was identified as a metabolite in the eubacterium *S. griseolosporeus* MF730-N6. Through the identification of a gene cluster responsible for biosynthesis of **113**, two cyclase-like sequences (Cyc1 and Cyc2) were also found.[73] Cyc1 converted GGPP into terpentedienol diphosphate

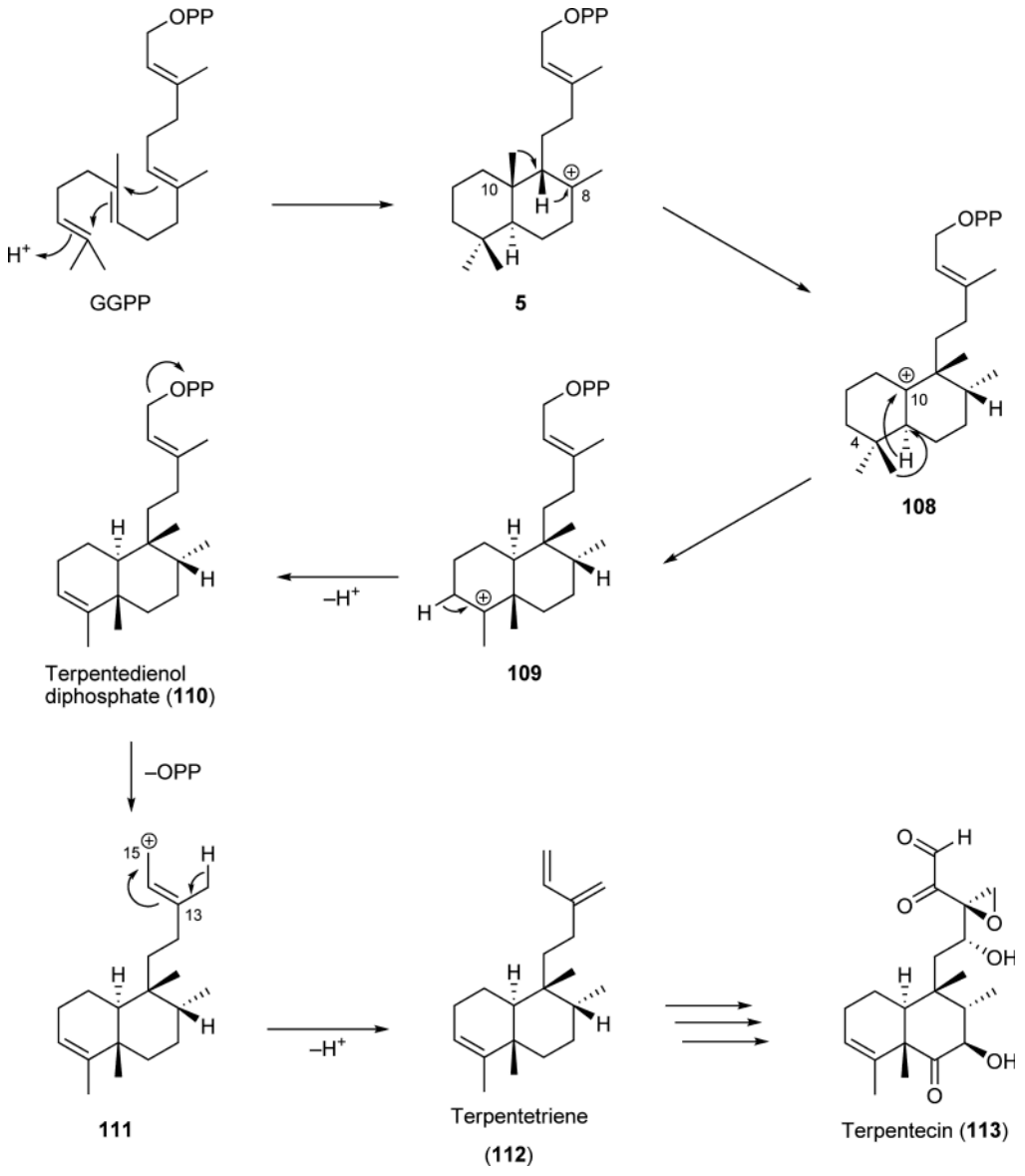


Scheme 21

(**110**), and Cyc2 converted **110** into terpentetriene (**112**) via the C-15 carbocation **111** (**Scheme 22**).[74] The amino acid sequences of Cyc1 (499 amino acids) and Cyc2 (311 amino acids) include a type B motif (DxDD) and a type A motif (DDxxD), and were similar to those of ORF2 (*ent*-CPP synthase; see Section 1.17.2.1) and ORF3 (*ent*-pimara-9(11),15-diene synthase; see Section 1.17.2.3.3) in *Streptomyces* sp. strain KO-3988, respectively. The type A Cyc2 possessed phosphatase activity, and used not only **110** but also GGPP and farnesyl diphosphate (FPP) as substrates. The proposed mechanism of the reaction from GGPP to **110** catalyzed by type B Cyc1 is as follows: the 1,2-hydride shift from C-9 to C-8 in the C-8 carbocation **5**, derived from GGPP (see **Scheme 1**), and successive 1,2-methyl shift from C-10 to C-9 gives the C-10 carbocation **108**. The 1,2-hydride shift from C-5 to C-10 in **108** and successive 1,2-methyl shift from C-4 to C-5 gives the C-4 carbocation **109**. Finally, **110**, a *syn-trans*-clerodane, is formed by the deprotonation of 3-H in **109**.

1.17.3.2 Halimane

A type B cyclase (Rv3377, 501 amino acids), whose primary structure was similar to that of eubacterial Cyc1 (see Section 1.17.3.1), was identified in *Mycobacterium tuberculosis* H37.[75] As shown in **Scheme 23**, Rv3377 converted GGPP into tuberculosinol diphosphate (**116**) via the C-8 carbocation **1** (see **Scheme 1**). The 1,2-hydride shift from C-9 to C-8 in **1** and 1,2-methyl shift from C-10 to C-9 produces the C-10 carbocation **114**. The 1,2-hydride shift from C-5 to C-10 in **114** gives the C-5 carbocation **115** and, finally, the deprotonation of 6-H in **115** gives **116**.

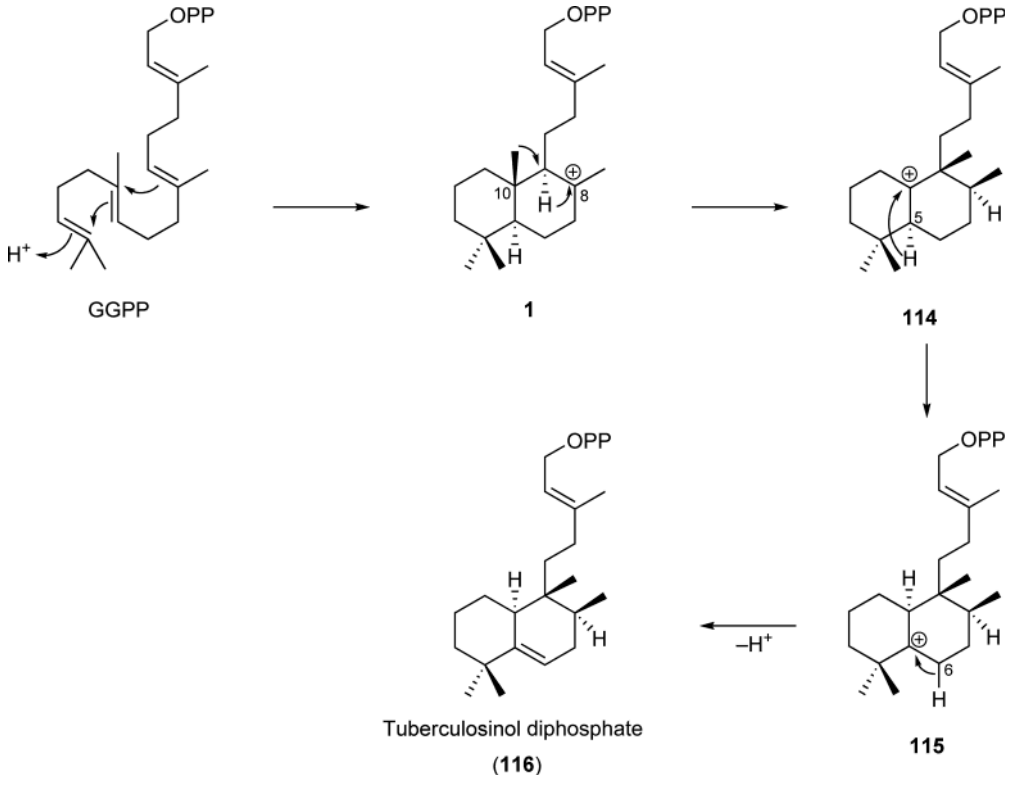


Scheme 22

1.17.4 Other types of Diterpenes

1.17.4.1 Taxane and Phomactane

Taxol (**126** in **Scheme 24**) is a potent anticancer drug that was originally isolated from the Pacific yew. It was shown that **126** is biosynthesized from GGPP through several steps via taxa-4(5),11(12)-diene (**120**) as an actual intermediate cyclic diterpene hydrocarbon, as reviewed by Croteau *et al.*[76] As described in Section 1.17.1, the taxadiene synthase (862 amino acids) gene was isolated from *T. brevifolia* (**Figure 1(e)**),[5] and several orthologs have been identified from other yew species, such as *Taxus chinensis*[77] and *Taxus* × *media*.[78] The mechanism of the reaction catalyzed by taxadiene synthase has been elucidated (**Scheme 24**).[76] The attack of C-14 on C-1 and successive attack of C-10 on C-15 after ionization of the diphosphate in GGPP (type A) produces a verticillen-12-yl cation (**117**). Intramolecular proton 1,5-migration from C-11 to C-7 gives the C-8



Scheme 23

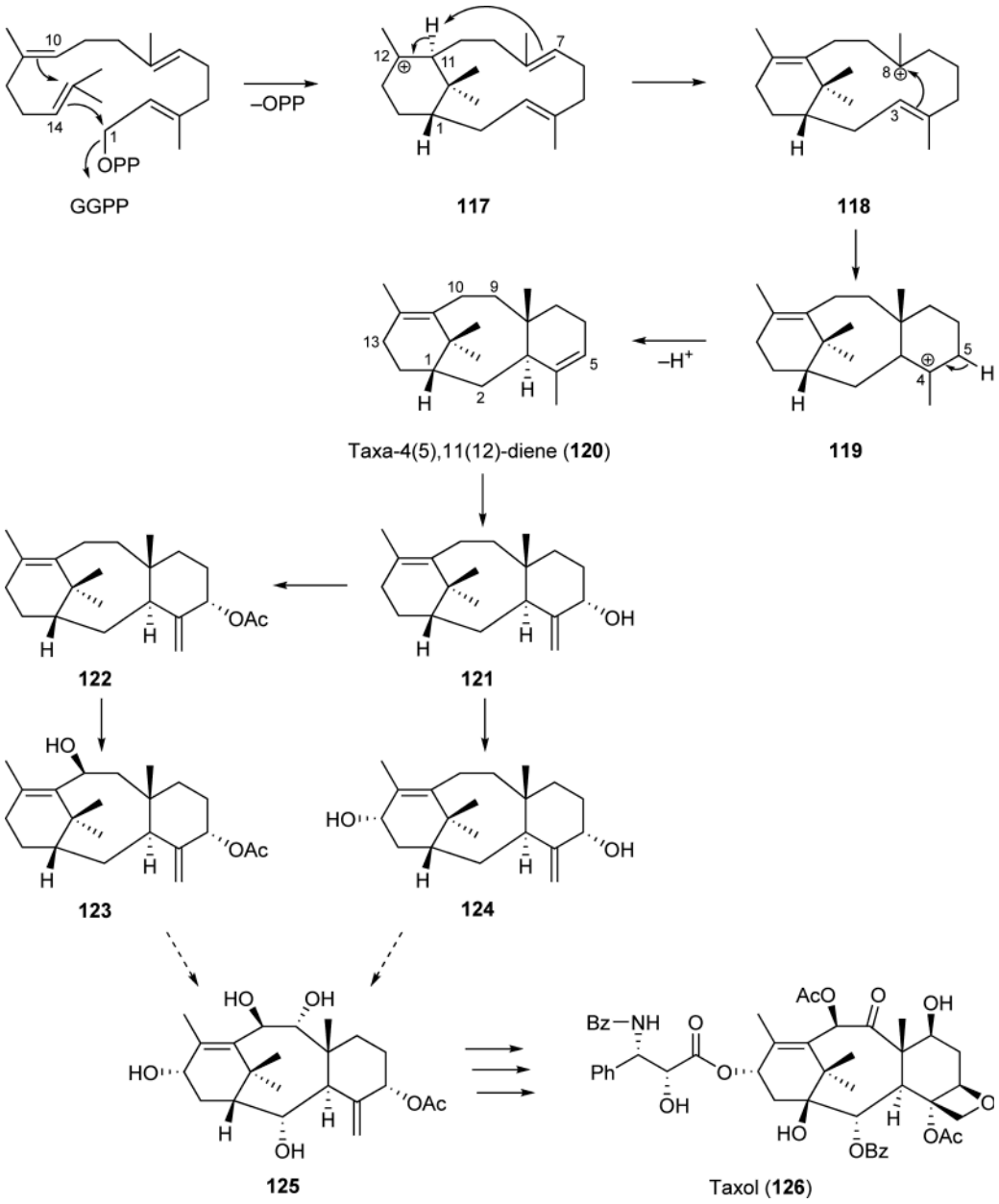
carbocation **118**. This step includes type B cyclization. Another pathway was recently suggested by theoretical calculations that a subsequent intramolecular proton migration from C-11 to C-3, and from C-3 to C-7 gives **118**.[79] The attack of C-3 on C-8 in **118** gives the C-4 carbocation **119**, and the deprotonation of 5-H in **119** provides **120**.

On the other hand, phomacta-1(14),3,7-triene (**130**) is also converted from verticillen-12-yl cation **117**, and the mechanism of this reaction has been investigated (**Scheme 25**).[80,81] Compound **130** is a biosynthetic precursor of phomactins (e.g., **131**), which are potent PAF antagonists. The 1,2-hydride shift from C-11 to C-12 in the intermediate compound **117** gives the C-11 carbocation **127**, and a 1,2-methyl shift from C-15 to C-11 in **127** provides the C-15 carbocation **128**. The 1,2-hydride shift from C-1 to C-15 in **128** gives the C-1 carbocation **129**, and the deprotonation of 14-H in **129** provides **130**. However, the diterpene cyclase responsible for the conversion of **130** from GGPP has not yet been identified.

Chemical modification enzymes, which are also involved in taxol (**126**) biosynthesis, have been identified (**Scheme 24**). Taxadiene hydroxylase, a member of the P-450 monooxygenase family, converts **120** into taxa-4(20),11(12)-dien-5α-ol (**121**) by 5α-hydroxylation.[82] An acetyltransferase that converts **121** to taxa-4(20),11(12)-dien-5α-ol-acetate (**122**) and a P-450 enzyme that converts **122** into taxa-4(20),11(12)-dien-5α,10β-diol-5-acetate (**123**) by 10β-hydroxylation have been cloned.[83,84] Moreover, P-450 taxoid 13α-hydroxylase, which converts **121** into taxa-4(20),11(12)-dien-5α,13α-diol (**124**), was also identified.[85] However, taxoid 10β-hydroxylase could not use **124** as a substrate and, in contrast, taxoid 13α-hydroxylase could not use **123**. Therefore, the precise order of hydroxylation and acetylation remains unclear, although **125** has been suggested as a hypothetical intermediate of **126**.[76]

1.17.4.2 Casbene and Cembratriene

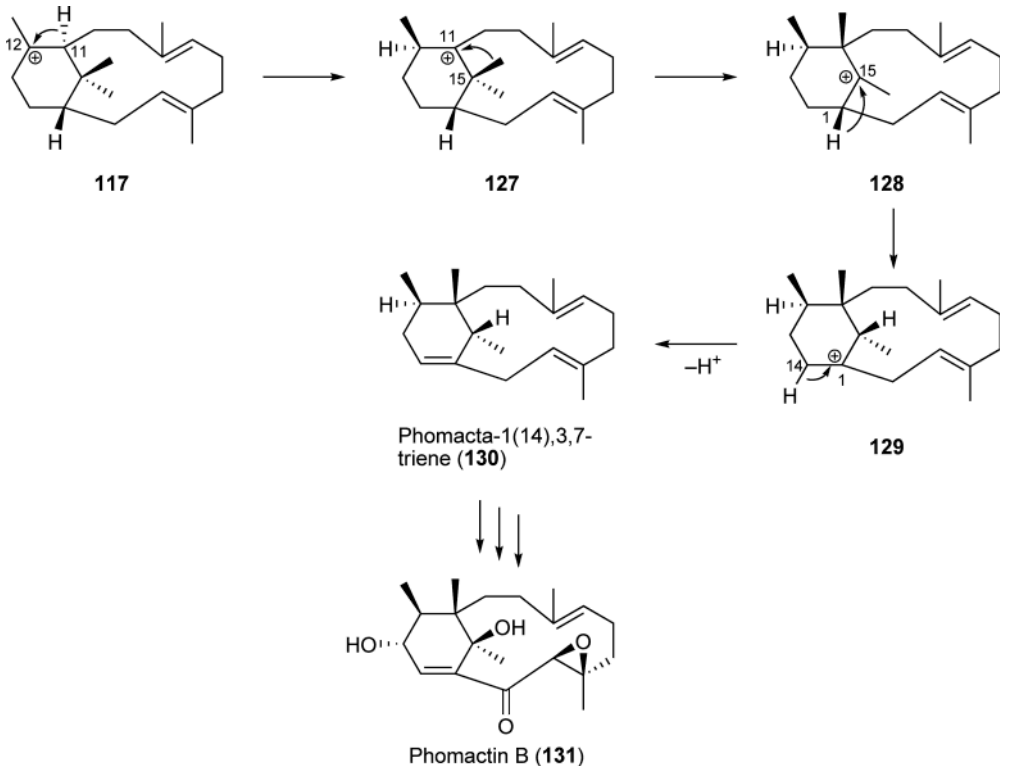
Casbene (**133** in **Scheme 26**) is a novel bicyclic diterpene hydrocarbon that serves as a phytoalexin in castor bean (*R. communis*). As described in Section 1.17.1, the cDNA for the casbene synthase that catalyzes the conversion of GGPP into **133** (**Scheme 26**) has been cloned (**Figure 1(c)**).[1] Following ionization of the



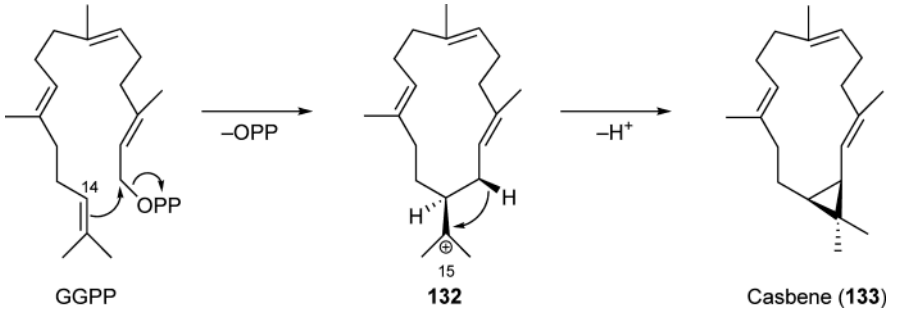
Scheme 24

diphosphate in GGPP, an attack on the *re*-face of the 14,15-double bond at C-1 provides the C-15 carbocation **132**. The stabilization of the intermediate **132** via proton abstraction gives **133**.

The type A diterpene cyclase gene, *CYC-1*, which is responsible for the biosynthesis of cembratriene 4,6α-diol isomers (**137** in **Scheme 27**), was identified from tobacco (*Nicotiana tabacum*).[86] Cembratriene-related compounds accumulate in trichomes. RNAi experiments have clearly indicated that CYC-1 (597 amino acids) converts GGPP into cembratriene 4-ol isomers (**136**). The 1,3-hydride shift from C-2 to C-15 of the intermediate compound **134** provides the C-4 carbocation **135**, and the nonstereospecific capture of water by **135** gives **136**. RNAi experiments have also shown that CYP71D16, a member of the P-450 monooxygenase family, catalyzes the conversion of **136** into **137** by 6α-hydroxylation.[86] The transcripts of both these genes accumulate substantially in tobacco trichomes.



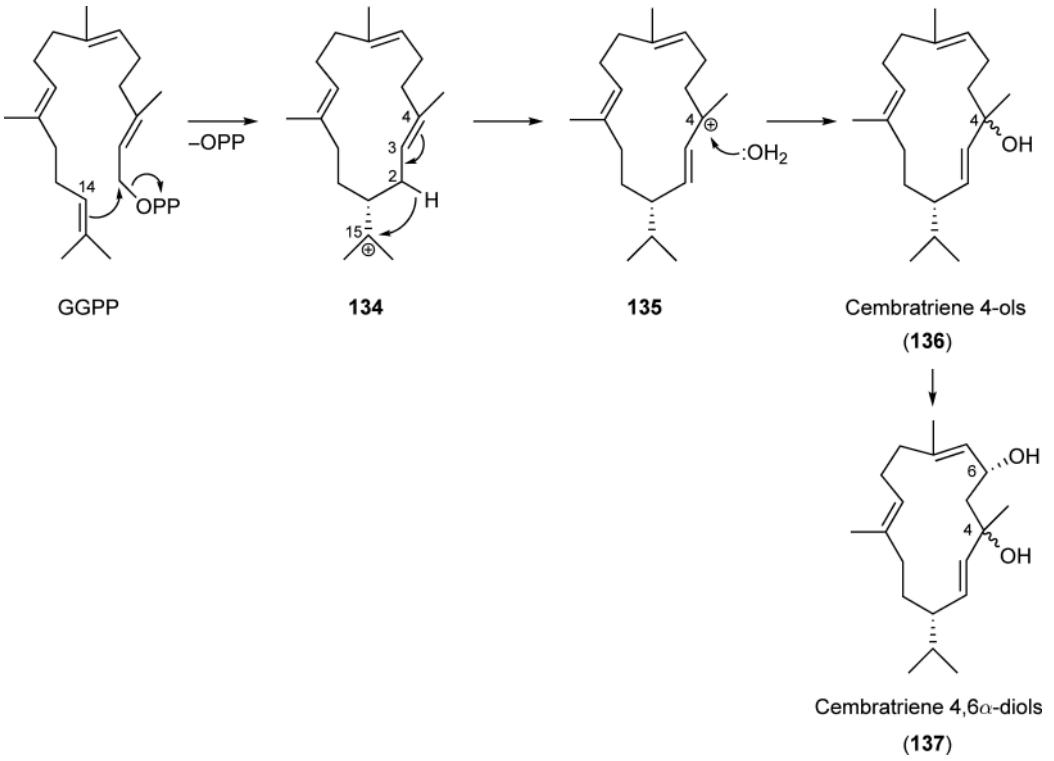
Scheme 25



Scheme 26

1.17.4.3 Fusicoccane

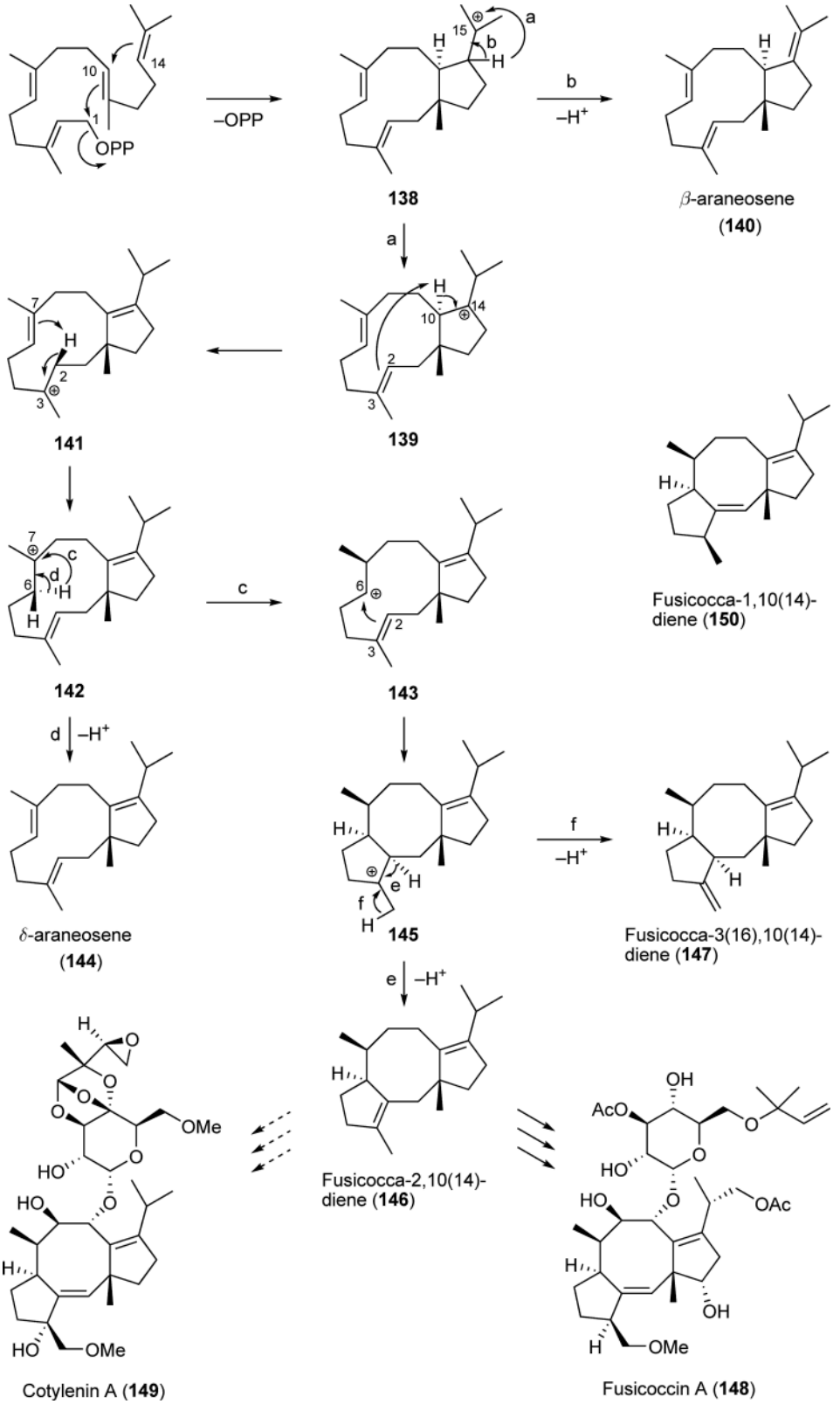
Fusicoccins (e.g., **148** in **Scheme 28**) are diterpene glucosides, produced by the plant-pathogenic fungus *P. amygdali*. Compound **148** exhibits auxin-like activity on plant cells via the continuous activation of proton-ATPase through the formation of a ternary complex consisting of 14-3-3 protein, **148**, and a phosphopeptide from the C-terminus of proton-ATPase.[87] Cotylenin A (**149**), a fusicoccane-related metabolite produced by *Cladosporium* sp. 501-7W, not only shows phytohormone-like activity, but also induces the differentiation of a series of human cells.[88] It was previously thought that **148** was derived from GGPP via fusicocca-1,10(14)-diene (**150**), a possible intermediate tricyclic hydrocarbon, as reviewed by MacMillan and Beale.[6] However, a detailed search of cyclic hydrocarbons and feeding experiments using labeled compounds in the mycelia of *P. amygdali* indicated that fusicocca-2,10(14)-diene (**146**) is a true biosynthetic intermediate hydrocarbon of **148**.[89–91] How **146** might also be an intermediate of **149** remains unexplained. Furthermore, a cDNA encoding fusicoccadiene



Scheme 27

synthase (PaFS, 719 amino acids) was isolated from the mycelia of *P. amygdali* and characterized.[92] PaFS converted GGPP into mainly **146**. The minor by-products β-araneosene (**140**), δ-araneosene (**144**), and fusicocca-3(16),10(14)-diene (**147**) were also produced, all of which were isolated from the mycelia. Experiments using recombinant PaFS and labeled GGPPs, together with previous information, have led to a newly proposed mechanism for the formation of **146** from GGPP (**Scheme 28**).[93] After ionization of the diphosphate in GGPP, the attack of the 10,11-double bond on C-1 and successive attack of the 14,15-double bond on C-10 gives the C-15 carbocation **138**. The 1,2 hydride shift from C-14 to C-15 in **138** (pathway a) gives the C-14 carbocation **139** or, alternatively, the deprotonation of 14-H in **138** (pathway b) gives **140**. A subsequent intramolecular proton 1,4-migration from C-10 to C-2β and from C-2β to C-6β gives the C-7 carbocation **142** via **141**. The possibility of alternative pathway cannot be excluded that 1,5-proton migration from C-10 to C-6β gives **142**. The 1,2-hydride shift from C-6α to C-7α in **142** (pathway c) gives the C-6 cation **143**, whereas the deprotonation of 6α-H in **142** (pathway d) gives **144**. The attack of C-2 on C-6 in **143** provides the C-3 carbocation **145**, and the subsequent deprotonation of 2-H (pathway e) and 16-H (pathway f) in **145** gives **146** and **147**, respectively.

PaFS possesses an unusual primary structure and multifunctionality.[92] It consists of two domains, the diterpene cyclase domain at the N-terminus and the prenyltransferase domain at the C-terminus, that are responsible for the synthesis of **146** from GGPP and GGPP synthesis from isoprene units, respectively (**Figure 3**). The amino acid sequence of the diterpene cyclase domain is more similar to those of aristolochene synthases in fungi, rather than other identified fungal diterpene cyclases, and includes a type A motif. The prenyltransferase domain has a DDxxD motif, which is generally included in the GGPP synthases. PaFS can produce **146** in any cell in which isoprene units or FPP for sterol synthesis is present, due to the ability to supply a GGPP substrate by itself. Homologous genes of a chimeric *PaFS* are found in the databases of fungi other than *P. amygdali*. In the flanking region of genomic DNA near the *PaFS* gene, a P-450-like gene, a dioxygenase-like gene, and a short-chain dehydrogenase/reductase-like gene were found, suggesting that some fusicoccin biosynthetic genes are clustered within the genome of *P. amygdali*.[92]


–OPP
b
–H+
138
β-araneosene
(140)
a
141
139
Fusicocca-1,10(14)-
diene (150)
c
142
143
d –H+
δ-araneosene
(144)
145
f
–H+
Fusicocca-3(16),10(14)-
diene (147)
e –H+
Fusicocca-2,10(14)-
diene (146)
Cotylenin A (149)
Fusicoccin A (148)

Scheme 28

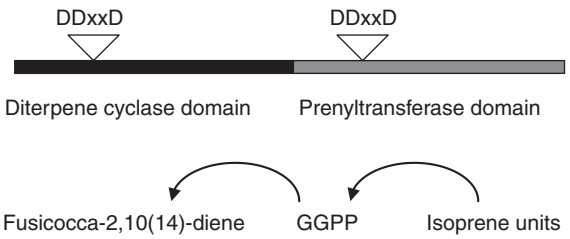
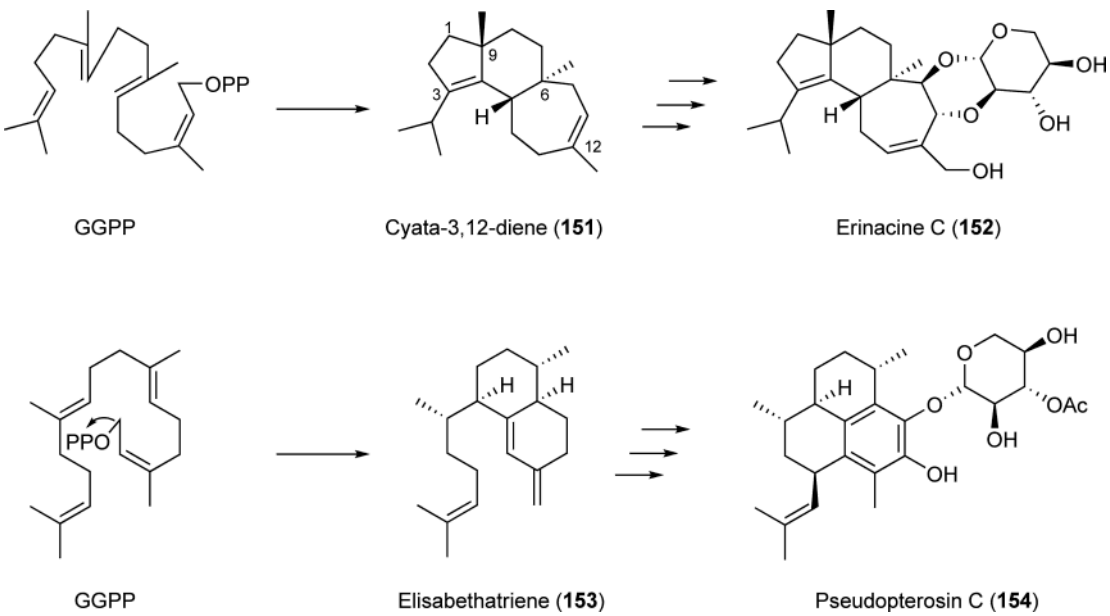


Figure 3 Schematic diagram of the primary structure of fusicoccadiene synthase (PaFS) in *Phomopsis amygdali*. Black bar and gray bar indicate diterpene cyclase domain and prenyltransferase domain, which are responsible for fusicocca-2,10(14)-diene synthesis and GGPP synthesis, respectively. Reverse triangles indicate aspartate-rich motifs.

1.17.5 Summary and Future Prospects

A large number of diterpenes have been isolated from various organisms. These diterpenes include not only antibiotics, but also physiologically active substances, such as phytohormones and drugs. Structural and functional diversity is achieved through variations in carbon skeleton formation and chemical modification. As described in this chapter, a variety of biosynthetic enzymes and genes for diterpene biosynthesis have been identified. In particular, a wide variety of diterpene carbon skeletons are derived from the common precursor GGPP by their corresponding cyclases. Accumulated information on diterpene synthases gives us two important insights. (1) Diversity: A single cyclase produces several by-products in addition to its main product, and a change in one or several amino acid residues contributes to a change in the products. This means that minor changes in the primary structure of a cyclase can yield cyclases that produce different products. In fact, rice produces a variety of diterpenes by the action of a multigene family of diterpene cyclases that may have evolved through gene duplication. (2) Efficiency: Several gene clusters for diterpene biosynthesis have been identified in bacteria, fungi, and rice, a higher plant. Diterpene cyclase fused with GGPP synthase was found in fungi.

A further insight into diterpene biosynthesis will require the identification of other diterpene cyclases, such as cyata-3,12-diene (**151**) synthase and elisabethatriene (**153**) synthase (**Scheme 29**). Compounds **151** and **153** are the possible biosynthetic intermediate hydrocarbons of erinacines (e.g., **152**) in *Hericium erinaceum*[94] and of



Scheme 29

pseudopterosins (e.g., **154**) in *Pseudopterogorgia elisabethae*,[95] respectively. To date, the X-ray structures of these diterpene cyclases have not been determined. More information, including 3D structures of cyclases, will provide a new insight into catalytic mechanisms and how to manipulate the products or stereochemical outcome. The identification of biosynthetic genes will enable us to produce diterpenes, which are produced in very small quantities in a given organism, using heterologous expression systems in familiar bacteria and yeast. Moreover, finding a gene cluster can lead to the discovery of a series of biosynthetic genes, the manipulation of which in heterologous or homologous cells will allow the production of structurally and functionally modified diterpenes.

Abbreviations

CPP	copalyl diphosphate
FPP	farnesyl diphosphate
GA2ox	gibberellin 2-oxidase
GA20ox	gibberellin 20-oxidase
GA3ox	gibberellin 3-oxidase
GA7ox	gibberellin 7-oxidase
GGPP	geranylgeranyl diphosphate
KAO	kaurenoic acid oxidase
PAF	platelet-activating factor
RNAi	RNA interference
UGT	UDP-glucosyltransferase
UV	ultraviolet

References

1. C. J. D. Mau; C. A. West, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8497–8501.
2. T.-P. Sun; Y. Kamiya, *Plant Cell* **1994**, *6*, 1509–1518.
3. S. Yamaguchi; T. Saito; H. Abe; H. Yamane; N. Murofushi; Y. Kamiya, *Plant J.* **1996**, *10*, 203–213.
4. B. Stofer Vogel; M. R. Wildung; G. Vogel; R. Croteau, *J. Biol. Chem.* **1996**, *271*, 23262–23268.
5. M. R. Wildung; R. Croteau, *J. Biol. Chem.* **1996**, *271*, 9201–9204.
6. J. MacMillan; M. H. Beale, Diterpene Biosynthesis. In *Comprehensive Natural Products Chemistry*; S. D. Barton, K. Nakanishi, Eds.; Elsevier Science Ltd.: Oxford, 1999; Vol. 2, pp 217–243.
7. S. Yamaguchi, *Annu. Rev. Plant Biol.* **2008**, *59*, 225–251.
8. K. Hayashi; H. Kawaide; M. Notomi; Y. Sakigi; A. Matsuo; H. Nozaki, *FEBS Lett.* **2006**, *580*, 6175–6181.
9. H. Kawaide; R. Imai; T. Sassa; Y. Kamiya, *J. Biol. Chem.* **1997**, *272*, 21706–21712.
10. T. Toyomasu; H. Kawaide; A. Ishizaki; S. Shinoda; M. Otuska; W. Mitsuhashi; T. Sassa, *Biosci. Biotechnol. Biochem.* **2000**, *64*, 660–664.
11. H. Kawaide; T. Sassa; Y. Kamiya, *J. Biol. Chem.* **2000**, *275*, 2276–2280.
12. R. J. Peters; R. B. Croteau, *Biochemistry* **2002**, *41*, 1836–1842.
13. T. Kawasaki; T. Kuzuyama; Y. Kuwamori; N. Matsuura; N. Itoh; K. Furihata; H. Seto; T. Dairi, *J. Antibiot.* **2004**, *57*, 739–747.
14. T. Kawasaki; Y. Hayashi; T. Kuzuyama; K. Furihata; N. Itoh; H. Seto; T. Dairi, *J. Bacteriol.* **2006**, *188*, 1236–1244.
15. K. Otomo; H. Kenmoku; H. Oikawa; W. A. König; H. Toshima; W. Mitsuhashi; H. Yamane; T. Sassa; T. Toyomasu, *Plant J.* **2004**, *39*, 886–893.
16. T. Sakamoto; K. Miura; H. Itoh; T. Tatsumi; M. Ueguchi-Tanaka; K. Ishiyama; M. Kobayashi; G. K. Agrawal; S. Takeda; K. Abe; A. Miyao; H. Horichika; H. Kitano; M. Ashikari; M. Matsuoka, *Plant Physiol.* **2004**, *134*, 1642–1653.
17. Y. Hayashi; T. Toyomasu; Y. Hirose; Y. Onodera; W. Mitsuhashi; H. Yamane; T. Sassa; T. Dairi, *Biosci. Biotechnol. Biochem.* **2008**, *72*, 523–530.
18. M. Xu; P. R. Wilderman; D. Morrone; J. Xu; A. Roy; M. Margis-Pinherio; N. M. Upadhyaya; R. M. Coates; R. J. Peters, *Phytochemistry* **2007**, *68*, 312–326.
19. K. Otomo; Y. Kanno; A. Motegi; H. Kenmoku; H. Yamane; W. Mitsuhashi; H. Oikawa; H. Toshima; H. Itoh; M. Matsuoka; T. Sassa; T. Toyomasu, *Biosci. Biotechnol. Biochem.* **2004**, *68*, 2001–2006.
20. Y. Kanno; K. Otomo; H. Kenmoku; W. Mitsuhashi; H. Yamane; H. Oikawa; H. Toshima; M. Matsuoka; T. Sassa; T. Toyomasu, *Biosci. Biotechnol. Biochem.* **2006**, *70*, 1702–1710.
21. E. M. Cho; A. Okada; H. Kenmoku; K. Otomo; T. Toyomasu; W. Mitsuhashi; T. Sassa; A. Yajima; G. Yabuta; K. Mori; H. Oikawa; H. Toshima; N. Shibuya; H. Nojiri; H. Omori; M. Nishiyama; H. Yamane, *Plant J.* **2004**, *37*, 1–8.

22. T. Nemoto; E. M. Cho; A. Okada; K. Okada; K. Otomo; Y. Kanno; T. Toyomasu; W. Mitsuhashi; T. Sassa; E. Minami; N. Shibuya; M. Nishiyama; H. Nojiri; H. Yamane, *FEBS Lett.* **2004**, *571*, 182–186.
23. R. J. Peters, *Phytochemistry* **2006**, *67*, 2307–2317.
24. T. Toyomasu, *Biosci. Biotechnol. Biochem.* **2008**, *72*, 1168–1175.
25. S. Prisic; M. Xu; P. R. Wilderman; R. J. Peters, *Plant Physiol.* **2004**, *136*, 4228–4236.
26. M. Xu; M. L. Hillwig; S. Prisic; R. M. Coates; R. J. Peters, *Plant J.* **2004**, *39*, 309–318.
27. T. Toyomasu; R. Niida; H. Kenmoku; Y. Kanno; S. Miura; C. Nakano; Y. Shiono; W. Mitsuhashi; H. Toshima; H. Oikawa; T. Hoshino; T. Dairi; N. Kato; T. Sassa, *Biosci. Biotechnol. Biochem.* **2008**, *72*, 1038–1047.
28. R. J. Peters; M. W. Ravn; R. M. Coates; R. B. Croteau, *J. Am. Chem. Soc.* **2001**, *123*, 8974–8978.
29. R. J. Peters; J. E. Flory; R. Jetter; M. M. Ravn; H.-J. Lee; R. M. Coates; R. B. Croteau, *Biochemistry* **2000**, *39*, 15592–15602.
30. D.-K. Ro; G. Arimura; S. Y. Lau; E. Piers; J. Bohlmann, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 8060–8065.
31. R. J. Peters; R. B. Croteau, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 580–584.
32. R. J. Peters; O. A. Carter; Y. Zhang; B. W. Matthews; R. B. Croteau, *Biochemistry,* **2003**, *42*, 2700–2707.
33. D. M. Martin; J. Fäldt; J. Bahlmann, *Plant Physiol.* **2004**, *135*, 1908–1927.
34. D.-K. Ro; J. Bohlmann, *Phytochemistry* **2006**, *67*, 1572–1578.
35. H. G. Schepmann; J. Pang; S. P. Matsuda, *Arch. Biochem. Biophys.* **2001**, *392*, 263–269.
36. C. I. Keeling; J. Bohlmann, *Phytochemistry* **2006**, *67*, 2415–2423.
37. P. R. Wilderman; R. J. Peters, *J. Am. Chem. Soc.* **2007**, *129*, 15736–15737.
38. C. I. Keeling; S. Weisshaar; R. P. C. Lin; J. Bohlmanan, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1085–1090.
39. H. Kenmoku; M. Tanaka; K. Ogiyama; N. Kato; T. Sassa, *Biosci. Biotechnol. Biochem.* **2004**, *68*, 1574–1577.
40. Y. Kono; A. Kojima; R. Nagai; M. Watanabe; T. Kawashima; T. Onizawa; T. Teraoka; M. Watanabe; H. Koshino; J. Umezawa; Y. Suzuki; A. Sakurai, *Phytochemistry* **2004**, *65*, 1291–1298.
41. M. Xu; P. R. Wilderman; R. J. Peters, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 7397–7401.
42. J. E. Brandle; P. G. Telmer, *Phytochemistry* **2007**, *68*, 1855–1863.
43. C. Ikeda; Y. Hayashi; N. Itoh; H. Seto; T. Dairi, *J. Biochem.* **2007**, *141*, 37–45.
44. S. Yamaguchi; T.-P. Sun; H. Kawaide; Y. Kamiya, *Plant Physiol.* **1998**, *116*, 1271–1278.
45. C. A. Helliwell; C. C. Sheldon; M. R. Olive; A. R. Walker; J. A. D. Zeevaart; W. J. Peacock; E. S. Dennis, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9019–9024.
46. C. A. Helliwell; A. Poole; W. J. Peacock; E. S. Dennis, *Plant Physiol.* **1999**, *119*, 507–510.
47. C. A. Helliwell; P. M. Chandler; A. Pool; E. S. Dennis; W. J. Peacock, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2065–2070.
48. T. Lange, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6553–6558.
49. A. L. Phillips; D. A. Ward; S. Uknes; N. E. Appleford; T. Lange; A. K. Huttly; P. Gaskin; J. E. Graebe; P. Hedden, *Plant Physiol.* **1995**, *108*, 1049–1057.
50. T. Lange; P. Hedden; J. E. Graebe, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8552–8556.
51. H. H. Chiang; I. Hwang; H. M. Goodman, *Plant Cell* **1995**, *7*, 195–201.
52. J. Williams; A. L. Phillips; P. Gaskin; P. Hedden, *Plant Physiol.* **1998**, *117*, 559–563.
53. S. G. Thomas; A. L. Phillips; P. Hedden, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4698–4703.
54. F. M. Scomburg; C. M. Bizzell; D. J. Lee; J. A. D. Zeevaart; R. M. Amasino, *Plant Cell* **2003**, *15*, 151–163.
55. Y. Zhou; T. Nomura; Y. Xu; Y. Zhang; Y. Peng; B. Mao; A. Hanada; H. Zhou; X. Wang; P. Li; X. Zhu; L. N. Mander; Y. Kamiya; S. Yamaguchi; Z. He, *Plant Cell* **2005**, *18*, 442–456.
56. M. Varbanvoa; S. Yamaguchi; Y. Yang; K. McKelvey; A. Hanada; R. Borochov; F. Yu; Y. Jikumaru; J. Ross; D. Cortes; C. J. Ma; J. P. Noel; L. N. Mander; V. Shulaev; Y. Kamiya; S. Rodermel; D. Weiss; E. Pichersky, *Plant Cell* **2007**, *19*, 32–45.
57. B. Tudzynski; K. Hölter, *Fungal Genet. Biol.* **1998**, *25*, 157–170.
58. B. Tudzynski; P. Hedden; E. Carrera; P. Gaskin, *Appl. Environ. Microbiol.* **2001**, *67*, 3514–3522.
59. M. C. Rojas; P. Hedden; P. Gaskin; B. Tudzynski, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5838–5843.
60. B. Tudzynski; M. C. Rojas; P. Gaskin; P. Hedden, *J. Biol. Chem.* **2002**, *277*, 21246–21253.
61. B. Tudzynski; M. Mihlan; M. C. Rojas; P. Linnemannstöns; P. Gaskin; P. Hedden, *J. Biol. Chem.* **2003**, *278*, 28635–28643.
62. H. Kawaide, *Biosci. Biotechnol. Biochem.* **2006**, *70*, 583–590.
63. J. E. Brandle; P. G. Telmer, *Phytochemistry* **2007**, *68*, 1855–1863.
64. A. S. Richman; M. Gijzen; A. N. Starratt; Z. Yang; J. E. Brandle, *Plant J.* **1999**, *19*, 411–421.
65. T. V. Humphrey; A. S. Richman; R. Menassa; J. E. Brandle, *Plant Mol. Biol.* **2006**, *61*, 47–62.
66. A. S. Richman; A. Swanson; T. Humphrey; R. Chapman; B. McGarvey; R. Pocs; J. Brandle, *Plant J.* **2005**, *41*, 56–67.
67. P. R. Wilderman; M. Xu; Y. Jin; R. M. Coates; R. J. Peters, *Plant Physiol.* **2004**, *135*, 2098–2105.
68. D. Morrone; Y. Jin; M. Xu; S.-Y. Choi; R. M. Coates; R. J. Peters, *Arch. Biochem. Biophys.* **2006**, *448*, 133–140.
69. K. Shimura; A. Okada; K. Okada; Y. Jikumaru; K.-W. Ko; T. Toyomasu; T. Sassa; M. Hasegawa; O. Kodama; N. Shibuya; J. Koga; H. Nojiri; H. Yamane, *J. Biol. Chem.* **2007**, *282*, 34013–34018.
70. H. Oikawa; T. Toyomasu; H. Toshima; S. Ohashi; H. Kawaide; Y. Kamiya; M. Ohtsuka; S. Shinoda; W. Mitsuhashi; T. Sassa, *J. Am. Chem. Soc.* **2001**, *123*, 5154–5155.
71. H. Oikawa; K. Nakamura; H. Toshima; T. Toyomasu; T. Sassa, *J. Am. Chem. Soc.* **2002**, *124*, 9145–9153.
72. T. Toyomasu; K. Nakaminami; H. Toshima; T. Mie; K. Watanabe; H. Ito; H. Matsui; W. Mitsuhashi; T. Sassa; H. Oikawa, *Biosci. Biotechnol. Biochem.* **2004**, *68*, 146–152.
73. T. Dairi; Y. Hamano; T. Kuzuyama; N. Itoh; K. Furihata; H. Seto, *J. Bacteriol.* **2001**, *183*, 6085–6094.
74. Y. Hamano; T. Kuzuyama; N. Itoh; K. Furihata; H. Seto; T. Dairi, *J. Biol. Chem.* **2002**, *277* (40), 37098–37104.
75. C. Nakano; T. Okamura; T. Sato; T. Dairi; T. Hoshino, *Chem. Commun.*1016–1018.
76. R. Croteau; R. E. B. Ketchum; R. M. Long; R. Kaspera; M. R. Wildung, *Phytochem. Rev.* **2006**, *5*, 75–97.
77. W. Wang; Q. Shi; T. Quyang; P. Zhu; K. Cheng, *Acta Bot. Sin.* **2002**, *44*, 181–187.
78. G. Kai; L. Zhao; L. Zhang; Z. Li; B. Guo; D. Zhao; X. Sun; Z. Miao; L. Tang, *J. Biochem. Mol. Biol.* **2005**, *38*, 668–675.
79. P. Gutta; D. J. Jantillo, *Org. Lett.* **2007**, *9*, 1069–1071.
80. T. Tokiwano; T. Endo; T. Tsukagoshi; H. Goto; E. Fukushi; H. Oikawa, *Org. Biomol. Chem.* **2005**, *3*, 2713–2722.

81. S. Y. Chow; H. J. Williams; J. D. Pennington; S. Nanda; J. H. Reibenspies; A. I. Scott, *Tetrahedron* **2007**, *63*, 6204–6209.
82. J. Hefner; S. M. Rebenstein; R. E. B. Ketchum; D. M. Gibson; R. M. Williams; R. Croteau, *Chem. Biol.* **1996**, *3*, 479–489.
83. K. Walker; A. Schoendorf; R. Croteau, *Arch. Biochem. Biophys.* **2000**, *374*, 371–380.
84. A. Schoendorf; C. D. Rithner; R. M. Williams; R. B. Croteau, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 1501–1506.
85. S. Jennewein; C. D. Rithner; R. M. Williams; R. B. Croteau, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13595–13600.
86. E. Wang; G. J. Wagner, *Planta* **2003**, *216*, 686–691.
87. M. Würtele; C. Jelich-Ottmann; A. Wittinghofer; C. Oecking, *EMBO J.* **2003**, *22*, 987–994.
88. Y. Honma; M. Akimoto, *Cancer Sci.* **2007**, *98*, 1643–1816.
89. N. Kato; C.-S. Zhang; T. Matsui; H. Iwabuchi; A. Mori; A. Ballio; T. Sassa, *J. Chem. Soc., Perkin Trans. 1* **1998**, 2473–2474.
90. N. Kato; C.-S. Zhang; N. Tajima; A. Mori; A. Graniti; T. Sassa, *Chem. Commun.* **1999**, 367–368.
91. T. Sassa; C.-S. Zhang; M. Sato; N. Tajima; N. Kato; A. Mori, *Tetrahedron Lett.* **2000**, *41*, 2401–2404.
92. T. Toyomasu; M. Tsukahara; A. Kaneko; R. Niida; W. Mitsuhashi; T. Dairi; N. Kato; T. Sassa, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 3084–3088.
93. T. Toyomasu; M. Tsukahara; H. Kenmoku; M. Anada; H. Nitta; W. Mitsuhashi; T. Sassa; N. Kato, *Org. Lett.* **2009**, *11*, 3044–3047.
94. H. Kenmoku; N. Kato; M. Shimada; M. Omoto; A. Mori; W. Mitsuhashi; T. Sassa, *Tetrahedron Lett.* **2001**, *42*, 7439–7442.
95. R. G. Kerr; A. C. Kohl; T. A. Ferns, *J. Ind. Microbiol. Biotechnol.* **2006**, *33*, 532–538.

Biographical Sketches

Tomonobu Toyomasu is an Associate Professor at Yamagata University in Japan. He was born in 1966. He received his B.S. (1989), M.S. (1991), and Ph.D. (1994) from the University of Tokyo. His Ph.D. thesis titled 'Studies on Mechanism of Germination of Photoblastic Lettuce Seed' investigated the regulation by light of endogenous levels of gibberellins and abscisic acid in lettuce seeds. He did postdoctoral research at RIKEN Frontier Research Program in 1995, and studied gibberellin biosynthetic enzyme genes in rice. He was appointed Assistant Professor at Yamagata University in 1995, and promoted to Associate Professor in 1998. He was involved in three main projects at Yamagata University: (1) regulation by light of metabolism of gibberellins and abscisic acid in photoblastic lettuce seeds; (2) biosynthesis of diterpene phytoalexins in rice; and (3) isolation and characterization of diterpene cyclase genes in fungi. He received The Japanese Society of Chemical Regulation of Plants Award for the Encouragement of Young Scientists in 2004 (project 1), and The Japan Bioscience, Biotechnology, and Agrochemistry Society Award for the Encouragement of Young Scientists in 2007 (projects 2 and 3).

Takeshi Sassa is a Professor Emeritus at the Faculty of Agriculture, Yamagata University. He received his Ph.D. (1969) in agricultural organic chemistry from Nagoya University. He spent 3 years (1967–69) at Nagoya University, working on microbial plant growth regulators, and moved as an Associate Professor to Faculty of Agriculture, Yamagata University in 1970. He worked as a Visiting Researcher with Professor R. P. Pharis at Calgary University in 1975. He received an award for the encouragement of young scientists for studies on plant growth substances produced by microorganisms from The Japan Society of Bioscience, Biotechnology, and Agrochemistry in 1975 and The Japanese Chemical Regulation of Plants Society Award for chemical studies on plant growth-regulating substances produced by true fungi in 1993. In 1977, he was promoted to Professor at the same Faculty. He retired in 2005, and then spent 2 years as a Visiting Professor at the same Faculty to continue the study on the biosynthesis including cyclases of bioactive diterpenoids. His present research focuses on the creation of a novel differentiation-inducing anticancer drug by chemical modification of fungal diterpene glycosides, cotylenin and fusicoccin.

1.18 Triterpenes

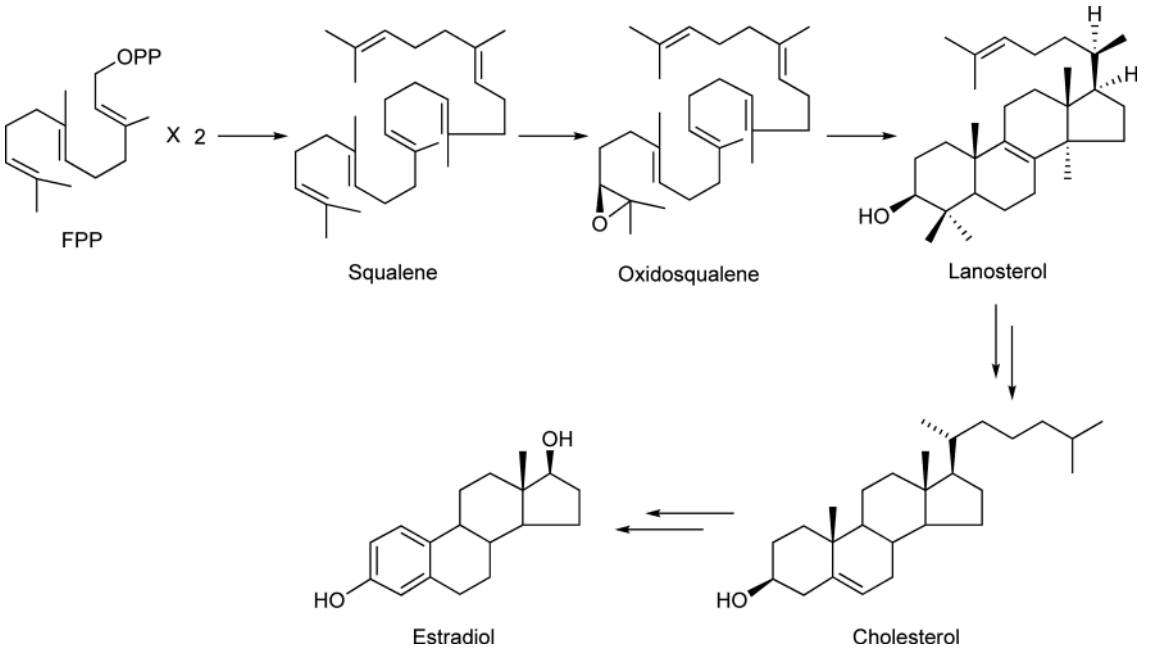
Tetsuo Kushiro and Yutaka Ebizuka, The University of Tokyo, Tokyo, Japan

© 2010 Elsevier Ltd. All rights reserved.

1.18.1	**Introduction**	673
1.18.2	**Cyclization Mechanism**	674
1.18.3	**Lanosterol/Cycloartenol Synthases**	676
1.18.3.1	Mechanism	676
1.18.3.2	Genes	679
1.18.3.3	Mutational Studies	680
1.18.3.4	Structure of Human Lanosterol Synthase	681
1.18.4	**Plant Triterpene Synthases**	684
1.18.4.1	β-Amyrin and Lupeol Synthases	684
1.18.4.1.1	Mechanism	684
1.18.4.1.2	Genes	685
1.18.4.1.3	Product specificity	687
1.18.4.2	Multifunctional Triterpene Synthase	688
1.18.4.3	*Arabidopsis* Triterpene Synthases	691
1.18.4.4	Other Triterpene Synthases	695
1.18.4.4.1	Isomultiflorenol synthase	695
1.18.4.4.2	Dammarenediol synthase	696
1.18.4.4.3	Cucurbitadienol synthase	696
1.18.4.4.4	Baccharis oxide synthase	699
1.18.4.4.5	Protostadienol synthase	700
1.18.4.4.6	Squalene cyclases from ferns	700
1.18.5	**Triterpene Tailoring Steps**	701
1.18.6	**Summary and Future Perspectives**	704
References		705

1.18.1 Introduction

Triterpenes are members of isoprenoids that are derived from a C30 precursor, squalene. Squalene itself is derived from two molecules of farnesyl diphosphate (FPP), which is the precursor of sesquiterpenes (Chapter 1.16). The unique tail-to-tail condensation of FPP is catalyzed by an enzyme squalene synthase.[1] Because of this symmetrical nature, squalene lacks diphosphate group (–OPP), and, therefore, its cyclization exclusively involves protonation (type B cyclization) to generate a carbocation, rather than an ionization, of the OPP group seen mainly in mono-, sesqui-, and diterpene biosynthesis. The majority of triterpenes found in nature are cyclic triterpenes with 1–5 ring systems. In some bacteria and protozoans, squalene is directly cyclized into pentacyclic triterpenes such as hopene by squalene:hopene cyclase (SHC). Details of this enzyme are described in Chapter 1.19. On the other hand, in animals, plants, and fungi, (3*S*)-2,3-oxidosqualene is cyclized into cyclic triterpenes by the enzyme oxidosqualene cyclase (OSC). An additional enzyme known as squalene epoxidase is required by these organisms to convert squalene into oxidosqualene, which involves a stereospecific epoxidation of the terminal olefin.[2] Because of the presence of an epoxide in oxidosqualene, all of the cyclic triterpenes derived from this precursor possess oxygen functionality at the C-3 position. These cyclic triterpenes are further converted into various metabolites including sterols, steroids, and saponins, which play an important physiological role in the cells (**Scheme 1**). Sterols and steroids are described in Chapter 1.21. Nearly 100 different skeletal types of cyclic triterpenes are known in nature, which underscores the potential ability of an OSC to produce large structural diversity from a single precursor. This chapter describes our current knowledge of this remarkable enzyme and its cyclization reaction mechanisms to generate diverse structures found in the triterpene class (**Scheme 1**).



Scheme 1

1.18.2 Cyclization Mechanism

The cyclization of oxidosqualene had been a subject of intensive research over the past 60 years.[3,4] It was considered to be the most complex organic transformations found in nature, which include formation of a multiring system with numerous stereogenic centers in a single transformation. It is not just remarkable but somewhat astonishing how nature has managed to carry out such a reaction in an aqueous solution that involves reactive carbocationic intermediates. It was, thus, of prime interest to chemists how OSC handles carbocationic intermediates, controls the conformation of the substrate, and guides through the reaction sequence to produce each specific product. The OSC reaction is composed of the same basic reactions found in mono-, sesqui-, and diterpene biosynthesis; however, OSC produces up to five ring systems. The cyclization reaction of OSC can be divided into the following four parts (**Figure 1**).

1. Initiation of a reaction. A proton attack on the epoxide moiety generates an initial tertiary carbocation. The resulting OH group resides on every triterpene produced by OSC except for a few cases.
2. Propagation of a reaction through cation-π cyclization. Once formed, the carbocation undergoes attack by a nearby π electron onto an empty 2p orbital of a carbocationic center to form a C–C σ bond.
3. Further propagation through 1,2-shifts. This process is known as Wagner–Meerwein rearrangements, which include hydride, methyl, and alkyl group migration that takes place suprafacially. Ring expansion is part of this process.
4. Termination of a reaction. Deprotonation from a carbon adjacent to a carbocationic center will generate an olefin. On the other hand, a water attack onto a carbocationic center will produce a hydroxyl group.

In terms of generating triterpene structural diversity, it is important to understand how OSC exerts its huge catalytic potential to create a multitude of diverse structures. At the same time, OSC navigates the cascade of carbocation-mediated cyclization and rearrangements into particular reaction pathway to produce each specific product(s).

The catalytic potential of OSC arises primarily from the active site cavity, which is highly hydrophobic in nature and common to all of the terpene cyclase class of enzymes. Numerous hydrophobic amino acid residues compose the cavity, which binds linear isoprenoid substrates and forces them to fold into a reactive

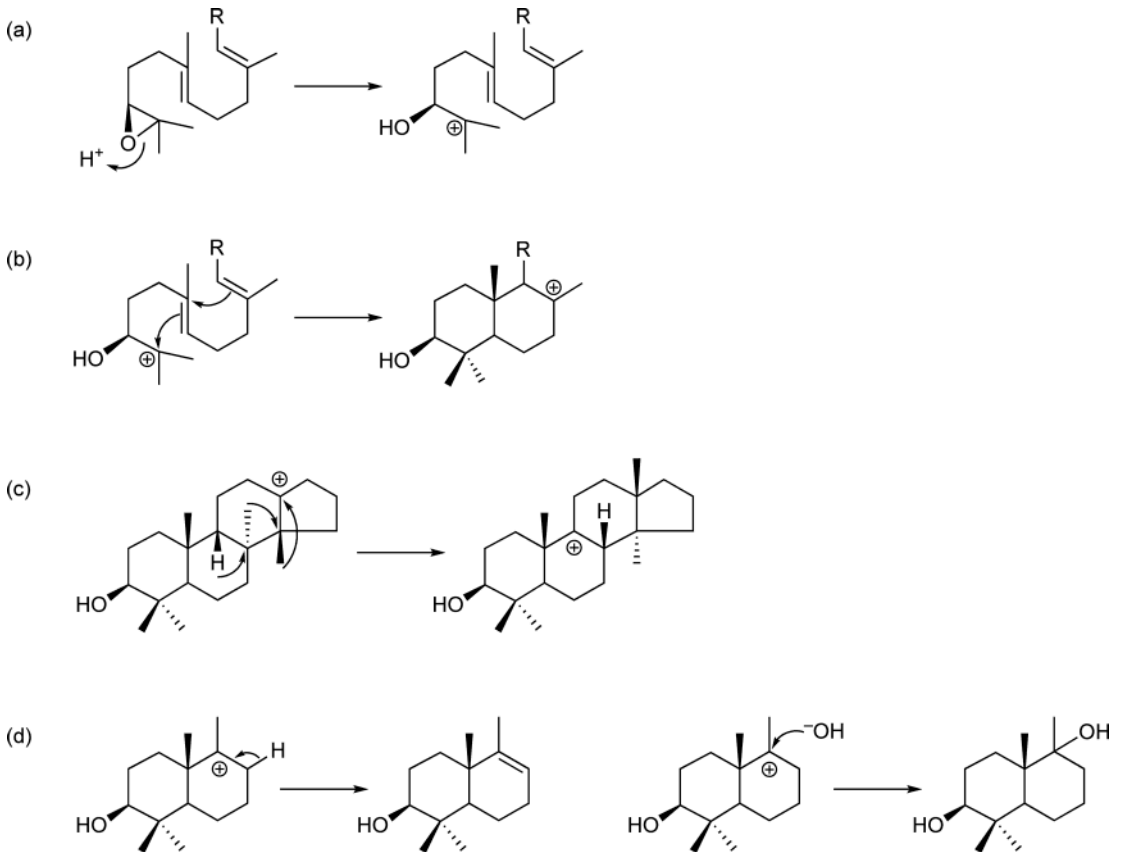


Figure 1 Four basic reactions found in OSC-catalyzed cyclization reaction.

conformation ready for cyclization cascade.[5] Steric effects are mainly responsible for the substrate folding whereas electronic effects are thought to be important for the stabilization of transition state carbocationic intermediates lowering the activation barrier to facilitate cation-π cyclization. However, calculations have suggested a rather small activation barrier for the cation-π cyclization, and there may be only a minimal assistance from the enzyme.[6] The hydrophobic cavity is also responsible for excluding solvent water molecules that may interfere with an intermediate carbocation to aberrantly quench the reaction. A strong catalytic acid is also necessary to generate a carbocation for initiation of the reaction as well as a catalytic base for terminating the reaction. The OSC reaction is highly exothermic (some 60 kcal mol^{-1} is estimated to be released during pentacyclic formation) and requires the enzyme to be rigid enough to support the protein structure during this process.

In addition, each reaction pathway leading to a particular product is determined by the number of cyclization, which dictates the number of rings formed, and by the extent to which hydride and methyl migration takes place. The number of cyclization is mainly controlled by the conformation of the substrate, that is, whether the neighboring olefin is ideally positioned for the attack of a carbocation for the next round of cyclization. The shape of the active site cavity may primarily be responsible for such a control, and, therefore, OSCs producing tetracyclic and pentacyclic structures would have different cavity shapes. In addition, stabilization of a particular carbocation may determine the lifetime of a carbocation; whether a competing reaction exists or not may also contribute to the reaction specificity. In fact, it has been considered that not only a thermodynamic control but also a kinetic control contributes to the observed product specificity.[6] That is, the reaction pathway is not determined by the relative stability of intermediate carbocationic species alone.

On the other hand, the extent of hydride and methyl migration is believed to be largely determined by the position of deprotonation. Because Wagner–Meerwein shifts can readily occur chemically in solution, there

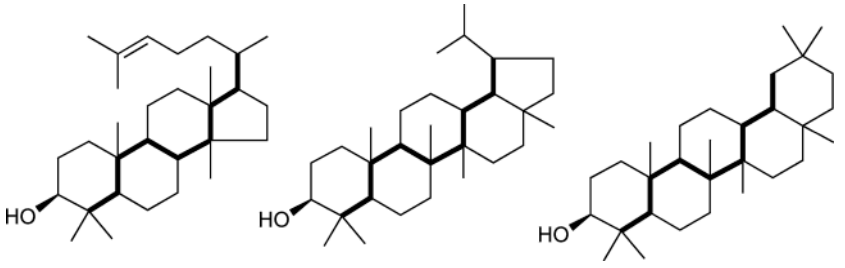


Figure 2 The path by which backbone rearrangements take place.

may be only minimal assistance from the enzyme during this process. What perhaps determines the extent of rearrangements is the position of the base (either a basic residue or a solvent water molecule) in the active site. Currently, no study has successfully identified a catalytic base among all the terpene cyclase families. The path by which these rearrangements take place on the triterpene skeleton is called *backbone rearrangement*. As depicted in **Figure 2**, in each skeletal type, rearrangements take place along the carbon atoms shown in bold lines, which gives only the tertiary carbocation during this process. Most of the products originating from each of these carbocationic centers have been identified in nature.[7]

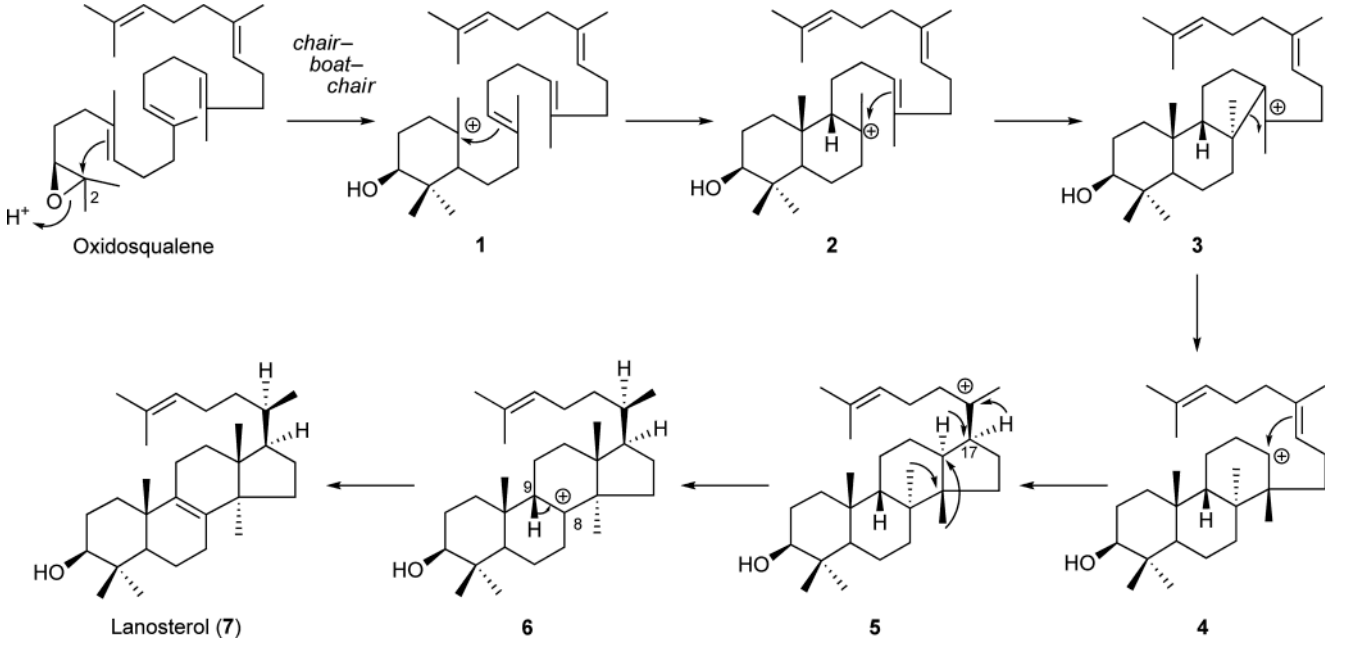
1.18.3 Lanosterol/Cycloartenol Synthases

1.18.3.1 Mechanism

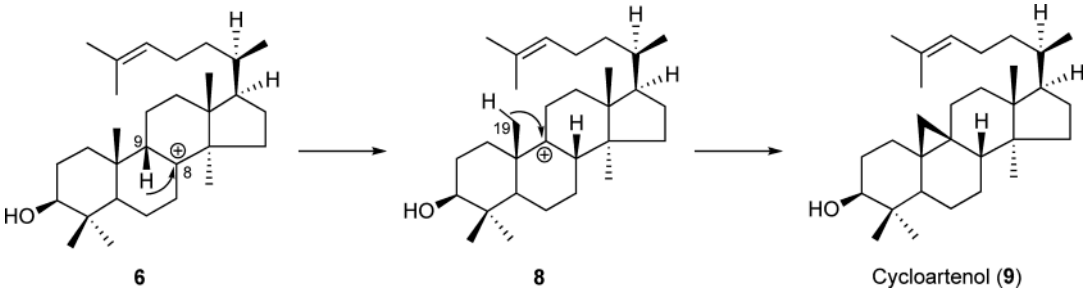
Lanosterol serves as a precursor of sterols such as cholesterol in mammals and ergosterol in yeast and fungi, whereas cycloartenol serves as a precursor of phytosterols in plants. Sterols are essential components of the cells; therefore, lanosterol synthase and cycloartenol synthase genes are essential genes in these organisms. In humans, cholesterol plays a central role in lipid physiology and is a precursor of various steroid hormones. Thus, lanosterol synthase is an ideal target for the development of clinical drugs that control cellular cholesterol levels.

Cyclization into lanosterol and cycloartenol is the most well-studied reaction among OSCs. The biosynthesis of lanosterol and cycloartenol involves oxidosqualene folded in a *chair–boat–chair* conformation in the active site. Protonation of the epoxide generates tertiary carbocation at C-2, which is then cyclized by the attack of a neighboring olefin to generate first a monocyclic carbocation **1** and then a bicyclic carbocation **2** (**Scheme 2**). Protonation triggered epoxide opening, and the first ring cyclization has been proposed to take place in a concerted manner.[8] The third ring cyclization initially forms a 6-6-5 tricyclic tertiary carbocation intermediate **3**, which then undergoes a ring expansion into six-membered C ring to form **4**. This is then followed by the fourth cyclization to give a tetracyclic intermediate, protosteryl cation **5** with a 17β side chain. From this stage, a series of Wagner–Meerwein 1,2-shifts of hydrides and methyl groups take place to give C-8 carbocation **6**, which then undergoes deprotonation from C-9 to give lanosterol (**7**) (**Scheme 2**). On the other hand, additional hydride shift from C-9 to C-8 in **6** will give C-9 carbocation **8**, which then undergoes a deprotonation from C-19 methyl group to form a cyclopropane ring of cycloartenol (**9**) (**Scheme 3**).

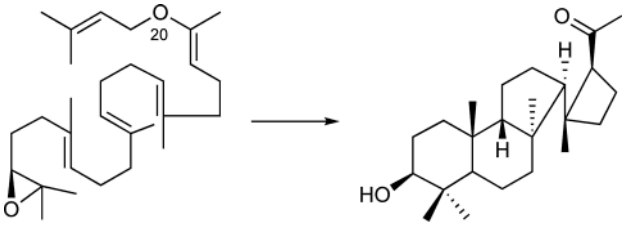
Initially, all of these cyclization processes were proposed to take place in a concerted manner; however, it is now generally accepted that cyclization is a stepwise process involving each discrete carbocationic intermediate. The *anti*-Markovnikov third ring cyclization was controversial for a long time because it involves an unstable secondary carbocation **4**. The existence of 6-6-5 fused intermediate **3** was suggested from the experiment using 20-oxa-2,3-oxidosqualene (**10**) as a substrate, which generated an unusual product **11** (**Scheme 4**).[9] Involvement of a ring expansion process at this stage is critical for proper formation of the 6-6-6-5 fused ring system seen in sterols. This problem arose primarily from the (oxido)squalene molecule which adopts a symmetrical structure due to tail-to-tail condensation of FPP. If it were a regular polyprenol formed through a head-to-tail condensation of isoprene units, a stable tertiary carbocation would be formed at the


chair–
boat–
chair
H+
O
2
HO
H
17
9
8
Oxidosqualene
1
2
3
4
5
6
Lanosterol (7)

Scheme 2



Scheme 3



Scheme 4

tricyclic stage as exemplified in **Figure 3**. In fact, any synthetic efforts to construct a polycyclic structure from oxidosqualene using Lewis acid generate only the 6-6-5 fused tricyclic system and nothing beyond.

The stereochemistry at C-17 during protosteryl cation **5** was also a controversy in the past, as it was originally proposed to have a 17α side chain (**Figure 4(a)**); it had been proposed that the following Wagner–Meerwein hydride shift takes place in a concerted manner with strict antistereochemistry.[3] This would then require ~120° rotation of a C–C bond between C-17 and C-20 prior to hydride shift from C-17 in order to install the correct C-20*R* stereochemistry. However, incubation of **10** with yeast microsomal preparation gave a tetracyclic methyl ketone **12** with 17β configuration (**Scheme 5**).[10] This implied that **5** already had 17β side chain and oxidosqualene is folded in a conformation as shown in **Figure 4(b)**. This would require only minimum rotation of ~60° to achieve 20*R* stereochemistry prior to the hydride shift from C-17. The strict antistereochemistry during the Wagner–Meerwein shift is not always necessary.

It has been documented in the past that methyl groups of oxidosqualene play an important role in the correct folding of the substrate in the active site of the enzyme. For example, 10,15-desmethyl analogue (**13**) was shown to cyclize into an unusual product **14**, which involves a *chair–chair–boat* conformation that resulted in aberrant cyclization (**Scheme 6**).[11] Thus, methyl groups at C-10 and C-15 are important for forcing B ring into a *boat* conformation.

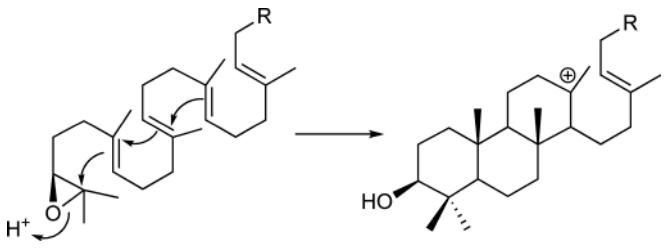
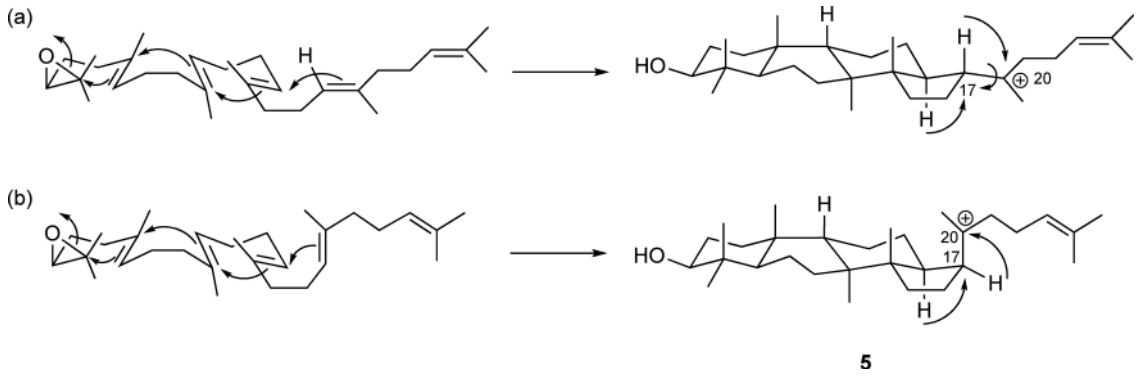
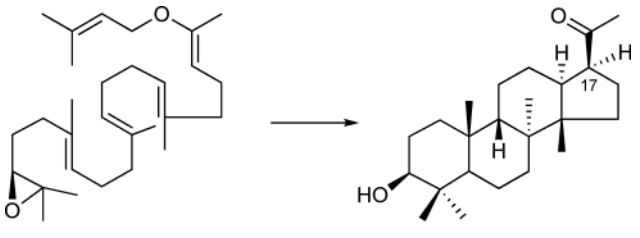


Figure 3 Hypothetical cyclization of polyprenol at tricyclic stage.

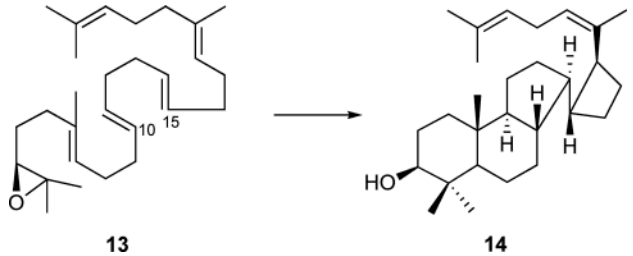


5

Figure 4 Conformation of oxidosqualene during *chair–boat–chair* cyclization to produce protosteryl cation (**5**). (a) Previously postulated conformation. (b) Currently proposed conformation.



Scheme 5



Scheme 6

1.18.3.2 Genes

The lanosterol synthase gene was first cloned from yeast *Candida albicans* (ERG7)[12–14] and later from *Saccharomyces cerevisiae* (ERG7),[15,16] *Schizosaccharomyces pombe*,[17] rat,[18,19] and human.[20,21] These genes code for a protein of ~730 amino acid residues with a molecular weight of ~83 kDa. Several repetitive motifs were identified and were termed QW motifs ($(K/R)(G/A)X_{2-3}(F/Y/W)(L/I/V)X_3QX_{2-5}GXW$).[22] The QW motif was also found in related bacterial SHC. It was initially thought that aromatic amino acids such as Trp in this motif might stabilize the intermediate carbocation during the cyclization reaction;[23] however, later this was proven to be wrong as these motifs were located on the surface of OSC to stabilize the protein structure.[24] Lanosterol synthase genes were further cloned from *Trypanosoma brucei*,[25] *T. cruzi*,[26] *Cephalosporium caerulens*,[27] *Pneumocystis carinii*,[28] and plants including *Arabidopsis thaliana*,[29,30] *Panax ginseng*,[30] and *Lotus japonicus*[31] (see below).

The cycloartenol synthase gene was first cloned from *A. thaliana* (AtCAS1) and found to code for a similar protein of 759 amino acid residues.[32] Sequence comparison of lanosterol and cycloartenol synthases showed 30–40% sequence identities and revealed several conserved sequences including those of QW motifs. Numerous cycloartenol synthase genes were identified later from *Pisum sativum*,[33] *P. ginseng*,[34] *Luffa cylindrica*,[35]

Dictyostelium discoideum,[36] *Glycyrrhiza glabra*,[37] *Costus speciosus*,[38] *Betula platyphylla*,[39] *Cucurbita pepo*,[40] *Ricinus communis*,[41] *L. japonicus*,[42] *Rhizophora stylosa*,[43] *Kandelia candel*,[43] and a fern, *Adiantum capillus-veneris*.[44]

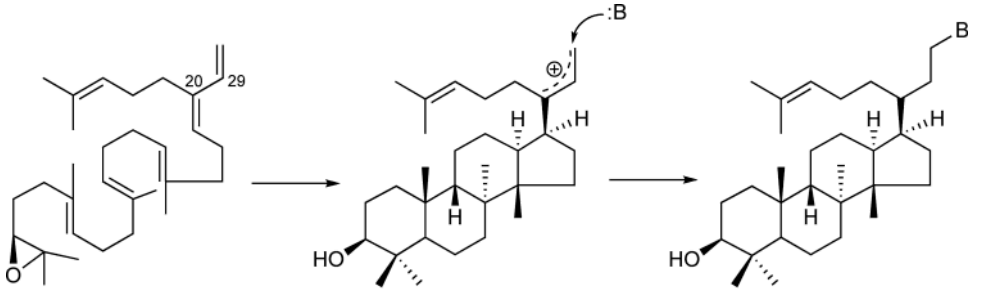
Interestingly, several OSC genes have been cloned from prokaryotes as well, including lanosterol synthase from *Methylococcus capsulatus*[45,46] and cycloartenol synthase from *Stigmatella aurantiaca*.[47] Sterols are normally absent in prokaryotes, and the existence of an OSC gene in these organisms raises an intriguing question regarding the evolution of an OSC gene and a sterol biosynthesis.

1.18.3.3 Mutational Studies

Earlier on, an effort to identify the active site residues was carried out using a mechanism-based irreversible inhibitor 29-methylidene-2,3-oxidosqualene (29-MOS) (**15**). This inhibitor was designed to trap the protosteryl cation through the allylic cation by introducing a vinyl group at C-20, which may then be covalently linked to a nearby nucleophilic residue in the active site (**Scheme 7**).[48] This study identified a so-called DCTAE motif that is strictly conserved among OSCs. The inhibitor was found to be attached to Asp residue of this motif, suggesting that the negative charge of Asp may stabilize the positive charge of the protosteryl cation intermediate. Later on, mutation of this Asp into the Asn of *S. cerevisiae* lanosterol synthase was found to abolish lanosterol synthase activity.[49] It was postulated that some acid residue such as Asp or Glu should be responsible for attacking an epoxide to initiate the cyclization reaction. In fact, mutation of all Asp and Glu residues that are conserved among five different lanosterol synthases from *C. albicans*, *S. cerevisiae*, *S. pombe*, rat, and human (six Asp residues and nine Glu residues in total) has shown that only the abovementioned Asp residue (Asp456 in *S. cerevisiae* ERG7 gene) was critical for lanosterol production.[49] Based on these results, Asp456 was proposed to be involved in the protonation of the epoxide rather than in stabilizing an intermediate carbocation.

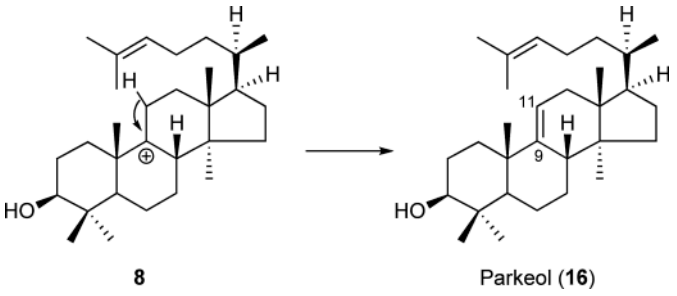
Further information regarding the active site residues came from a random mutagenesis study. Mutant cycloartenol synthase (AtCAS1) clones that allowed yeast lanosterol synthase mutant (ergosterol auxotroph) to grow in the absence of exogenous ergosterol supply were selected. These clones were expected to produce lanosterol. One such clone was identified to have I481V mutation (AtCAS1 numbering).[50] This residue is located two residues upstream of the DCTAE motif. Careful analysis of the mutant showed the production of cycloartenol (**9**), lanosterol (**7**), and parkeol (**16**) to be in a 55:24:21 ratio. Parkeol (**16**) arose from C-9 carbocation **8** through deprotonation from C-11 (**Scheme 8**). Therefore, the residue at this position was proposed to control the final deprotonation step during C-8 or C-9 carbocation intermediate. In fact, all cycloartenol synthases identified so far possess Ile at this position. Conversely, a reverse mutation on lanosterol synthase, V454I mutation (*S. cerevisiae* ERG7 numbering), produced only lanosterol.[51] Therefore, Ile at this position in the context of AtCAS1 sequence is important for cycloartenol formation but not in an ERG background. Further substitution at this position by Ala or Gly resulted in monocyclic achilleol A (**17**) and camelliol C (**18**) produced through a monocyclic carbocation **1** (**Scheme 9**),[51,52] indicating that the residue at this position is also important for proper folding and cyclization beyond the monocyclic stage.

Numerous mutagenesis studies were further carried out, aided by the X-ray crystal structure of human lanosterol synthase, and have been are reviewed extensively.[53–55] These studies revealed several residues important for catalysis and product specificity, including Tyr410, His477, and Tyr532 of *A. thaliana* AtCAS1 and Trp232, His234, Thr384, Phe445, Tyr510, Phe699, and Tyr707 of *S. cerevisiae* ERG7. Most notably, just two

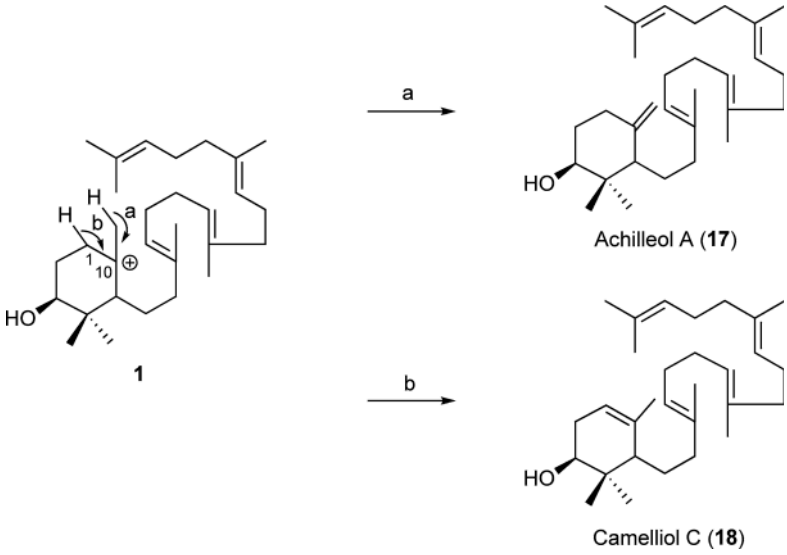


29-MOS (**15**)

Scheme 7



Scheme 8



Scheme 9

mutations on AtCAS1, H477N/I481V, have converted cycloartenol synthase into nearly perfect lanosterol synthase (lanosterol:parkeol = 99:1).[56] According to the homology-modeled structure of AtCAS1, His477 is not located in an active site but in a second sphere located behind the active site residue Tyr410. It is remarkable how a small change in the amino acid residues of OSC can alter the cyclization reaction completely to produce other triterpenes.

1.18.3.4 Structure of Human Lanosterol Synthase

The X-ray crystal structure of human lanosterol synthase complexed with the product lanosterol was solved at 2.1 Å resolution (**Figure 5**).[24] The overall structure consists of two α/α barrel domains connected by several loops. The membrane-binding region is located in domain 2, which renders OSC as a monotopic membrane protein. The active site cavity is located in the center of the protein between the two domains and is connected to the surface through a narrow channel that protrudes into the membrane-binding region. This channel allows the substrate oxidosqualene to enter the hydrophobic active site from the membrane. Overall, the structure of OSC resembles that of SHC, despite showing only ~20% sequence identity. OSC has a smaller active site cavity than SHC, which may be important for controlling the substrate in the unstable B-ring *boat* conformation as opposed to *chair* in SHC, which requires minimal assistance from the enzyme.

The active site cavity is mainly hydrophobic, consisting of aromatic residues and a polar cap by Asp455 at the bottom of the cavity (**Figure 6**). Protonation of the substrate is mediated by this Asp residue, which is

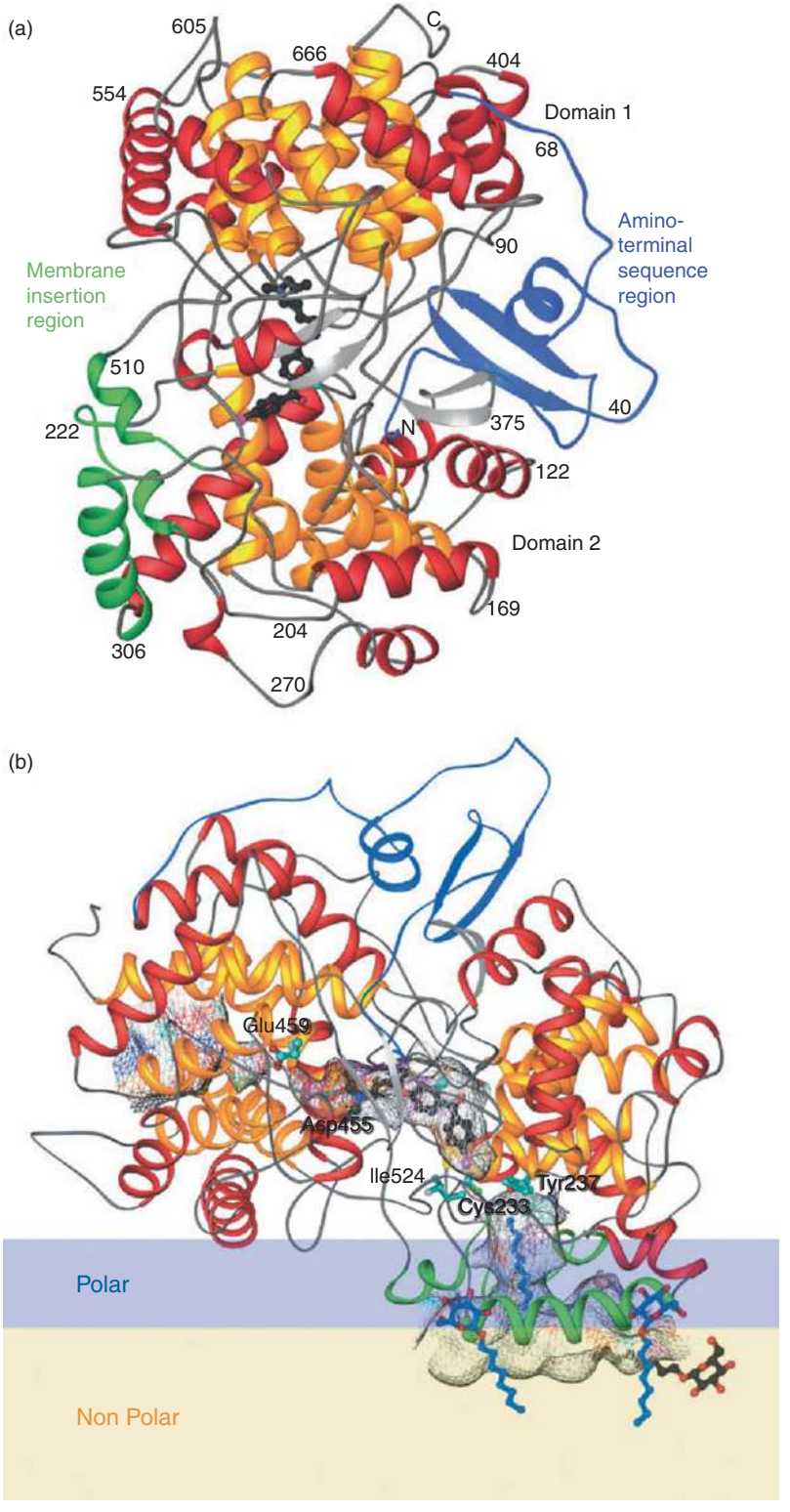


Figure 5 Crystal structure of human lanosterol synthase. (a) Ribbon diagram of the overall structure. The bound inhibitor molecule (Ro 48-8071) is shown in black and indicates the position of the active site. (b) The structure modeled within the membrane lipid bilayer. Detergent molecules found in the crystal are in blue.

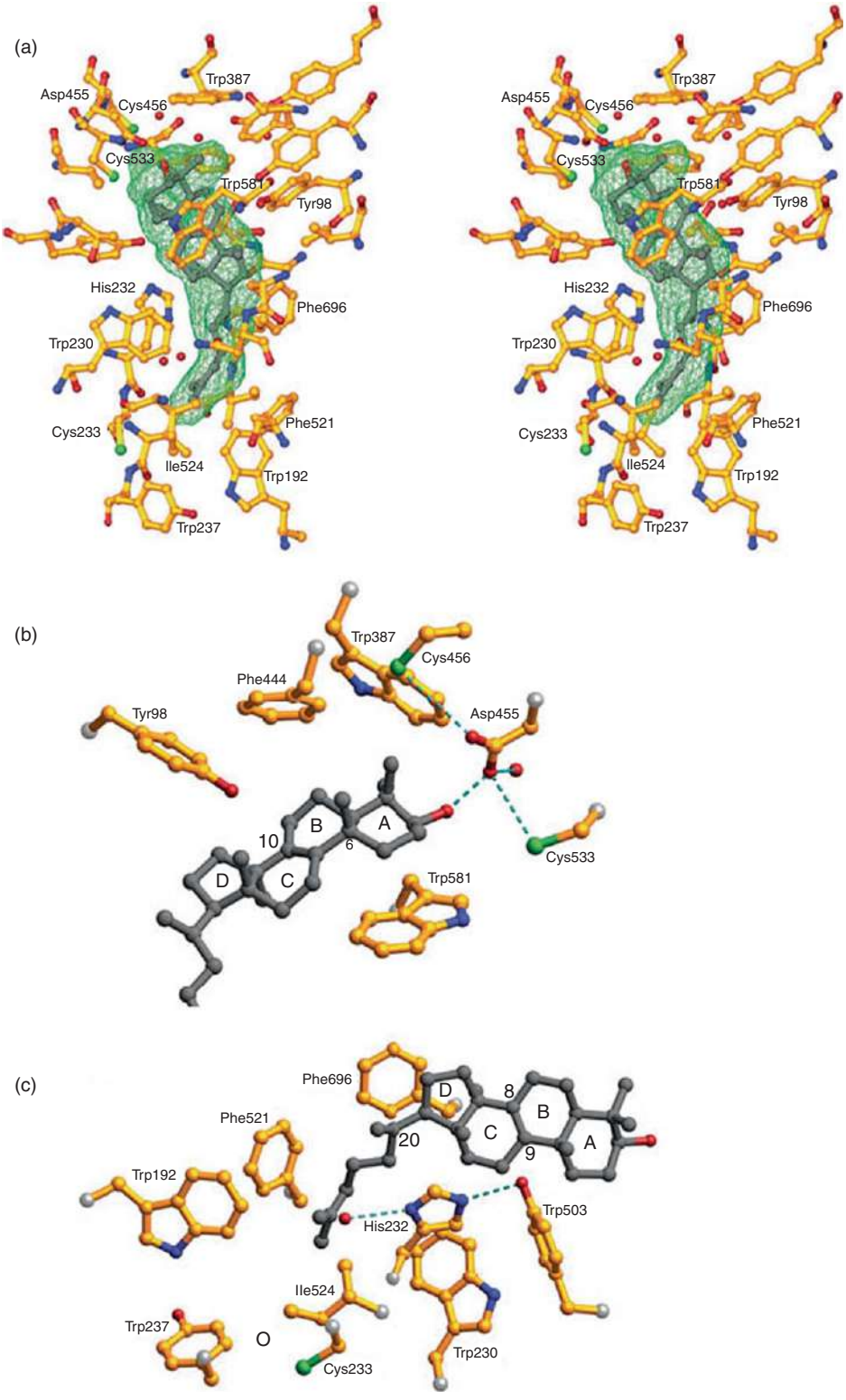


Figure 6 Stereoview of the active site cavity with bound lanosterol molecule. (a) Overall view of the active site. (b) The catalytic acid Asp455 with two Cys residues and aromatic residues that are postulated to be involved in the formation of A and B rings is shown. (c) Residues that are postulated to be involved in the formation of C and D rings are shown including His232 and Tyr503 hydrogen bond dyad, which plays a critical role in the final deprotonation process.

hydrogen bonded to nearby Cys residues that enhance the acidity of the Asp. The intermediate carbocations are stabilized through cation-π interaction by the aromatic residues that are suitably positioned to stabilize each different carbocationic intermediate. These hydrophobic residues also force the linear substrate to fold into a reactive conformation to allow cation-π cyclization to take place. The structure together with mutational studies has postulated the following picture for lanosterol formation.

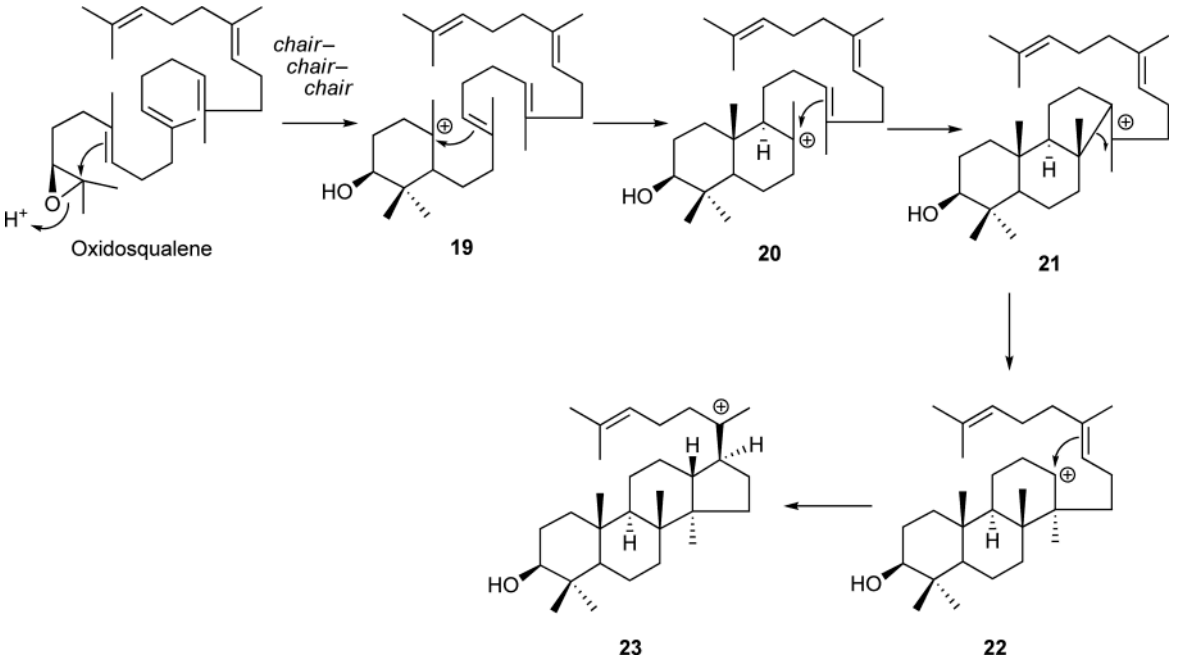
The Asp455 residue assisted by two Cys residues (Cys456 and Cys533) is responsible for the initial proton attack on the epoxide to generate a carbocation. After the first ring formation, the monocyclic C-10 carbocation is stabilized by the Tyr503 residue. The second bicyclic C-8 carbocation is then stabilized by the Tyr704 residue, followed by the third 6-6-5 tricyclic C-14 carbocation by Tyr98 and His232 residues. The Tyr98 residue was also postulated to force the substrate into B-ring *boat* conformation. The fourth 6-6-6-5 protosteryl C-20 carbocation is then stabilized by His232 and Phe696 residues. The same Phe696 residue also stabilizes the first hydride rearranged C-17 carbocation as well. The Trp230 residue in the vicinity of His232 plays an important role during the hydride and methyl group shifts up to C-8 carbocation. Finally, both His232 and Tyr503 residues, which form a hydrogen bond dyad, play a critical role in the final deprotonation process. Mutations in these residues result in aberrant termination of the reaction to produce truncated products such as mono-, bi-, and tricyclic products as well as alternately deprotonated products.

1.18.4 Plant Triterpene Synthases

1.18.4.1 β-Amyrin and Lupeol Synthases

1.18.4.1.1 Mechanism

The majority of triterpenes found in nature are produced by higher plants. Over 100 different skeletal types of triterpenes are known.[7] These triterpenes can be classified according to the cyclization mechanism described below. In contrast to lanosterol and cycloartenol, most plant triterpenes are formed through oxidosqualene folded in a *chair–chair–chair* conformation in the active site of OSC. Similar to lanosterol and cycloartenol described above, cyclization initiated by a protonation of the epoxide generates carbocation intermediates **19–22** that result in tetracyclic dammarenyl cation **23** (**Scheme 10**). Comparing the structure of **23** with that of



Scheme 10

the protosteryl cation (**5**) points to the differences in the stereochemistry around the C ring, which is due to the difference in B-ring conformation during cyclization (*chair* vs *boat*). Again, a ring expansion process is presumably involved during the third ring cyclization from the initially formed 6-6-5 fused system (**21**) to the 6-6-6 tricyclic intermediate (**22**) that eventually furnishes the 6-6-6-5 fused ring system of **23** (**Scheme 10**). A large number of triterpenes are derived from dammarenyl cation (**23**) such as dammarane, euphane, and tirucallane.

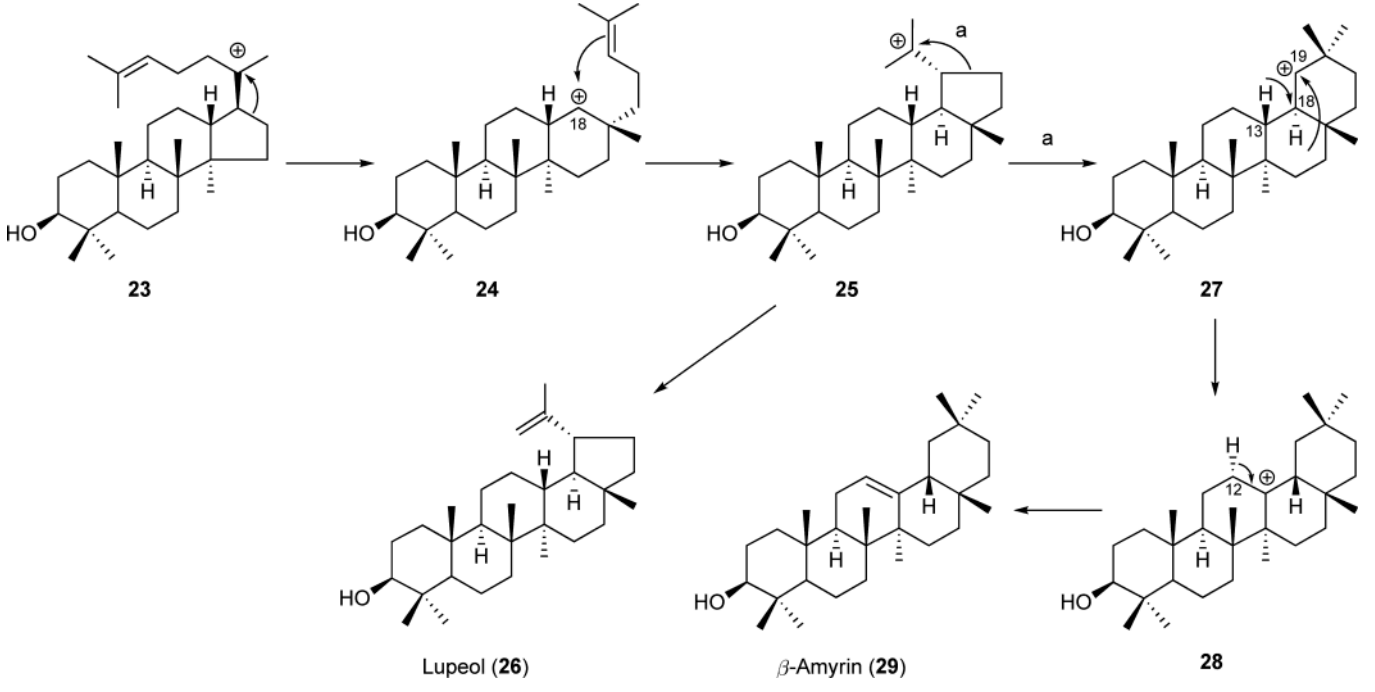
The most abundant triterpene found among higher plants by far is oleanane, which is derived from β-amyrin. β-Amyrin formation involves expansion of the D-ring of the dammarenyl cation from a five- to a six-membered ring, similar to C-ring expansion, which results in secondary C-18 baccharenyl cation (**24**) (**Scheme 11**). This is the basic carbon skeleton of the baccharane-type triterpenes. A subsequent fifth cyclization forms the five-membered E-ring carbocation intermediate known as a lupanyl cation (**25**). All lupane triterpenes originate from this intermediate, such as lupeol a (**26**), produced by deprotonation from one of the methyl groups. Another E-ring expansion would give the 6-6-6-6-6 cation **27** with a secondary C-19 carbocation, which is then followed by two hydride shifts from C-18 to C-19 and C-13 to C-18 to form an oleanyl cation (**28**). Final deprotonation of H-12α will produce β-amyrin (**29**). This transformation furnishes five ring systems and eight stereogenic centers in a single reaction.

1.18.4.1.2 Genes

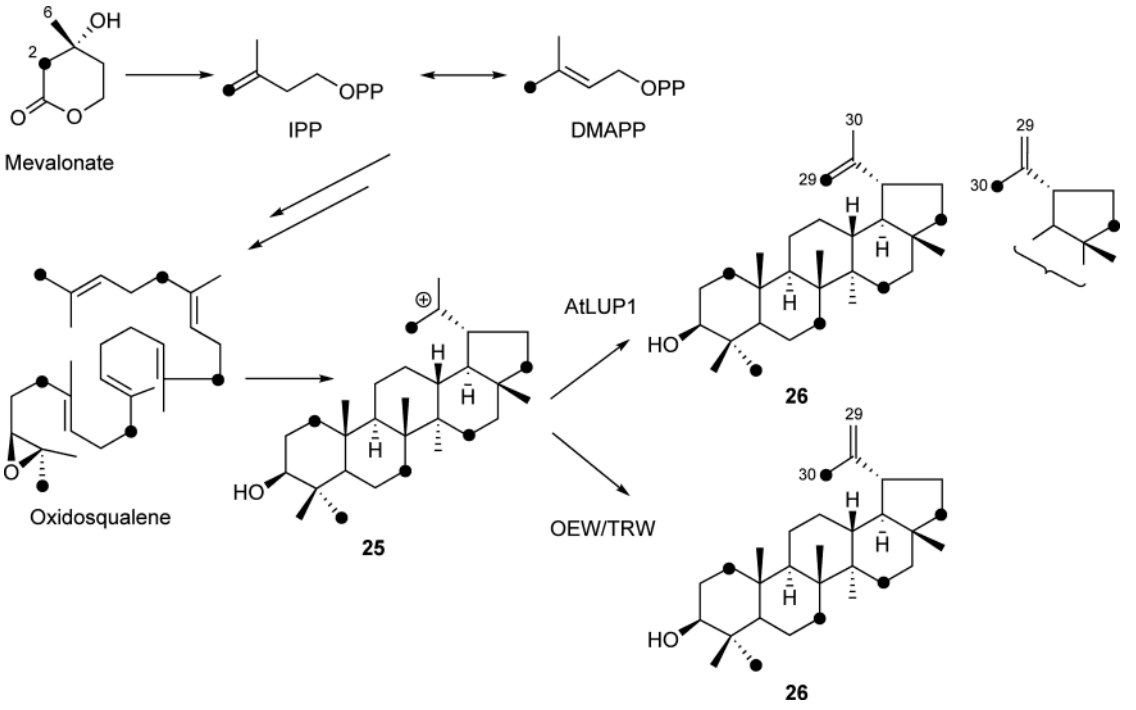
The first triterpene synthase genes identified from plants were lupeol synthase from *A. thaliana* (AtLUP1)[57] and β-amyrin synthase from *P. ginseng* (PNY).[34] Both genes showed 30–40% synthase identities to animal and fungal lanosterol synthases, ~60% to cycloartenol synthase AtCAS1 and 70% to each other. They all contain QW motifs and the DCTAE motifs as well. Therefore, close evolutionary relationships between OSCs involved in triterpene biosynthesis and sterol biosynthesis have been revealed for the first time. The oleanane-type triterpenes that originate from β-amyrin are the most abundant triterpenes found in plants by far. β-Amyrin synthase genes have since been identified from a number of plants including *P. sativum*,[58] *G. glabra*,[59] *Avena strigosa*,[60] *B. platyphylla*,[39] *Medicago truncatula*,[61,62] *Euphorbia tirucalli*,[63] *L. japonicus*,[42] *Saponaria vaccaria*,[64] *Bruguiera gymnorrhiza*,[65] *Artemisia annua*,[66] *Aster sedifolius*,[67] and *A. thaliana*.[68] Except for *A. strigosa* bAS, all these genes share a substantial sequence identity (~80%) with each other and form a distinct clade in the phylogenetic tree. The monocot gene appears to diverge substantially from dicot genes.

AtLUP1 was initially described as lupeol synthase[57] but was later found to produce multiple products, including 3β,20-dihydroxylupane (**30**) (see Section 1.18.4.3).[69] A second type of lupeol synthase genes was cloned from *Olea europaea* (OEW) and *Taraxacum officinale* (TRW).[70] These lupeol synthases produced lupeol as the sole product. A sequence comparison between AtLUP1 showed ~60% amino acid sequence identity. It was later found that the second type of lupeol synthases was more widely found among such plants as including *B. platyphylla*,[39] *G. glabra*,[71] *R. communis*,[41] *L. japonicus*,[42] and *B. gymnorrhiza*.[65] Similar to β-amyrin synthases, these genes also form a distinct clade in the phylogenetic tree, demonstrating that OSCs with the same function exhibit a high sequence identity regardless of plant species.

A distinct stereochemical course of deprotonation from **25** during lupeol formation was identified between AtLUP1 and OEW/TRW.[72,73] Two terminal methyl groups of lupeol have a different biosynthetic origin, as one is derived from C-2 of mevalonate and the other from C-6 of mevalonate. These two methyl groups can be distinguished by feeding [1,2-$^{13}C_2$]acetate. According to the conventional mevalonate pathway, incorporating this doubly labeled acetate would form a triterpene in which carbon signals originating from C-2 of mevalonate appear as a singlet upon ^{13}C-NMR measurements, whereas those originating from C-6 appear with an accompanying doublet due to direct ^{13}C–^{13}C coupling resulting from retention of an intact acetate unit in the mevalonate molecule (**Scheme 12**). The feeding of [1,2-$^{13}C_2$]acetate into an AtLUP1-expressing yeast resulted in lupeol, where both C-29 and C-30 signals were accompanied by a doublet.[72] This indicated that deprotonation took place from both methyl groups. On the other hand, lupeol formed from OEW and TRW showed peaks in which C-29 appeared with an accompanying doublet with C-30 as a singlet.[73] Therefore, deprotonation took place only from the methyl group derived from C-6 of mevalonate, which corresponds to (*Z*)-methyl of oxidosqualene. These results indicated that OEW and TRW have strict control during the deprotonation step, whereas AtLUP1 appears to lack this control. Further study on the stereochemical course of a water addition to lupanyl cation (**25**) during 3β,20-dihydroxylupane (**30**) formation by AtLUP1 was carried out.[74]



Scheme 11

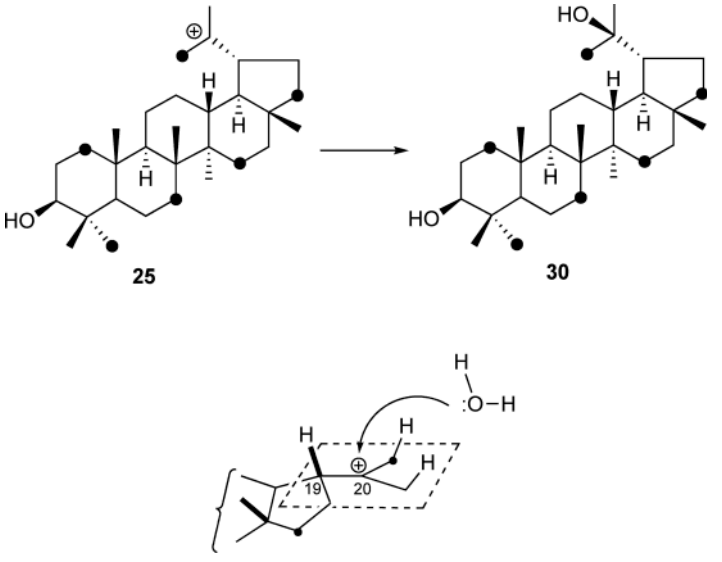


Scheme 12

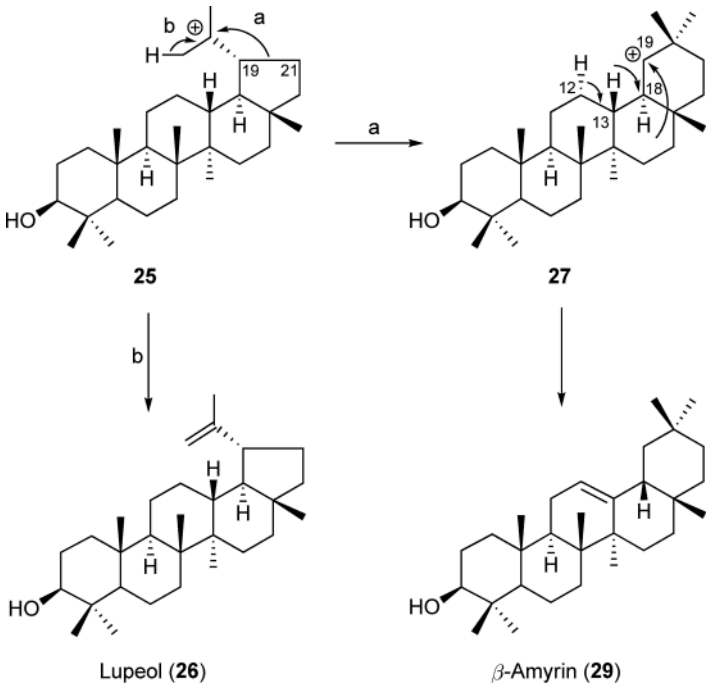
The result showed stereospecific addition of a water molecule onto **25** discriminating the plane of the isopropyl cation moiety (**Scheme 13**). Taken together, AtLUP1 is a multifunctional OSC mainly producing lupeol (**26**) and 3,20-dihydroxylupane (**30**), whereas OEW and TRW are dedicated lupeol synthases.

1.18.4.1.3 Product specificity

By using lupeol synthase AtLUP1 and β-amyrin synthase PNY, amino acid residues that are responsible for controlling lupeol and β-amyrin formation during the cyclization reaction were investigated.[72,75] As mentioned above, formation of both triterpenes takes the same route up to the lupanyl cation stage **25**, where



Scheme 13



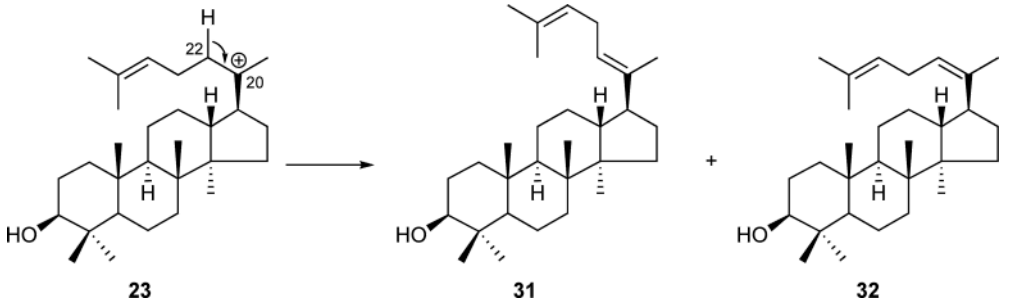
Scheme 14

deprotonation from a methyl group produces lupeol (**26**), whereas the C19–C21 bond migration to form a six-membered E ring, and two consecutive hydride migrations and a deprotonation of H-12α will produce β-amyrin (**29**) (**Scheme 14**). Any residues that control these two reactions should act at the lupanyl cation stage and divert the reaction path into each one.

Various chimeric proteins were made between AtLUP1 and PNY, and they have shown that the region important for β-amyrin formation resides on the 80 amino acid sequence region from Cys260 to Trp340 of PNY in the N-terminal half.[72] Sequence comparison of this region between several lupeol and β-amyrin synthases pointed to a sequence motif ^{258}MWCYC^{262} in PNY (MLCYC in OEW and ILCYS in AtLUP1). A Trp259 in this motif is strictly conserved among β-amyrin synthases, whereas it is conserved as Leu in all lupeol synthases. Mutation of the W259L of the PNY produced lupeol as a major product, whereas mutation of L256W of OEW produced β-amyrin as a major product (75%).[75] It is remarkable that OEW, which solely produces lupeol, already encompassed the potential ability to produce β-amyrin. This was evident from a single mutation that converted it to a nearly accurate β-amyrin synthase. Therefore, this Trp plays a critical role in controlling the reaction at the lupanyl cation stage to divert the reaction into β-amyrin pathway. The π electron of Trp may stabilize the secondary carbocation **27** to favor the E-ring expansion process, although a steric effect also cannot be ruled out. A Tyr261 of PNY in this motif is also highly conserved among lupeol and β-amyrin synthases, whereas it is His in all lanosterol and cycloartenol synthases. Therefore, involvement of this residue in regulating the fifth ring cyclization was suggested. Mutation of Y261H of PNY completely abolished β-amyrin production and produced novel tetracyclic triterpenes **31** and **32**.[75] These are dammarane types produced by a deprotonation from C-22 of dammarenyl cation **23** to form a mixture of E/Z isomer of $\Delta^{20(22)}$ (**Scheme 15**). The His residue, in place of Tyr, which would stabilize the secondary baccharenyl cation, might have served as a base to abstract proton from C-22 at the dammarenyl cation stage. These studies demonstrated that product specificity is controlled by relatively few residues of OSC.

1.18.4.2 Multifunctional Triterpene Synthase

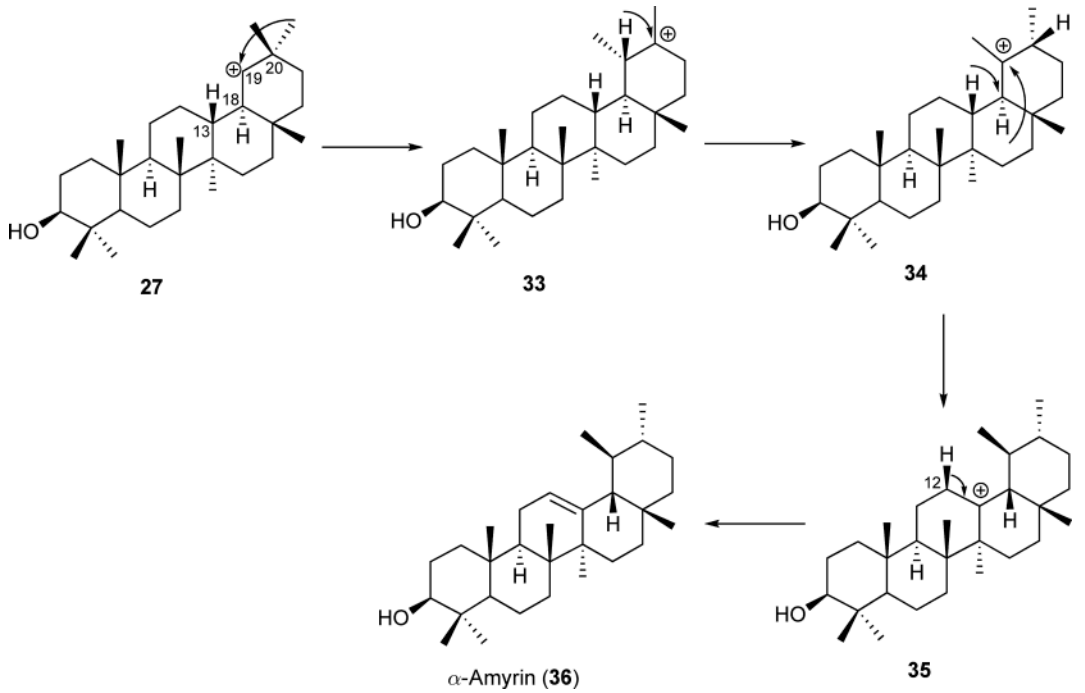
Besides lupeol and β-amyrin synthases, the cloning of OSCs from higher plants has revealed the existence of multifunctional triterpene synthases that produce several products simultaneously. These results argue



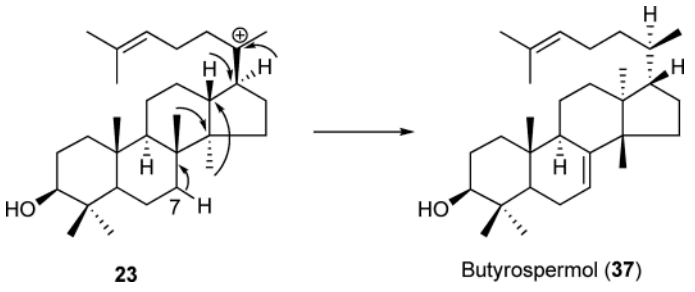
Scheme 15

against the existence of product-specific OSCs for each of the 100 different triterpene skeletons found in nature so far. Therefore, the number of OSC genes is probably much smaller than the actual number of triterpene skeletons found.

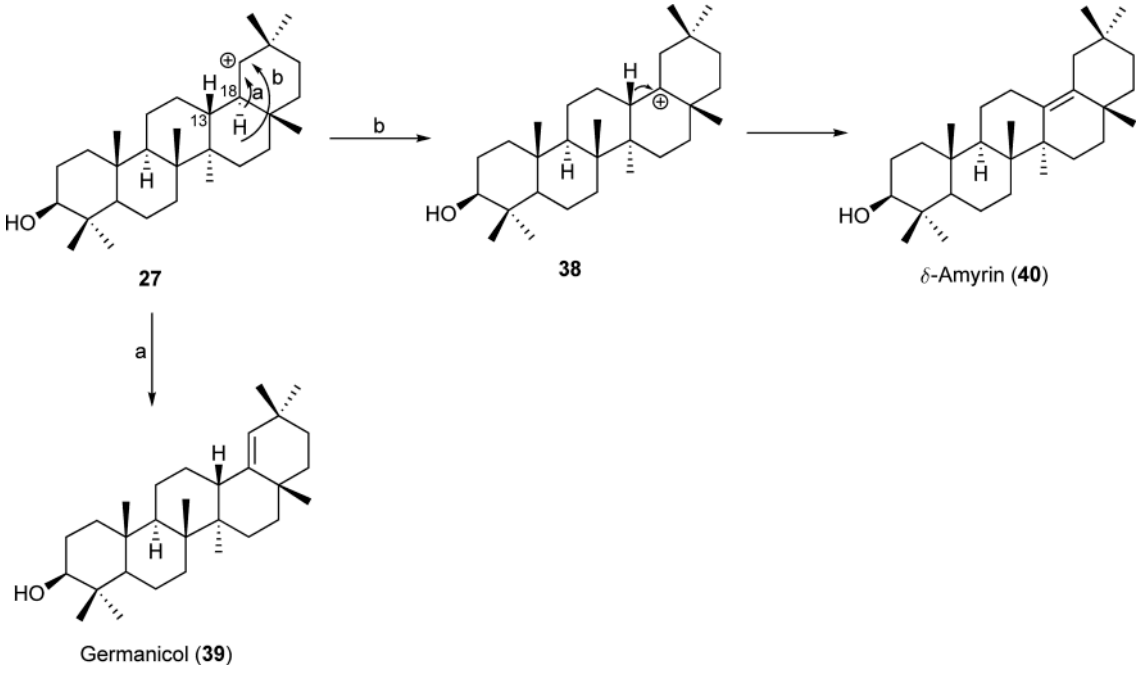
The triterpene synthase that produces α-amyrin, the basic carbon skeleton of ursane-type triterpenes, which are the second major types found in higher plants, was identified from *P. sativum* (PSM).[58] However, it was found to be a multifunctional OSC producing α-amyrin (**36**) (50%) and β-amyrin (**29**) (34%) in nearly equal amounts, together with several minor triterpenes including δ-amyrin (**40**) (8%), Ψ-taraxasterol (**41**) (5.3%), butyrospermol (**37**) (3.4%), lupeol (**26**) (1.3%), germanicol (**39**) (<1%), and taraxasterol (**42**) (<1%). α-amyrin (**36**) formation involves methyl group migration from C-20 during carbocation **27** to result in a taraxasteryl cation (**33**), from which three consecutive hydride shifts from C-19 to C-20, C-18 to C-19, and C-13 to C-18 take place to result in a ursanyl cation (**35**). A final deprotonation of H-12β will produce α-amyrin (**36**) (**Scheme 16**). On the other hand, butyrospermol (**37**) is formed from a dammarenyl cation (**23**) through two hydride and two methyl group shifts and a final deprotonation from C-7 (**Scheme 17**), whereas germanicol (**39**) and δ-amyrin (**40**) are formed from a deprotonation from cation **27** and germanicyl cation (**38**), respectively (**Scheme 18**). Both Ψ-taraxasterol (**41**) and taraxasterol (**42**) arose from a taraxasteryl cation (**33**) by a



Scheme 16

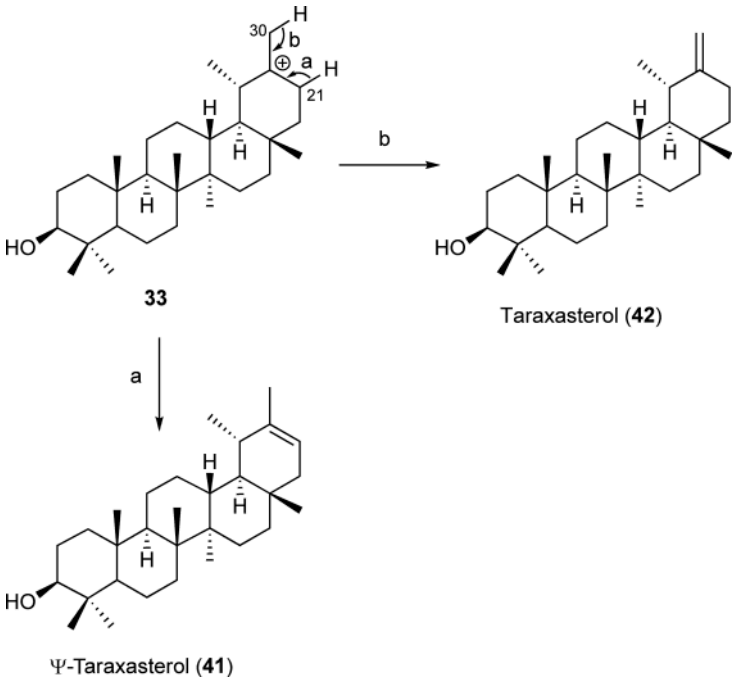


Scheme 17

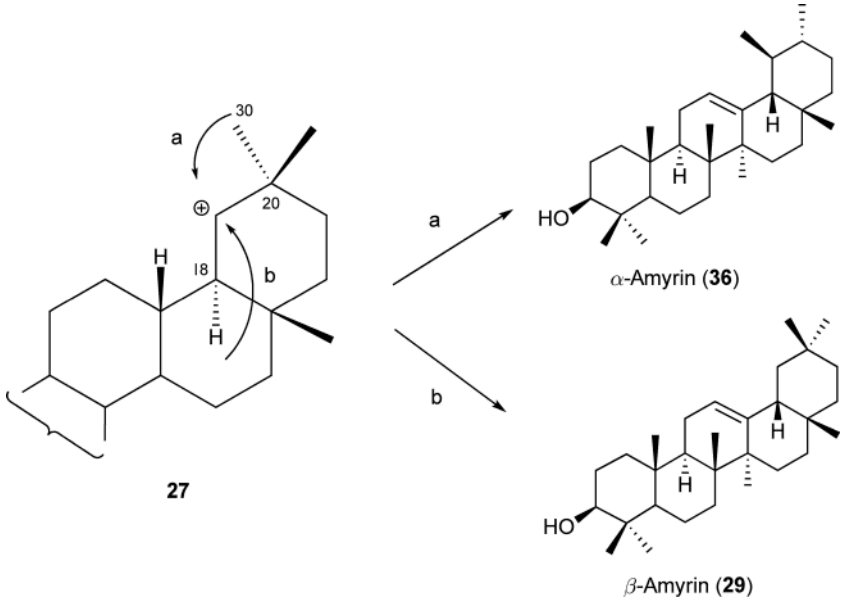


Scheme 18

deprotonation from C-21 and from methyl group C-30, respectively (**Scheme 19**). PSM and AtLUP1 are one of the first multifunctional triterpene synthases to be identified. A product-specific β-amyrin synthase PSY was also found from *P. sativum* and exhibited 78% sequence identity with PSM.[58] A remarkable sequence identity and a difference in the product outcome pose an intriguing question as to how plants have evolved such OSCs to produce a myriad of triterpene products. Formation of α- and β-amyrin, two of the most popular triterpenes found in higher plants, shares the common carbocationic intermediate **27**. A methyl group shift from C-20 (C-30 methyl group) would eventually lead to α-amyrin, whereas a hydride shift from C-18 would give β-amyrin (**Scheme 20**). Therefore, in PSM, competition between methyl and hydride migration takes place to produce both triterpenes in nearly the same amounts. Formation of other minor products indicates that the reaction control is not strict in PSM and that these products arose from termination of a reaction at several intermediate carbocationic species. The existence of such multifunctional enzymes is characteristic of terpene cyclases in general and implies that nature has allowed to evolve such nonspecific cyclases that lost control over the cyclization reaction. This may have an advantage in plants that utilize a myriad of terpenoid products for defense purposes. Until now, none of the α-amyrin-specific OSCs has been cloned. This may imply that α-amyrin is always produced together with β-amyrin by multifunctional OSCs. Several multifunctional OSC genes have been cloned from *C. speciosus*,[38] *L. japonicus*,[42,62] *K. candel*,[76] *O. europaea*,[77] and *R. stylosa*.[65]



Scheme 19



Scheme 20

1.18.4.3 *Arabidopsis* Triterpene Synthases

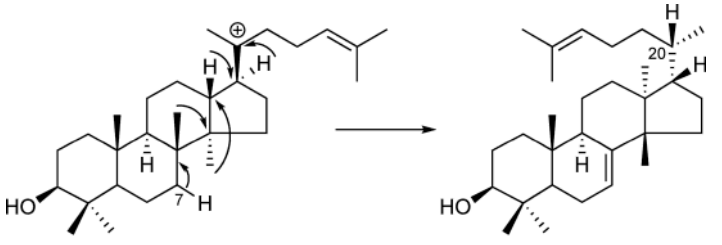
The whole genome sequencing of *A. thaliana* has revealed the existence of 13 OSC genes in the genome. Serving as a model plant, it was of general interest to identify the function of these OSCs. Most of these genes have now been functionally characterized and have uncovered a remarkable catalytic potential of OSC that had not been seen before. These *A. thaliana* OSC genes form distinct branches on the phylogenetic tree, which indicates that these genes evolved independently in this plant lineage.

AtCAS1 (At2g07050) is the first identified cycloartenol synthase from any plants that are responsible for sterol biosynthesis.[32] The gene was found to be essential for sterol biosynthesis in *A. thaliana*. A knockout of the gene caused growth arrest of shoot and root meristems and abnormal growth of leaves as well as albino phenotype in inflorescence shoots resulting from defect in plastid biogenesis.[78]

AtLUP1 (At1g78970) is a multifunctional triterpene synthase that mainly produces lupeol (**26**) (39%) and 3β,20-dihydroxylupane (**30**) (39%), together with minor amounts of β-amyrin (**29**) (8%), germanicol (**39**) (7%), Ψ-taraxasterol (**41**) (4%), and taraxasterol (**42**) (3%).[69] Therefore, AtLUP1 mainly forms a lupanyl cation (**25**); however, minor amounts of cation **27** and taraxasteryl cation (**33**) are also formed.

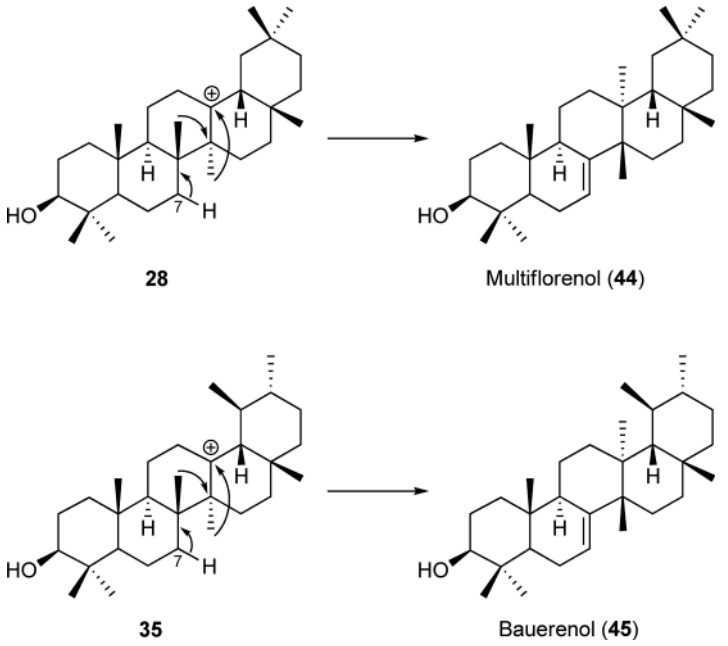
At1g78960 (LUP2) is a multifunctional triterpene synthase that produces nine triterpenes in total,[79] including lupeol (**26**), β-amyrin (**29**), α-amyrin (**36**), taraxasterol (**42**), Ψ-taraxasterol (**41**), butyrospermol (**37**), tirucalla-7,24-diene-3β-ol (**43**), multiflorenol (**44**), and bauerenol (**45**). Although production of taraxasterol and β-amyrin was slightly higher than the other triterpenes, the product ratio of all nine triterpenes was more or less the same. Tetracyclic tirucalla-7,24-diene-3β-ol (**43**) is derived from the dammarenyl cation (**23**) similar to butyrospermol but differs in the stereochemistry at C-20 (**Scheme 21**). Perhaps, the side-chain orientation is different between the two prior to H-17α hydride migration. On the other hand, multiflorenol (**44**) and bauerenol (**45**) are rather rare triterpenes found in plants, and they are derived from oleanyl (**28**) and ursanyl cation (**35**) intermediates through hydride and methyl migration. Again, the reactions are terminated by a deprotonation from C-7 (**Scheme 22**). Therefore, this OSC significantly lacks control over carbocationic intermediates and allows the reaction to terminate at various intermediate stages, for example, dammarenyl cation (**23**), lupanyl cation (**25**), taraxasteryl cation (**33**), oleanyl cation (**28**), and ursanyl cation (**35**). Four out of nine products arose from a deprotonation from C-7, which may suggest that this OSC has to have a strong basic residue around this position. The multiplicity of triterpenes produced by this OSC is unprecedented. Because this OSC exhibited a 79.4% sequence identity with AtLUP1, a region responsible for such a product multiplicity was sought.[79] Several chimeric proteins were made, and it was found that a C-terminal half and a part of an N-terminus were important for the observed multiplicity in At1g78960.

At1g78955 (CAMS1, LUP3) produces monocyclic camelliol C (**18**) (98%) as a major product, together with minor amounts of achilleol A (**17**) (2%) and β-amyrin (**29**) (0.2%).[80] These monocyclic products were derived from cation **1** through a deprotonation from C-1 for camelliol C (**18**) and depronation from a C-10 methyl group (C-25) for achilleol A (**17**) (**Scheme 9**). This is the first triterpene synthase that produces monocyclic triterpene as a major product. Such monocyclic products were produced in several lanosterol and cycloartenol synthase mutants mentioned above whose normal cyclization was aberrantly disrupted at a monocyclic stage. Interestingly, sequence comparison has indicated that this OSC harbors Ala at a position two residues upstream of the DCTAE motif where most plant triterpene synthases and lanosterol synthases contain Val, and whereas cycloartenol synthases contain Ile. This Ile was shown to be important for cycloartenol (**9**) formation as described above. Because overall sequence identity shows CAMS1 to be more closely related to LUP1 and LUP2 (71%), which produce polycyclic triterpenes, this monocyclic-producing OSC may have evolved from a polycyclic-producing OSC one through loss-of-function mutation such as Val to Ala at the critical residue that controls the reaction at a monocyclic intermediate stage. Production of minor amounts of β-amyrin (**29**) was suggested to be an evolutionary remnant of this event.



23 Tirucalla-7,24-diene-3β-ol (**43**)

Scheme 21



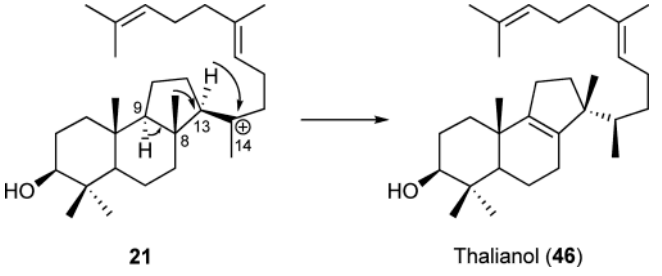
Scheme 22

At1g78950 (AtBAS) produces almost exclusively β-amyrin (**29**) together with trace amounts of some triterpenes including butyrospermol (**37**).[68] Therefore, AtBAS was designated as the β-amyrin synthase of *A. thaliana*. LUP1, LUP2, CAMS1, and AtBAS are all located in tandem on chromosome 1 and may have evolved through a tandem duplication, as they all exhibit a high sequence identity to each other.

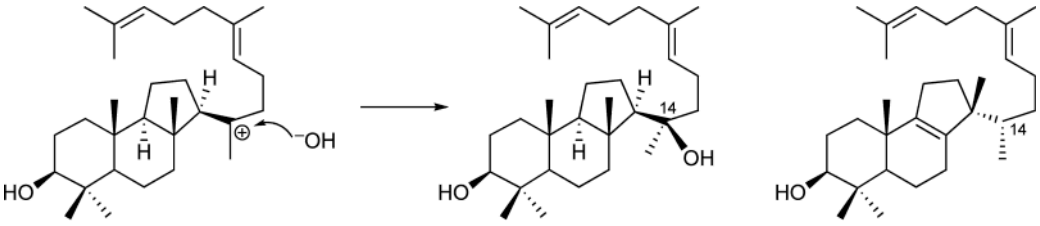
Notably, At1g66960 (LUP5) produces three products, of which one has been identified as tirucalla-7,21-dien-3β-ol (**43**).[81]

At5g48010 (THA1, PEN4) produces a tricyclic malabaricane-type triterpene named thalianol (3*S*,13*S*,14*R*)-malabarica-8,17,21-trien-3-ol (**46**) as the sole product.[82] Thalianol (**46**) is derived from 6-6-5 fused tricyclic carbocation **21**, which undergoes a hydride shift from C-13 to C-14 and a methyl shift from C-8 to C-13, followed by a deprotonation from C-9 (**Scheme 23**). Thalianol is a novel triterpene and such a rearranged malabaricane is very rare in plants. A C-3 analog has been identified from ferns.

At4g15340 produces a similar tricyclic malabaricane-type triterpene with two hydroxyl groups named arabidiol (3*S*,13*R*)-malabarica-17,21-dien-3,14-diol (**47**).[83] Arabidiol (**47**) is derived from the same tricyclic carbocation **21** by the addition of water without any rearrangements (**Scheme 24**). The stereochemistry of the tertiary alcohol was deduced to be 14*R* as shown in **Scheme 24**, which suggests that a water attack on a carbocation takes place from the sterically less hindered side of carbocation **21**.[84] This OSC also produces 14-epithalianol (**48**) as a minor product.[84]



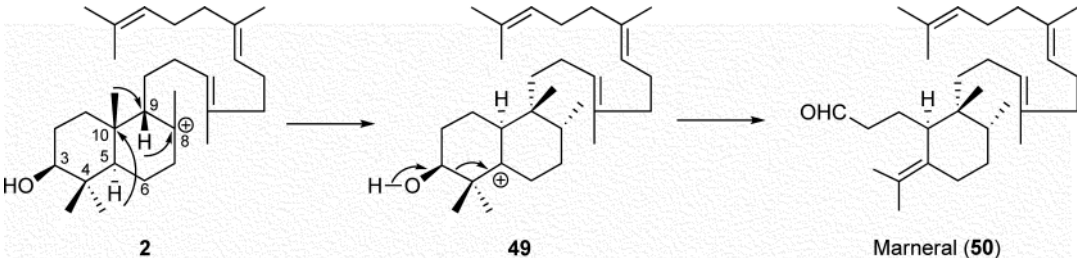
Scheme 23



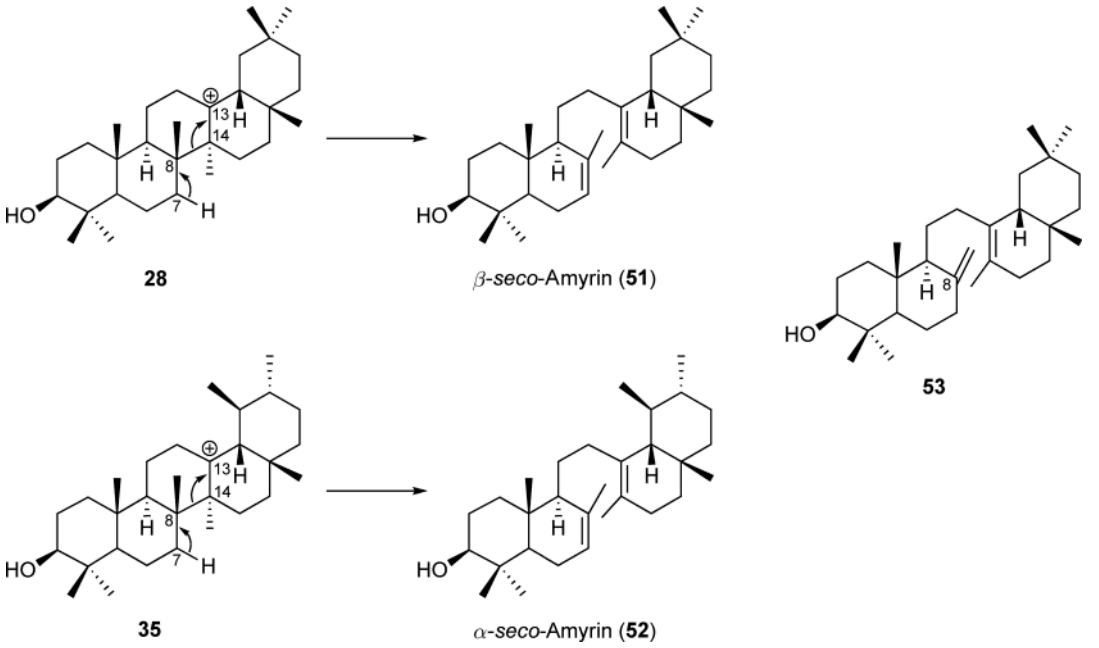
Scheme 24

Surprisingly, At3g45130 (LSS1, LAS1) produces lanosterol (**7**) as the sole product and is the first lanosterol synthase to be identified from the plant kingdom.[29,30] It had been long believed that plants produce cycloartenol (**9**) as a cyclization product of oxidosqualene and that cycloartenol serves as a precursor for all plant sterols, although some plants were found to contain lanosterol. This lanosterol was believed to be derived from cycloartenol through rearrangement of the cyclopropane ring to form a Δ^8 structure. With the discovery of lanosterol synthase in plants, the physiological role of lanosterol in plants is highly intriguing and may have a role in supporting the biosynthesis of some phytosterols. Recent studies, however, indicated a very minor contribution of LAS1 toward sterol biosynthesis in *A. thaliana*. The majority of are still produced by the cycloartenol synthase, AtCAS1.[85] Lanosterol synthase genes have also been identified from *P. ginseng*[30] and *L. japonicus*.[31] These plant lanosterol synthases share a high sequence identity with cycloartenol synthases from plants (~65%) than with fungal (~35%) or mammalian (~40%) lanosterol synthases, indicating a close evolutionary relationship with plant cycloartenol synthases. However, all critical residues that dictate cycloartenol or lanosterol formation were found the lanosterol type. For example, both Asn and Val residues were found in the place that corresponds to His477 and Ile481, respectively, of AtCAS1. As mentioned above, mutations in these two residues of AtCAS1 (H477N/I481V) have converted cycloartenol synthase into a nearly perfect lanosterol synthase.[56]

At5g42600 (MRN1, PEN5) produces *seco* triterpene aldehyde **50** named marneral,[86] the basic carbon skeleton of iridals, and arose through a ring-A cleavage of a bicyclic skeleton. The cyclization mechanism can be rationalized as follows (**Scheme 25**). The formation of AB rings involves oxidosqualene folded in *chair–boat* conformation. The B-ring boat conformation is rather unusual among plant triterpenes. The intermediate carbocation **2** then undergoes hydride shift from C-9 to C-8 followed by a methyl group shift from C-10 to C-9 and a hydride shift from C-5 to C-10 to result in carbocation **49**. This then undergoes Grob fragmentation through hyperconjugation of C3–C4 σ bond with C-5 carbocation to give *seco* aldehyde **50**. It is somewhat remarkable that this OSC allows only the C-5 carbocation to undergo Grob fragmentation, despite other pathways such as a deprotonation from C-6 to give a Δ^5 structure or a 4β-methyl migration from C-4 followed by 3α-hydride shift to give a 3-keto structure. It has been proposed that it is the B-ring conformation and the flexibility of a presumed Asp residue that abstracts a proton from a C-3 hydroxyl group that are important factors for this Grob fragmentation selectivity. This was the first observation of OSC catalyzing a preformed ring cleavage to produce a *seco* triterpene. Iridals have been only found from Iridaceae plants. In addition to **50**, this OSC also produces minor amounts of achilleol A (**17**) and camelliol C (**18**).



Scheme 25



Scheme 26

At1g78500 (PEN6) produces a mixture of triterpenes, but most notably produces *seco* triterpenes **51** and **52**. These are oleanane and ursane types of C-ring *seco* triterpenes.[87] These *seco* triterpenes arose from oleanyl (**28**) and ursanyl (**35**) cations through Grob fragmentation, which results in C-ring cleavage between C-8 and C-14 to form a C13–C14 double bond followed by a deprotonation from C-7 to give a C7–C8 double bond (**Scheme 26**). These are other examples of OSC catalyzing the formation of *seco* triterpenes. Compounds **51** and **52** are *β-seco*-amyrin and *α-seco*-amyrin, respectively. Only the *β-seco*-amyrin-type triterpene (**53**) has been isolated from plants, *Stevia viscida* and *S. eupatoria*, with an *exo*-methylene at C-8. Besides **51** and **52**, this OSC produces lupeol (**26**), α-amyrin (**36**), and bauerenol (**45**) as well.[81]

Finding these *seco* triterpene-producing OSCs has led to the proposal that other natural *seco* triterpenes may also be products of OSCs. Such products include graminol A (**54**),[88] helianol (**55**),[89] sasanquol (**56**),[90] camelliol A (**57**),[91] and camelliol B (**58**)[91] (**Figure 7**). Notably, camelliol B can be regarded as an A/B/C-ring *seco* triterpene.

At4g15370 (BARS1) produces almost exclusively baruol (**61**) (90%) and 22 minor triterpenes (0.02–3% each).[92] Baruol (**61**) is a baccharane-type triterpene derived from baccharenyl cation **24** through extensive hydride and methyl shifts (six in total) to generate C-5 carbocation **60**, which is followed by a deprotonation from C-6 (**Scheme 27**). The minor products include lemmaphylladienol (**62**) and columbiol (**63**), which arose from the intermediate carbocation **59** through a deprotonation from C-7 and C-9, respectively, and sasanqual (**64**). Sasanqual is another example of an A-ring *seco* triterpene formed from OSC. The ability of this OSC to generate such a multitude of products is remarkable.

1.18.4.4 Other Triterpene Synthases

1.18.4.4.1 Isomultiflorenol synthase

Isomultiflorenol synthase (IMS) gene was cloned from *Luffa cylindrica* cell suspension culture,[93] which is known to produce bryonolic acid, a C-29 oxidized form of isomultiflorenol. Isomultiflorenol (**65**) arose from oleanyl cation (**28**) through two methyl group shifts to generate the C-8 carbocation from which a deprotonation from C-9 takes place to give Δ^8 (**Scheme 28**). IMS exhibited a 65% sequence identity with *β*-amyrin synthase (PNY) and a 57% sequence identity with lupeol synthase (OEW). A rather high sequence identity with *β*-amyrin synthase is reasonable considering that **65** is formed from the same intermediate oleanyl cation (**28**).

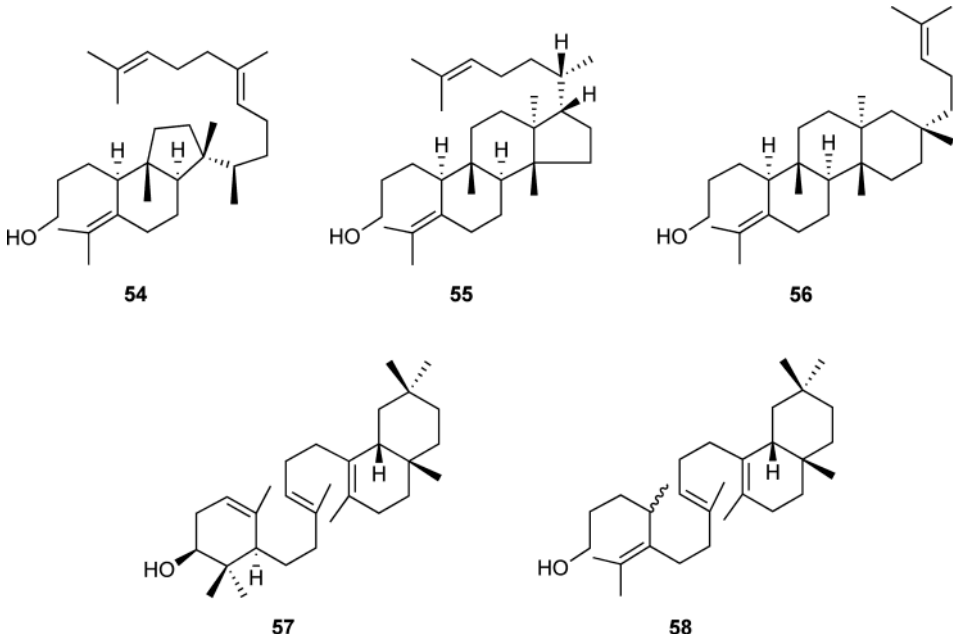


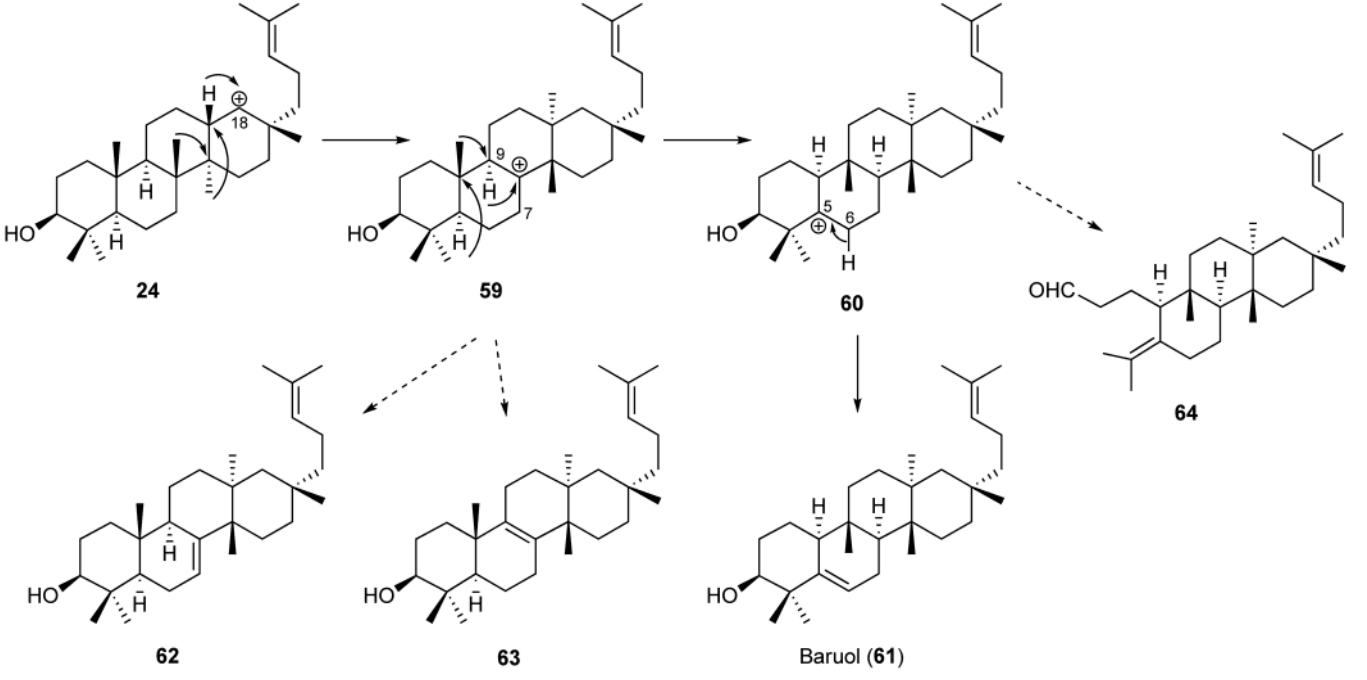
Figure 7 Structures of natural *seco* triterpenes found in plants.

1.18.4.4.2 Dammarenediol synthase

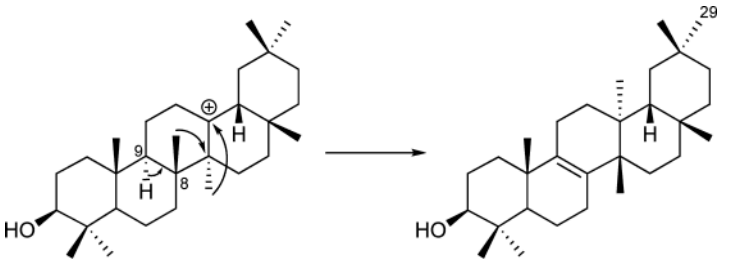
(20*S*)-Dammarenediol (**66**) is the basic skeleton of triterpene saponins from *P. ginseng*, known as ginsenosides, the most popular and widely used Chinese traditional medicine. It is derived from tetracyclic dammarenyl cation (**23**) by a stereospecific water addition onto C-20 (**Scheme 29**). The stereochemistry at C-20 of the enzymatically formed product was rigorously proven to be 20*S*.[94] All the naturally occurring ginsenosides have a 20*S* configuration; however, when *P. ginseng* is treated with steam water to produce red ginseng, some ginsenosides with a 20*R* configuration are produced. This epimerization at C-20 through dehydration and rehydration is assisted by a 12β-hydroxyl group.[95] Dammarenediol synthase (PNA) was cloned from *P. ginseng* hairy root culture[96] and exhibited ~57% sequence identity with β-amyrin synthase (PNY) and lupeol synthase (OEW). This is one of the rare cases in OSC that the reaction is terminated by adding water to a carbocationic center. Other OSCs known to produce such hydroxylated products are AtLUP1, which gives 3β,20-dihydroxylupane (**30**), and At4g15340, which gives arabidiol (**47**).

1.18.4.4.3 Cucurbitadienol synthase

Cucurbitane types are exceptional plant triterpenes that arose from protosteryl cation (**5**), which is formed by oxidosqualene folded in a *chair–boat–chair* conformation as in lanosterol (**7**) and cycloartenol (**9**). Cucurbitadienol (**68**) biosynthesis follows the same pathway with cycloartenol (**9**) up to C-9 carbocation intermediate **8**, where a further methyl shift from C-10 to C-9 and a hydride shift from C-5 to C-10 give C-5 carbocation **67**. A deprotonation from C-6 will furnish Δ^5 (**Scheme 30**). Cucurbitacins are highly oxygenated triterpenes characteristic of Cucurbitaceae plants and are often associated with a bitter taste such as in cucumber or *Momordica charantia*. Cucurbitadienol synthase (CPQ) was cloned from pumpkin *Cucurbita pepo*[40], and the amino acid sequence of CPQ showed a high sequence identity (~68%) with cycloartenol synthases, which is reasonable considering the resemblance of the cyclization mechanism. It contains Ile at the position two residues upstream of the DCTAE motif (corresponds to Ile481 of AtCAS1). CPQ and cycloartenol synthases are the only OSCs that harbor Ile at this position and presumably this Ile plays an important role in the generation of C-9 carbocation and beyond.

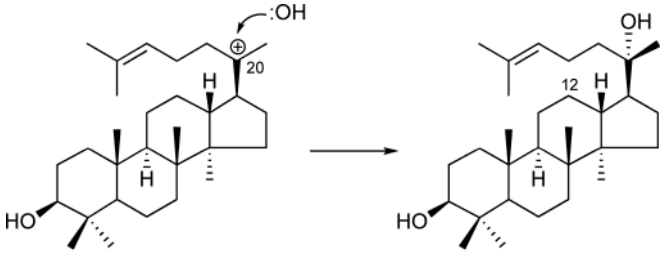


Scheme 27



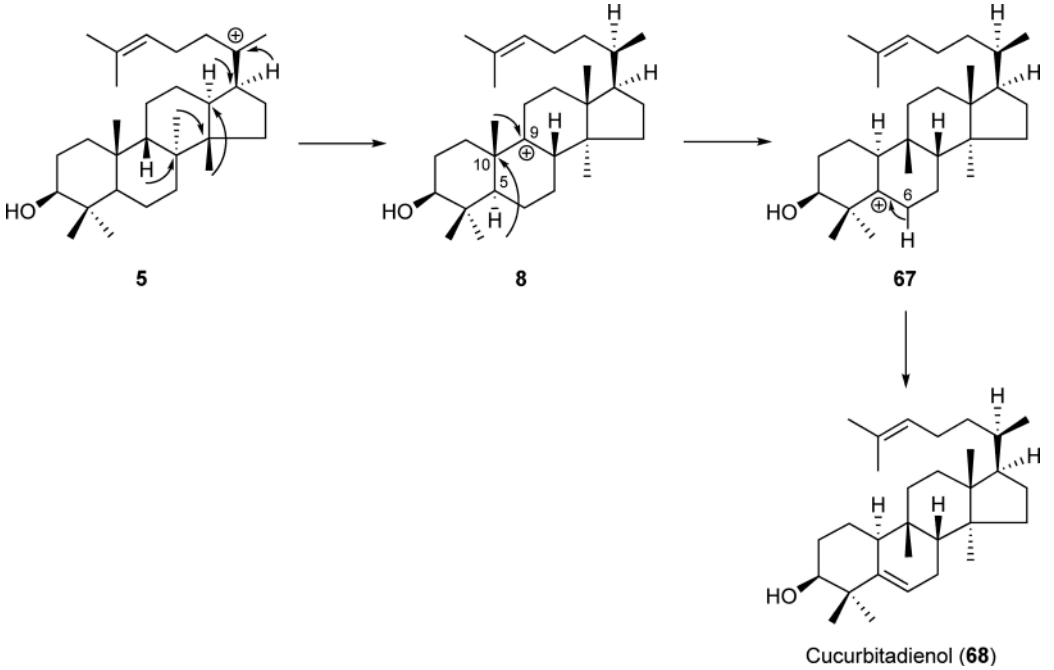
28 Isomultiflorenol (**65**)

Scheme 28



23 (20*S*)-Dammarenediol (**66**)

Scheme 29



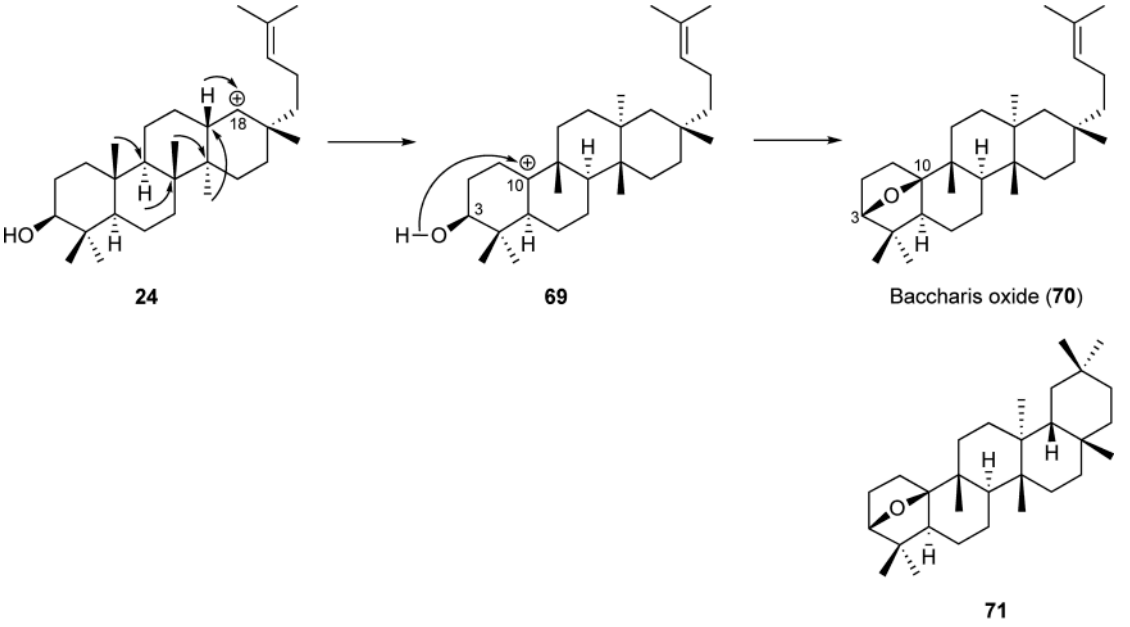
5 **8** **67**

Cucurbitadienol (**68**)

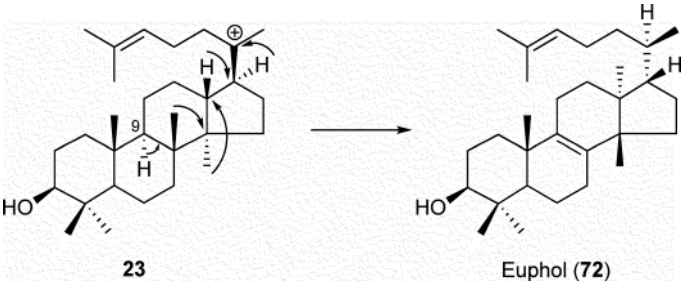
Scheme 30

1.18.4.4.4 Baccharis oxide synthase

The baccharis oxide (**70**) belongs to a baccharane-type triterpene that has an unusual C3–C10 oxide bridge over the A ring. The baccharis oxide is derived from baccharenyl cation (**24**) by a sequence of methyl and hydride shifts to give C-10 carbocation **69** to which the OH group at C-3 attacks to terminate the reaction (**Scheme 31**). Such an ether bridge in triterpene is found only in dendropanoxide (**71**), which is derived from an oleanyl cation (**28**).[97] The baccharis oxide has been isolated from Asteraceae plants. The baccharis oxide synthase (BOS) gene was cloned from *S. rebaudiana* and was also found to produce minor amounts of achilleol A (**17**), camelliol C (**18**), baruol (**61**), sasanquol (**56**), euphol (**72**), butyrospermol (**37**), and β-amyrin (**29**).[98] Production of **70** was 75% of the total products. Euphol (**72**) belongs to the euphane type, the same as butyrospermol (**37**), and arose from a dammarenyl cation (**23**) through two hydride and two methyl group shifts and a final deprotonation from C-9 (**Scheme 32**). This gene showed an ~70% sequence identity with β-amyrin synthases and an ~57% identity with lupeol synthases. BOS is characterized in having a Thr residue at the position two residues upstream of the DCTAE motif, which is normally Val or Ile (or Ala in *A. thaliana* CAMS1). Whether or not this Thr plays a role in oxide bridge formation remains to be studied. Evidently, this OSC forces the intermediate to adopt an A-ring *boat* conformation in order to allow an attack of the C-3 OH group onto the C-10 carbocation.



Scheme 31



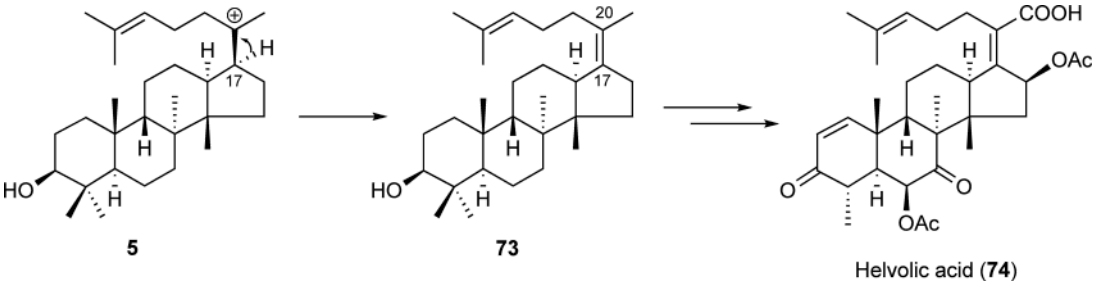
Scheme 32

1.18.4.4.5 Protostadienol synthase

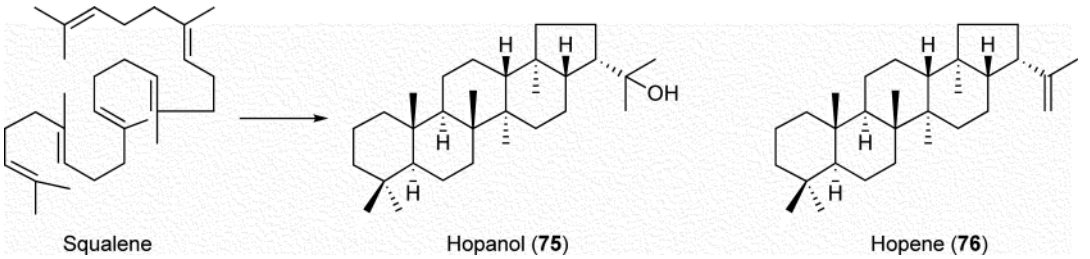
A fungal OSC that produces protosta-17(20)*Z*,24-dien-3β-ol (**73**), which is the basic carbon skeleton of fusidanes, was cloned from the fungus *Aspergillus fumigatus*.[99] Compound **73** arose from protosteryl cation (**5**) through from depronation C-17 without any hydride and methyl group shifts (**Scheme 33**). The characteristic *Z* configuration of $\Delta^{17(20)}$ that is found among fusidanes indicates that the side-chain moiety is folded in such a way that it does not require a substantial bond rotation between C-17 and C-20. This side-chain geometry is presumably conserved in lanosterol and cycloartenol synthases, as they all share 20*R* stereochemistry that can be achieved by a hydride migration from the α-face of C-17 without a significant bond rotation. This OSC gene was found clustered with eight other genes in the *A. fumigatus* genome, which includes genes for cytochrome P450 and short-chain dehydrogenase/reductase (SDR), which are presumed to be involved in the biosynthesis of helvolic acid (**74**). Fusidanes are clinically used antibiotics effective against Gram-positive bacteria such as *Staphylococcus aureus*.

1.18.4.4.6 Squalene cyclases from ferns

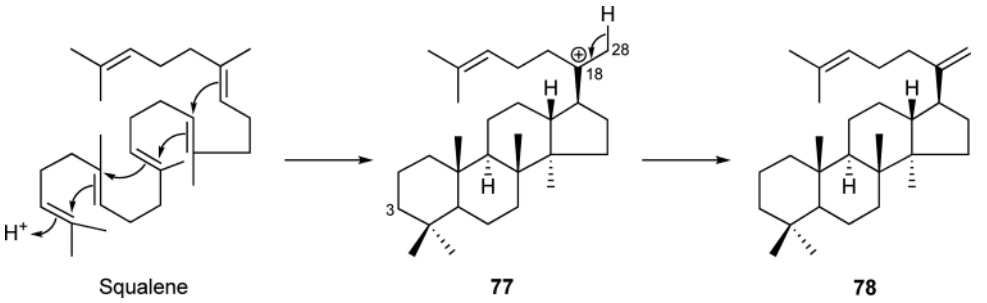
Ferns are known to produce a variety of triterpenes that lack C-3 oxygen functionality including hopanoids and migrated hopanoids. These triterpenes are believed to be derived from a cyclization of squalene. Therefore, ferns are expected to contain many squalene cyclase (SC) genes. The first SC gene from a fern *A. capillus-veneris* was identified and found to produce hopanol (**75**) as the sole product (**Scheme 34**).[44] This was in sharp contrast with bacterial SCs, which usually produce a mixture of hopene (**76**) and hopanol (**75**) in various ratios (Chapter 1.19). A cycloartenol synthase gene was also cloned from this fern, demonstrating that ferns contain both SC and OSC. Another SC gene was identified from *Dryopteris crassirhizoma* that produced dammara-18(28),21-diene (**78**).[100] This is the first demonstration of an SC that produces triterpene other than hopanoids. The dammarane-type triterpene **78** arose from C-3 deoxy dammarenyl cation **77** by a deprotonation from C-28 (**Scheme 35**). This SC exhibited 66% sequence identity with the one cloned from *A. capillus-veneris*. The fern SC sequences formed a distinct clade in the phylogenetic tree away from the bacterial sequences, indicating an independent evolution in the fern lineage.



Scheme 33



Scheme 34



Scheme 35

1.18.5 Triterpene Tailoring Steps

The formation of mature biologically active triterpenoids involves further modification of triterpene skeletons generated by OSCs, including an oxidation and a glycosylation. Triterpene saponins are found in many pharmaceutically important medicinal plants and exhibit rich biological activities. Despite extensive studies on OSCs, only a few limited studies are available for triterpene tailoring steps. Triterpene oxidation is generally thought to be catalyzed by cytochrome P450 monooxygenases (P450s).

The first P450 catalyzing the oxygenation of a triterpene substrate was identified from *G. max*, CYP93E1.[101] This P450 catalyzed a hydroxylation at C-24 of β-amyrin (**29**) and sophoradiol (**79**) to produce olean-12-ene-3β,24-diol (**80**) and soyasapogenol B (**81**), respectively (**Scheme 36**). This CYP93E1 is responsible for the soyasaponin biosynthesis in *G. max*. Soyasaponins, such as soyasaponin I (**82**) and soyasapogenol B, are known to possess strong hepatoprotective activity.

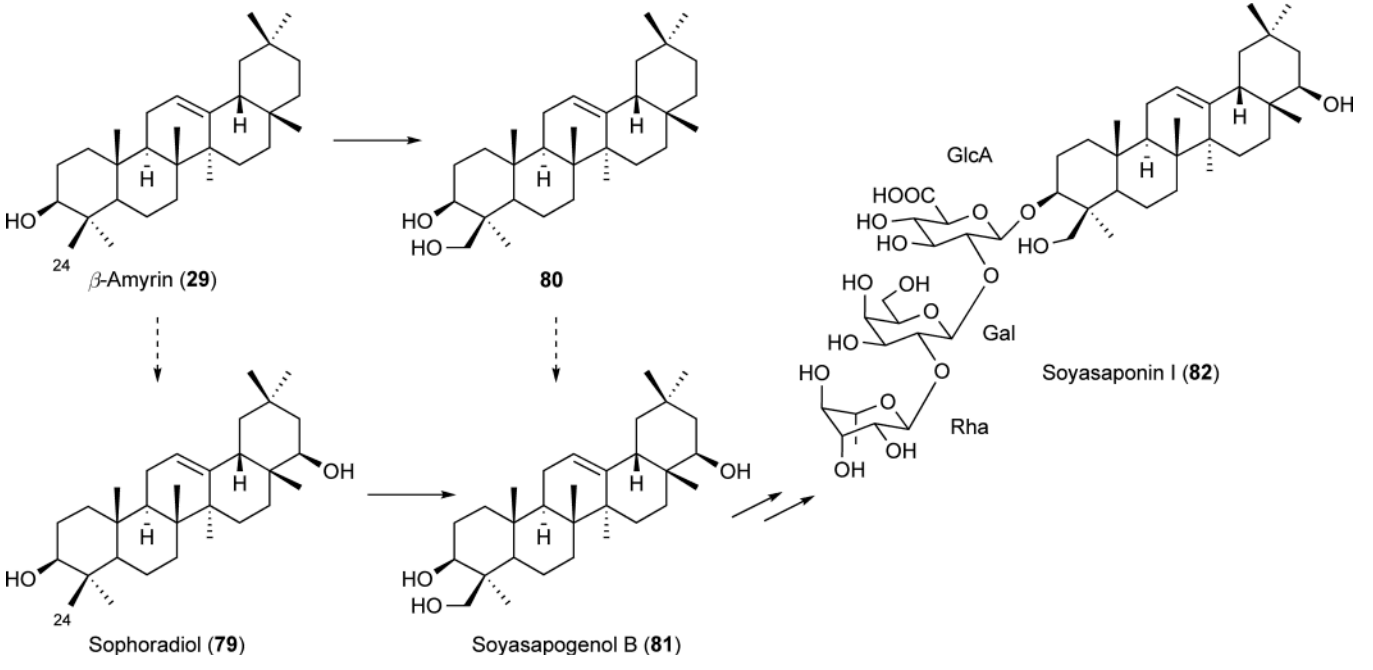
A second P450 gene that catalyzes triterpene oxygenation was identified from *G. uralensis*, CYP88D6.[102] This P450 catalyzed an oxygenation at C-11 of β-amyrin (**29**) to produce 11-oxo-β-amyrin (**84**) through 11α-hydroxy-β-amyrin (**83**) (**Scheme 37**). The two-step oxygenation that gives the 11-keto functionality is a characteristic step toward the biosynthesis of glycyrrhizin (**85**), a natural sweet compound that which possesses anti-inflammatory, immunomodulatory, antiulcer, antiallergy, and antiviral activities.

Fungal genes that modify protosta-17(20)*Z*,24-dien-3β-ol (**73**) were identified in the presumed helvolic acid biosynthetic gene cluster from *A. fumigatus*. A P450, CYP5081A1, was shown to catalyze the oxygenation of the 4β-methyl group to produce a diol (**86**) and a carboxylic acid (**87**) derivative (**Scheme 38**).[99] Another gene in the cluster that codes for an SDR was found to catalyze a dehydrogenation of the 3β-hydroxyl group of **73** to produce a 3-keto derivative (**88**).[99] These genes presumably function together to catalyze the demethylation of the 4β-methyl group during helvolic acid biosynthesis.

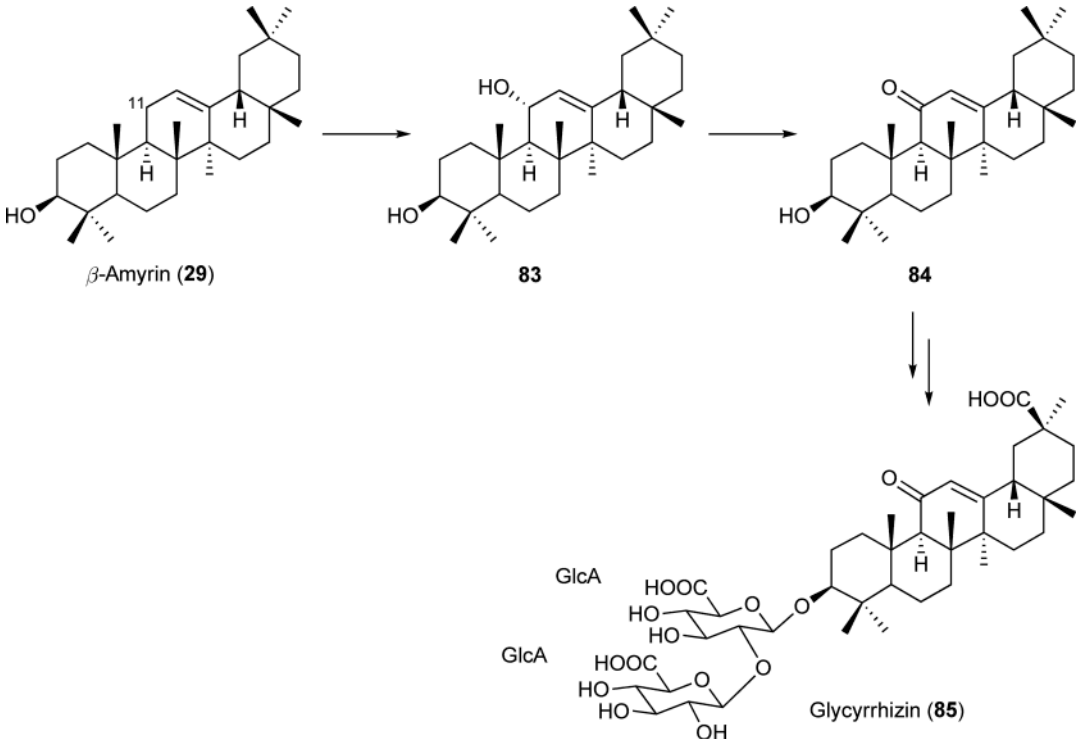
In oat *A. strigosa*, a P450 gene CYP51H10 was implicated as being involved in the biosynthesis of oleanane-type antifungal triterpene saponins, avenacins.[103] Mutation of this gene in oat resulted in a saponin-deficient phenotype. Therefore, CYP51H10 is expected to function in some oxidative modification steps of β-amyrin; however, the exact function of this gene remains to be studied.

In *A. thaliana*, two P450 genes were implicated as being involved in a triterpene tailoring reaction. In the vicinity of an OSC gene At5g48010 (thalianol synthase) in the genome, two P450 genes (CYP708A2 and CYP705A5) were found whose expression patterns were similar to the OSC gene. Both inactivation and overexpression studies of these genes resulted in accumulation of peaks on GC–MS, which were suggested to be thalianol (**46**) derivatives.[104] The exact function of these P450s also remains to be studied.

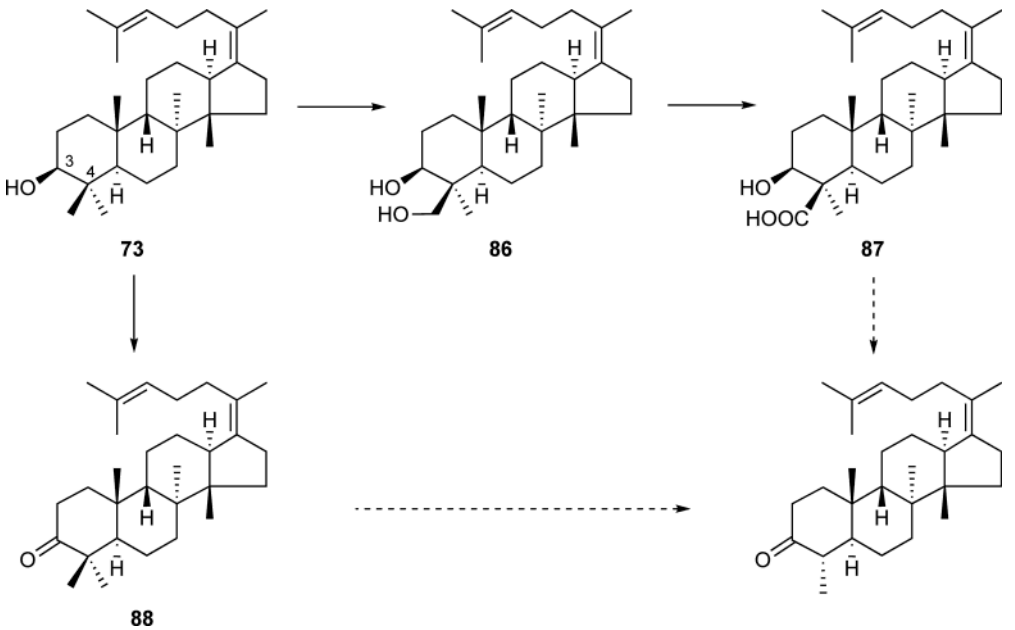
A glycosyltransferase gene responsible for a triterpene saponin biosynthesis was identified from *S. vaccaria*.[64] The plant family I glycosyltransferase, UGT74M1, was found to catalyze the attachment of glucose from uridine 5′-diphosphate-glucose (UDP-glucose) onto C-28 carboxylic acid of gypsogenic acid (**89**) to produce monoglucoside (**90**) (**Scheme 39**). This glycosyltransferase is presumably involved in the biosynthesis of the saponin vaccaroside B found in this plant.


β-Amyrin (29)
80
Sophoradiol (79)
Soyasapogenol B (81)
GlcA
Gal
Rha
Soyasaponin I (82)

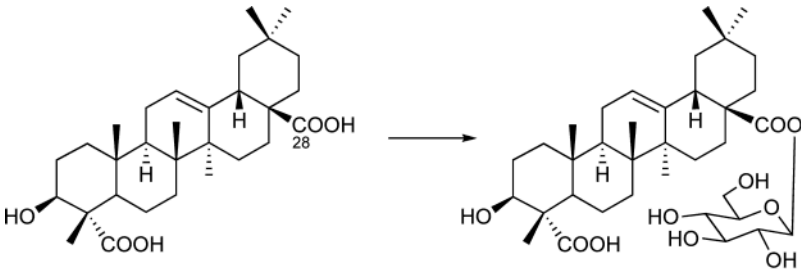
Scheme 36



Scheme 37



Scheme 38

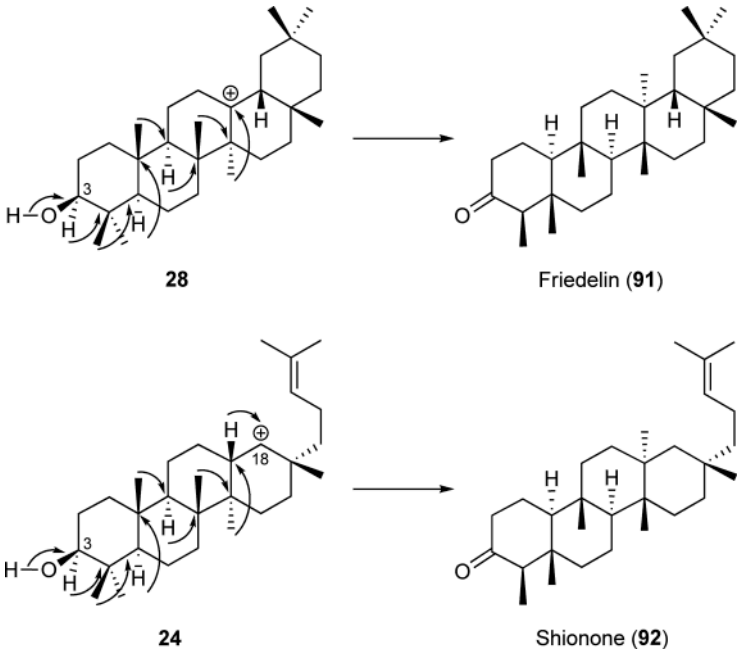


Gypsogenic acid (**89**) **90**

Scheme 39

1.18.6 Summary and Future Perspectives

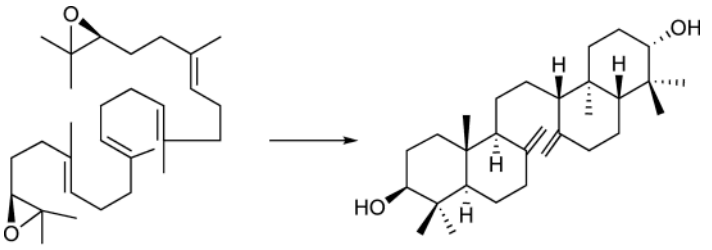
The cyclization of oxidosqualene into various triterpenes has fascinated chemists for more than 60 years due to its remarkable chemistry and its ability to produce hundreds of different products from a single precursor. In recent years, significant progress in our understanding of the enzyme OSC has been made. These studies have uncovered the chemistry, enzymology, and genetics of OSCs to reveal the molecular mechanism of the catalysis and their biological roles in the cells. Notably, the first crystal structure of human lanosterol synthase has allowed us to observe the detailed structure of the active site. In addition, cloning of various plant OSC genes has revealed the existence of both product-specific and multifunctional OSCs in plants, which accounts for the generation of significant numbers of triterpene carbon skeletons found in nature. Our further understanding of such triterpene biosynthesis requires us to provide a detailed structural analysis of plant OSCs and to identify critical residues governing their product specificity. In this line, further cloning of OSC genes responsible for triterpenes that arise from mechanistically novel pathways would be valuable. Such triterpenes include friedelin (**91**), shionone (**92**), and onocerin (**94**). Friedelin (**91**) and shionone (**92**) arise from an oleanyl cation (**28**) and a baccharenyl cation (**24**), respectively, from extensive methyl and hydride shifts all the way up to C-3 without intervention by deprotonation; they are terminated with a ketone formation (**Scheme 40**). Studies on



28 Friedelin (**91**)

24 Shionone (**92**)

Scheme 40



93 Onocerin (94)

Scheme 41

these OSCs would give us further insights into mechanisms that control Wagner–Meerwein shifts. On the other hand, onocerin (94) presumably arises from 2,3,22,23-dioxidosqualene (93) cyclized from both ends producing a bicyclic structure each (Scheme 41). Whether such products arise within a single active site or require the release of a singly cyclized product and recapture for another second round of cyclization is intriguing. In addition, triterpene tailoring steps are important for modifying triterpene skeletons to exhibit various biological activities. Currently, only a few limited studies are available for such triterpene tailoring steps. Identification of more genes involved in oxygenation and glycosylation is necessary for future engineering of triterpene biosynthesis to produce useful bioactive triterpenes.

Abbreviations

29-MOS	29-methylidene-2,3-oxidosqualene
BOS	baccharis oxide synthase
FPP	farnesyl diphosphate
IMS	isomultiflorenol synthase
OSC	oxidosqualene cyclase
P450	cytochrome P450 monooxygenase
SC	squalene cyclase
SDR	short-chain dehydrogenase/reductase
SHC	squalene:hopene cyclase

References

1. I. Shechter; G. Guan, *Comprehensive Natural Products Chemistry*; Elsevier: Oxford, 1999; Vol. 2, Chapter 9, pp 245–266.
2. I. Abe; G. D. Prestwich, *Comprehensive Natural Products Chemistry*; Elsevier: Oxford, 1999; Vol. 2, Chapter 10, pp 267–298.
3. A. Eschenmoser; D. Arigoni, *Helv. Chim. Acta* **2005**, *88*, 3011.
4. I. Abe; M. Rohmer; G. D. Prestwich, *Chem. Rev.* **1993**, *93*, 2189.
5. D. W. Christianson, *Chem. Rev.* **2006**, *106*, 3412.
6. R. Rajamani; J. Gao, *J. Am. Chem. Soc.* **2003**, *125*, 12768.
7. R. Xu; G. C. Fazio; S. P. T. Matsuda, *Phytochemistry* **2004**, *65*, 261.
8. E. J. Corey; H. Cheng; C. H. Baker; S. P. T. Matsuda; D. Li; X. Song, *J. Am. Chem. Soc.* **1997**, *119*, 1277.
9. E. J. Corey; S. C. Virgil; H. Cheng; C. H. Baker; S. P. T. Matsuda; V. Singh; S. Sarshar, *J. Am. Chem. Soc.* **1995**, *117*, 11819.
10. E. J. Corey; S. C. Virgil, *J. Am. Chem. Soc.* **1991**, *113*, 4025.
11. E. J. Corey; S. C. Virgil; D. R. Liu; S. Sarshar, *J. Am. Chem. Soc.* **1992**, *114*, 1524.
12. R. Kelly; S. M. Miller; M. H. Lai; D. R. Kirsch, *Gene* **1990**, *87*, 177.
13. C. J. Buntel; J. H. Griffin, *J. Am. Chem. Soc.* **1992**, *114*, 9711.
14. C. A. Roessner; C. Min; S. H. Hardin; L. W. Harris-Haller; J. C. McCollum; A. I. Scott, *Gene* **1993**, *127*, 149.
15. E. J. Corey; S. P. T. Matsuda; B. Bartel, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 2211.
16. Z. Shi; C. J. Buntel; J. H. Griffin, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7370.
17. E. J. Corey; S. P. T. Matusda; C. H. Baker; A. Y. Ting; H. Cheng, *Biochem. Biophys. Res. Commun.* **1996**, *219*, 327.
18. M. Kusano; M. Shibuya; U. Sankawa; Y. Ebizuka, *Biol. Pharm. Bull.* **1995**, *18*, 195.

19. I. Abe; G. D. Prestwich, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9274.
20. C. K. Sung; M. Shibuya; U. Sankawa; Y. Ebizuka, *Biol. Pharm. Bull.* **1995**, *18*, 1459.
21. C. H. Baker; S. P. T. Matsuda; D. R. Liu; E. J. Corey, *Biochem. Biophys. Res. Commun.* **1995**, *213*, 154.
22. K. Poralla; A. Hewelt; G. D. Prestwich; I. Abe; I. Reipen; G. Sprenger, *Trends Biochem. Sci.* **1994**, *19*, 157.
23. K. Poralla, *Bioorg. Med. Chem. Lett.* **1994**, *4*, 285.
24. R. Thoma; T. Schulz-Gasch; B. D'Arcy; J. Benz; J. Aebi; H. Dehmlow; M. Hennig; M. Stihle; A. Ruf, *Nature* **2004**, *432*, 118.
25. F. S. Buckner; L. N. Ngyuen; B. M. Joubert; S. P. T. Matsuda, *Mol. Biochem. Parasitol.* **2000**, *110*, 399.
26. B. M. Joubert; F. S. Buckner; S. P. T. Matsuda, *Org. Lett.* **2001**, *3*, 1957.
27. I. Abe; K. Naito; Y. Takagi; H. Noguchi, *Biochim. Biophys. Acta* **2001**, *1522*, 67.
28. P. Milla; F. Viola; S. Oliaro-Bosso; F. Rocco; L. Cattel; B. M. Joubert; R. J. LeClair; S. P. T. Matsuda; G. Balliano, *Lipids* **2002**, *37*, 1171.
29. M. D. Kolesnikova; Q. Xiong; S. Lodeiro; L. Hua; S. P. T. Matsuda, *Arch. Biochem. Biophys.* **2006**, *447*, 87.
30. M. Suzuki; T. Xiang; K. Ohyama; H. Seki; K. Saito; T. Muranaka; H. Hayashi; Y. Katsube; T. Kushiro; M. Shibuya; Y. Ebizuka, *Plant Cell Physiol.* **2006**, *47*, 565.
31. S. Sawai; T. Akashi; N. Sakurai; H. Suzuki; D. Shibata; S. Ayabe; T. Aoki, *Plant Cell Physiol.* **2006**, *47*, 673.
32. E. J. Corey; S. P. T. Matsuda; B. Bartel, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11628.
33. M. Morita; M. Shibuya; M.-S. Lee; U. Sankawa; Y. Ebizuka, *Biol. Pharm. Bull.* **1997**, *20*, 770.
34. T. Kushiro; M. Shibuya; Y. Ebizuka, *Eur. J. Biochem.* **1998**, *256*, 238.
35. H. Hayashi; N. Hiraoka; Y. Ikeshiro; K. Yazaki; S. Tanaka; T. Kushiro; M. Shibuya; Y. Ebizuka, *Plant Physiol.* **1999**, *121*, 1384.
36. S. M. Godzina; M. A. Lovato; M. M. Meyer; K. A. Foster; W. K. Wilson; W. Gu; E. L. de Hostos; S. P. T. Matsuda, *Lipids* **2000**, *35*, 249.
37. H. Hayashi; N. Hiraoka; Y. Ikeshiro; T. Kushiro; M. Morita; M. Shibuya; Y. Ebizuka, *Biol. Pharm. Bull.* **2000**, *23*, 231.
38. N. Kawano; K. Ichinose; Y. Ebizuka, *Biol. Pharm. Bull.* **2002**, *25*, 477.
39. H. Zhang; M. Shibuya; S. Yokota; Y. Ebizuka, *Biol. Pharm. Bull.* **2003**, *26*, 642.
40. M. Shibuya; S. Adachi; Y. Ebizuka, *Tetrahedron* **2004**, *60*, 6995.
41. O. Guhling; B. Hobl; T. Yeats; R. Jetter, *Arch. Biochem. Biophys.* **2006**, *448*, 60.
42. S. Sawai; T. Shindo; S. Sato; T. Kaneko; S. Tabata; S. Ayabe; T. Aoki, *Plant Sci.* **2006**, *170*, 247.
43. M. Basyuni; H. Oku; E. Tsujimoto; S. Baba, *Biosci. Biotechnol. Biochem.* **2007**, *71*, 1788.
44. J. Shinozaki; M. Shibuya; K. Masuda; Y. Ebizuka, *FEBS Lett.* **2008**, *582*, 310.
45. D. C. Lamb; C. J. Jackson; A. G. S. Warrilow; N. J. Manning; D. E. Kelly; S. L. Kelly, *Mol. Biol. Evol.* **2007**, *24*, 1714.
46. C. Nakano; A. Motegi; T. Sato; M. Onodera; T. Hoshino, *Biosci. Biotechnol. Biochem.* **2007**, *71*, 2543.
47. H. B. Bode; B. Zeggel; B. Silakowski; S. C. Wenzel; H. Reichenbach; R. Müller, *Mol. Microbiol.* **2003**, *47*, 471.
48. I. Abe; G. D. Prestwich, *J. Biol. Chem.* **1994**, *269*, 802.
49. E. J. Corey; H. Cheng; C. H. Baker; S. P. T. Matsuda; D. Li; X. Song, *J. Am. Chem. Soc.* **1997**, *119*, 1289.
50. E. A. Hart; L. Hua; L. B. Darr; W. K. Wilson; J. Pang; S. P. T. Matsuda, *J. Am. Chem. Soc.* **1999**, *121*, 9887.
51. B. M. Joubert; L. Hua; S. P. T. Matsuda, *Org. Lett.* **2000**, *2*, 339.
52. S. P. T. Matsuda; L. B. Darr; E. A. Hart; J. B. R. Herrera; K. E. McCann; M. M. Meyer; J. Pang; H. G. Schepmann, *Org. Lett.* **2000**, *2*, 2261.
53. M. J. R. Segura; B. E. Jackson; S. P. T. Matsuda, *Nat. Prod. Rep.* **2003**, *20*, 304.
54. I. Abe, *Nat. Prod. Rep.* **2007**, *24*, 1311.
55. T. K. Wu; C. H. Chang; Y. T. Liu; T. T. Wang, *Chem. Rec.* **2008**, *8*, 302.
56. S. Lodeiro; T. Schulz-Gasch; S. P. T. Matsuda, *J. Am. Chem. Soc.* **2005**, *127*, 14132.
57. J. B. R. Herrera; B. Bartel; W. K. Wilson; S. P. T. Matsuda, *Phytochemistry* **1998**, *49*, 1905.
58. M. Morita; M. Shibuya; T. Kushiro; K. Masuda; Y. Ebizuka, *Eur. J. Biochem.* **2000**, *267*, 3453.
59. H. Hayashi; P. Huang; A. Kirakosyan; K. Inoue; N. Hiraoka; Y. Ikeshiro; T. Kushiro; M. Shibuya; Y. Ebizuka, *Biol. Pharm. Bull.* **2001**, *24*, 912.
60. K. Haralampidis; G. Bryan; X. Qi; K. Papadopoulou; S. Bakht; R. Melton; A. Osbourn, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13431.
61. H. Suzuki; L. Achnine; R. Xu; S. P. T. Matsuda; R. A. Dixon, *Plant J.* **2002**, *32*, 1033.
62. I. Iturbe-Ormaetxe; K. Haralampidis; K. Papadopoulou; A. E. Osbourn, *Plant Mol. Biol.* **2003**, *51*, 731.
63. M. Kajikawa; K. T. Yamato; H. Fukuzawa; Y. Sakai; H. Uchida; K. Ohyama, *Phytochemistry* **2005**, *66*, 1759.
64. D. Meesapyodsuk; J. Balsevich; D. W. Reed; P. S. Covello, *Plant Physiol.* **2007**, *143*, 959.
65. M. Basyuni; H. Oku; E. Tsujimoto; K. Kinjo; S. Baba; K. Takara, *FEBS J.* **2007**, *274*, 5028.
66. J. Kirby; D. W. Romanini; E. M. Paradise; J. D. Keasling, *FEBS J.* **2008**, *275*, 1852.
67. M. Cammareri; M. F. Consiglio; P. Pecchia; G. Corea; V. Lanzotti; J. I. Ibeas; A. Tava; C. Conicella, *Plant Sci.* **2008**, *175*, 255.
68. M. Shibuya; Y. Katsube; M. Otsuka; H. Zhang; P. Tansakul; T. Xiang; Y. Ebizuka, *Plant Physiol. Biochem.* **2009**, *47*, 26.
69. M. J. R. Segura; M. M. Meyer; S. P. T. Matsuda, *Org. Lett.* **2000**, *2*, 2257.
70. M. Shibuya; H. Zhang; A. Endo; K. Shishikura; T. Kushiro; Y. Ebizuka, *Eur. J. Biochem.* **1999**, *266*, 302.
71. H. Hayashi; P. Huang; S. Takada; M. Obinata; K. Inoue; M. Shibuya; Y. Ebizuka, *Biol. Pharm. Bull.* **2004**, *27*, 1086.
72. T. Kushiro; M. Shibuya; Y. Ebizuka, *J. Am. Chem. Soc.* **1999**, *121*, 1208.
73. T. Kushiro; M. Shibuya; Y. Ebizuka, *Tetrahedron Lett.* **1999**, *40*, 5553.
74. T. Kushiro; M. Hoshino; T. Tsutsumi; K. Kawai; M. Shiro; M. Shibuya; Y. Ebizuka, *Org. Lett.* **2006**, *8*, 5589.
75. T. Kushiro; M. Shibuya; K. Masuda; Y. Ebizuka, *J. Am. Chem. Soc.* **2000**, *122*, 6816.
76. M. Basyuni; H. Oku; M. Inafuku; S. Baba; H. Iwasaki; K. Oshiro; T. Okabe; M. Shibuya; Y. Ebizuka, *Phytochemistry* **2006**, *67*, 2517.
77. H. Saimaru; Y. Orihara; P. Tansakul; Y. H. Kang; M. Shibuya; Y. Ebizuka, *Chem. Pharm. Bull.* **2007**, *55*, 784.
78. E. Babiychuk; P. Bouvier-Navé; V. Compagnon; M. Suzuki; T. Muranaka; M. Van Montagu; S. Kushnir; H. Schaller, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3163.
79. T. Kushiro; M. Shibuya; K. Masuda; Y. Ebizuka, *Tetrahedron Lett.* **2000**, *41*, 7705.

80. M. D. Kolesnikova; W. K. Wilson; D. A. Lynch; A. C. Obermeyer; S. P. T. Matsuda, *Org. Lett.* **2007**, *9*, 5223.
81. Y. Ebizuka; Y. Katsube; T. Tsutsumi; T. Kushiro; M. Shibuya, *Pure Appl. Chem.* **2003**, *75*, 369.
82. G. C. Fazio; R. Xu; S. P. T. Matsuda, *J. Am. Chem. Soc.* **2004**, *126*, 5678.
83. T. Xiang; M. Shibuya; Y. Katsube; T. Tsutsumi; M. Otsuka; H. Zhang; K. Masuda; Y. Ebizuka, *Org. Lett.* **2006**, *8*, 2835.
84. M. D. Kolesnikova; A. C. Obermeyer; W. K. Wilson; D. A. Lynch; Q. Xiong; S. P. T. Matsuda, *Org. Lett.* **2007**, *9*, 2183.
85. K. Ohyama; M. Suzuki; J. Kikuchi; K. Saito; T. Muranaka, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 725.
86. Q. Xiong; W. K. Wilson; S. P. T. Matsuda, *Angew. Chem. Int. Ed.* **2006**, *45*, 1285.
87. M. Shibuya; T. Xiang; Y. Katsube; M. Otsuka; H. Zhang; Y. Ebizuka, *J. Am. Chem. Soc.* **2007**, *129*, 1450.
88. T. Akihisa; K. Koike; Y. Kimura; N. Sashida; T. Matsumoto; M. Ukiya; T. Nikaido, *Lipids* **1999**, *34*, 1151.
89. T. Akihisa; Y. Kimura; K. Koike; T. Shibata; Z. Yoshida; T. Nikaido; T. Tamura, *J. Nat. Prod.* **1998**, *61*, 409.
90. T. Akihisa; K. Yasukawa; Y. Kimura; S. Yamanouchi; T. Tamura, *Phytochemistry* **1998**, *48*, 301.
91. T. Akihisa; K. Arai; Y. Kimura; K. Koike; W. C. M. C. Kokke; T. Shibata; T. Nikaido, *J. Nat. Prod.* **1999**, *62*, 265.
92. S. Lodeiro; Q. Xiong; W. K. Wilson; M. D. Kolesnikova; C. S. Onak; S. P. T. Matsuda, *J. Am. Chem. Soc.* **2007**, *129*, 11213.
93. H. Hayashi; P. Huang; K. Inoue; N. Hiraoka; Y. Ikeshiro; K. Yazaki; S. Tanaka; T. Kushiro; M. Shibuya; Y. Ebizuka, *Eur. J. Biochem.* **2001**, *268*, 6311.
94. T. Kushiro; Y. Ohno; M. Shibuya; Y. Ebizuka, *Biol. Pharm. Bull.* **1997**, *20*, 292.
95. S. Shibata; O. Tanaka; J. Shoji; H. Saito, *Economic and Medicinal Plant Research*; Academic Press: London, 1985; Vol. 1, p 217.
96. P. Tansakul; M. Shibuya; T. Kushiro; Y. Ebizuka, *FEBS Lett.* **2006**, *580*, 5143.
97. J. D. White; J. Fayos; J. Clardy, *J. Chem. Soc. Chem. Commun.* **1973**, 357.
98. M. Shibuya; A. Sagara; A. Saitoh; T. Kushiro; Y. Ebizuka, *Org. Lett.* **2008**, *10*, 5071.
99. H. Mitsuguchi; Y. Seshime; I. Fujii; M. Shibuya; Y. Ebizuka; T. Kushiro, *J. Am. Chem. Soc.* **2009**, *131*, 6402.
100. J. Shinozaki; M. Shibuya; K. Masuda; Y. Ebizuka, *Phytochemistry* **2008**, *69*, 2559.
101. M. Shibuya; M. Hoshino; Y. Katsube; H. Hayashi; T. Kushiro; Y. Ebizuka, *FEBS J.* **2006**, *273*, 948.
102. H. Seki; K. Ohyama; S. Sawai; M. Mizutani; T. Ohnishi; H. Sudo; T. Akashi; T. Aoki; K. Saito; T. Muranaka, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 14204.
103. X. Qi; S. Bakht; B. Qin; M. Leggett; A. Hemmings; F. Mellon; J. Eagles; D. Werck-Reichhart; H. Schaller; A. Lesot; R. Melton; A. Osbourn, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18848.
104. B. Field; A. E. Osbourn, *Science* **2008**, *320*, 543.

Biographical Sketches

Tetsuo Kushiro received his B.S. in chemistry from the Tokyo Institute of Technology (1993) and Ph.D. from the Graduate School of Pharmaceutical Sciences, The University of Tokyo (1998), where he studied with Professor Yutaka Ebizuka. He carried out JSPS postdoctoral research at the same laboratory (1998–2000) and then at The Scripps Research Institute with Professor Paul Schimmel (2000–02). He became research associate at Plant Science Center, RIKEN Institute, working with Dr. Yuji Kamiya (2002–05). He was then appointed assistant professor at The University of Tokyo in 2005. His current research includes biosynthesis of terpenoids in both plants and fungi, especially triterpenoids and meroterpenoids.

Yutaka Ebizuka is a Professor at the Graduate School of Pharmaceutical Sciences, The University of Tokyo in Japan. He received his Ph.D. from the Graduate School of Pharmaceutical Sciences, The University of Tokyo in 1974. After 2-year postdoctoral experience in Department of Chemistry and Department of Botany at the University of British Columbia in Vancouver, he returned to The University of Tokyo as an assistant professor. He has been Professor of Natural Products Chemistry since 1995. He has received several distinctions, which include the awards from the Japanese Society for Plant Cell and Molecular Biology (2001), Japanese Society of Pharmacognosy (2006), and the Pharmaceutical Society of Japan (2007). His research interests include biosynthetic studies of natural products from polyketide and terpenoid pathways.

1.19 Bacterial Squalene Cyclase

Ikuro Abe, The University of Tokyo, Hongo, Tokyo, Japan

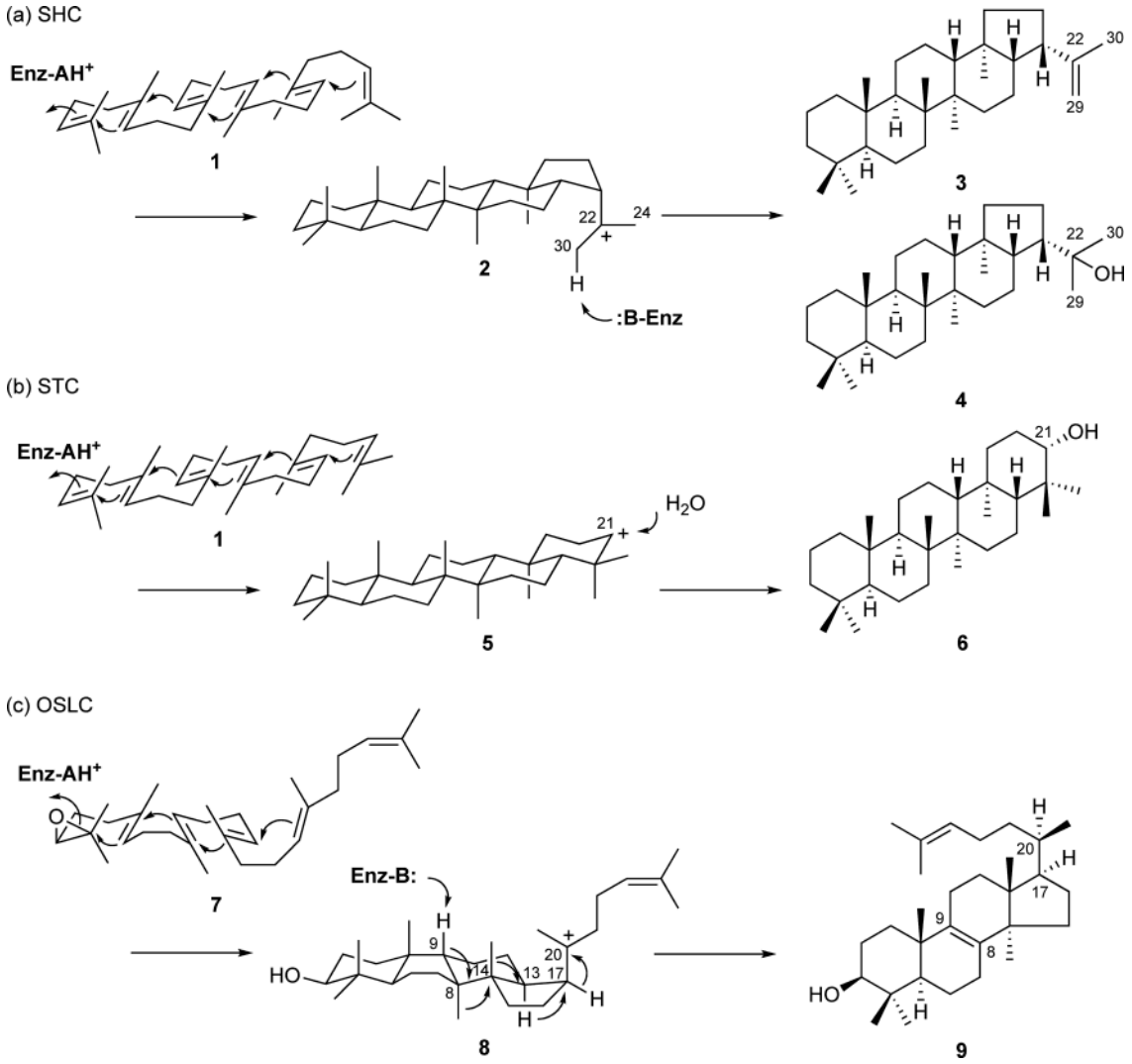
© 2010 Elsevier Ltd. All rights reserved.

1.19.1	**Introduction**	709
1.19.2	**The Reaction Mechanism**	711
1.19.3	**The Enzyme Structure**	712
1.19.4	**Site-Directed Mutagenesis**	716
1.19.4.1	Residues for Initiation and Termination	716
1.19.4.2	Residues for Cation Stabilization	717
1.19.5	**Substrate Recognition and Catalytic Potential**	719
1.19.5.1	Squalene and Oxidosqualene	719
1.19.5.2	Dihydrosqualene	720
1.19.5.3	Analogues with Various Chain Lengths	720
1.19.5.4	Heteroaromatic Ring-Containing Analogues	722
1.19.5.5	Methylidene-Extended Analogues	724
1.19.5.6	Fluorine and Sulfur Analogues	726
1.19.5.7	Desmethylsqualenes and Other Analogues	727
1.19.6	**Conclusions**	729
References		730

1.19.1 Introduction

In bacteria, sterols are usually absent, but some bacteria and protozoans produce pentacylic triterpenes, which are regarded as sterol surrogates in these organisms.[1–3] In this case, squalene (**1**) is cyclized directly (without prior epoxidation) to pentacyclic 3-deoxytriterpenes with the hopane and gammacerane skeletons such as hopene (**3**) and tetrahymanol (**6**). Squalene:hopene cyclase (SHC) from the thermoacidophilic bacterium *Alicyclobacillus acidocaldarius* has been well studied, and it was established that the enzyme binds squalene (**1**) in all-*chair* conformation, initiating the sequential carbon–carbon bonds forming reaction by an acid-catalyzed protonation of a terminal double bond. The polycyclization first produces the hopanyl C-22 carbocation (**2**) with a five-membered E-ring, which is followed by either regiospecific proton elimination from the terminal *Z*-methyl (Me-30) group or addition of H_2O at the cationic center yielding hop-22(29)-ene (**3**) (80%) or hopan-22-ol (**4**) (20%) (**Scheme 1(a)**).[4–8] On the other hand, the formation of tetrahymanol (**6**) by squalene:tetrahymanol cyclase (STC) from the protozoan *Tetrahymena pyriformis* is initiated by a proton attack on a terminal double bond of squalene, folded in all-*chair* conformation, but followed by addition of H_2O to the resulting gammaceranyl C-21 cation (**5**) (**Scheme 1(b)**).[9] No methyl or hydride migration takes place during the sequential ring-forming reactions. In contrast, in animals and fungi, eukaryotic oxidosqualene:lanosterol cyclase (OSLC) catalyzes mechanistically more complex reaction.[4,6,7,10–13] Thus, the formation of lanosterol (**9**), which is the biosynthetic precursor of cholesterol, is initiated by an oxirane ring opening of (3*S*)-2,3-oxidosqualene (**7**) folded in *chair–boat–chair* conformation. The cyclization reaction first produces the 6.6.6.5-fused tetracyclic protosteryl C-20 cation (**8**) with a 17β-side chain, which then undergoes transformation into the final product through a series of 1,2-methyl and hydride shifts (H-17$\alpha \rightarrow$ 20, H-13$\alpha \rightarrow$ 17α, CH_3-14$\beta \rightarrow$ 13β, CH_3-8$\alpha \rightarrow$ 14α) followed by elimination of H-9 to generate the Δ^8 double bond (**Scheme 1(c)**).

The squalene cyclases (SCs) and oxidosqualene cyclases (OSCs) thus catalyze the production and stabilization of polycyclic carbocations and direct the enzyme-templated formation of new carbon–carbon bonds in a regio- and stereochemically defined manner. During the biosynthesis of the hopane skeleton, 13 covalent bonds are broken or formed, and 9 chiral centers are constructed in a single reaction. The bacterial SCs have been



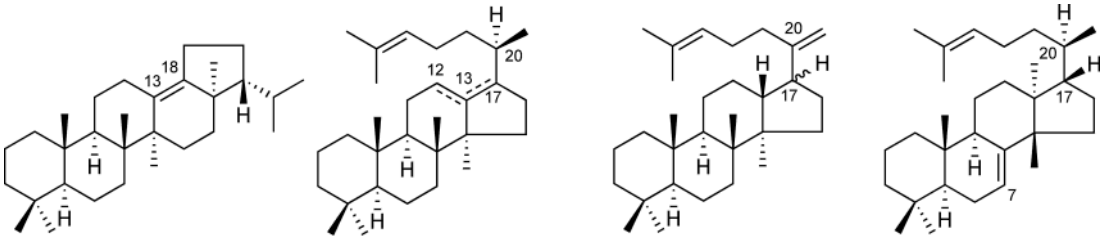
Scheme 1 Proposed mechanism for (a) squalene:hopene cyclase (SHC), (b) squalene:tetrahymanol cyclase (STC), and (c) oxidosqualene:lanosterol cyclase (OSLC).

considered to be rather primitive compared to the eukaryotic OSCs that catalyze mechanistically and conformationally more complex ring-forming reactions.[1,2,14] First, the direct cyclization of squalene in bacteria is an anaerobic process, while, in eukaryotes, squalene is first oxidized by squalene epoxidase using molecular oxygen. The folding of the squalene molecule in less strained all-*chair* conformation would require minimal enzymatic assistance in the aqueous environment of a cell. Further, the cyclization into hopene or tetrahymanol proceeds without carbon skeletal rearrangement, whereas eukaryotic OSCs have to direct and control the apparently complex rearrangement reactions. Moreover, the bacterial SCs exhibit remarkable substrate tolerance and catalytic plasticity, for example, they can cyclize not only squalene, the natural substrate, but also both enantiomers of oxidosqualene. This is in sharp contrast to the eukaryotic OSCs that have a rigorous substrate specificity as they do not accept intact squalene or (3*R*)-2,3-oxidosqualene, but cyclize (3*S*)-2,3-oxidosqualene specifically. Finally, on the basis of crystal structures of *A. acidocaldarius* SHC, the hydrophobic active site cavity appears to have enough space to accept bulky substrate analogues.[15] These remarkable properties of the bacterial SCs have enabled the recent precursor-directed and structure-based engineered biosynthesis of unnatural novel triterpene scaffolds.[7]

SC and OSC enzymes convert the C_{30} acyclic polyenes into various skeletal types of cyclic triterpenes; nearly 200 different skeletons are known from natural sources or enzymatic reactions.[16] The relationships among enzyme structure, the cyclization mechanism, and the products formed are of great interest. Earlier, Ourisson and coworkers proposed a hypothesis that the primitive bacterial SCs are evolved from the eukaryotic OSCs.[1,2,14] In principle, only small modifications of the active site geometry of the enzyme could lead to the formation of dramatically different cyclization products. It is now accepted that the enzymes initiate the cyclization cascade by presenting a general acid to protonate the terminal double bond or oxirane ring of the substrate with participation by a neighboring π-bond. The enzymes also provide a template that chaperones the flexible substrate and partially cyclized intermediate carbocations through a series of precise conformations leading to one specific cyclization product, by shielding the cationic intermediates from premature addition of water or elimination.[6] Further, the sequential ring-forming reactions are accelerated by stabilization of the intermediate cations by electron-rich environment.[6] Thus, the hydrophobic active site lined with conserved aromatic residues would have a molecular geometry that allows cyclization to proceed with a minimum of conformational changes in the substrate once the reaction is initiated. In addition, other steric and electrostatic factors, such as cation–π interactions involving the incipient carbocation and the electron-rich faces of aromatic side chains of Phe, Tyr, or Trp, could also shelter transient cationic species in analogy to known receptor–ligand and enzyme–substrate complexes.[17,18] Finally, recent advances on crystallographic and structure-based mutagenesis studies as well as utilization of synthetic active site-targeted probes have begun to reveal intimate structural details of the catalytic mechanism of the enzyme-templated sequential ring-forming reactions.[7,8]

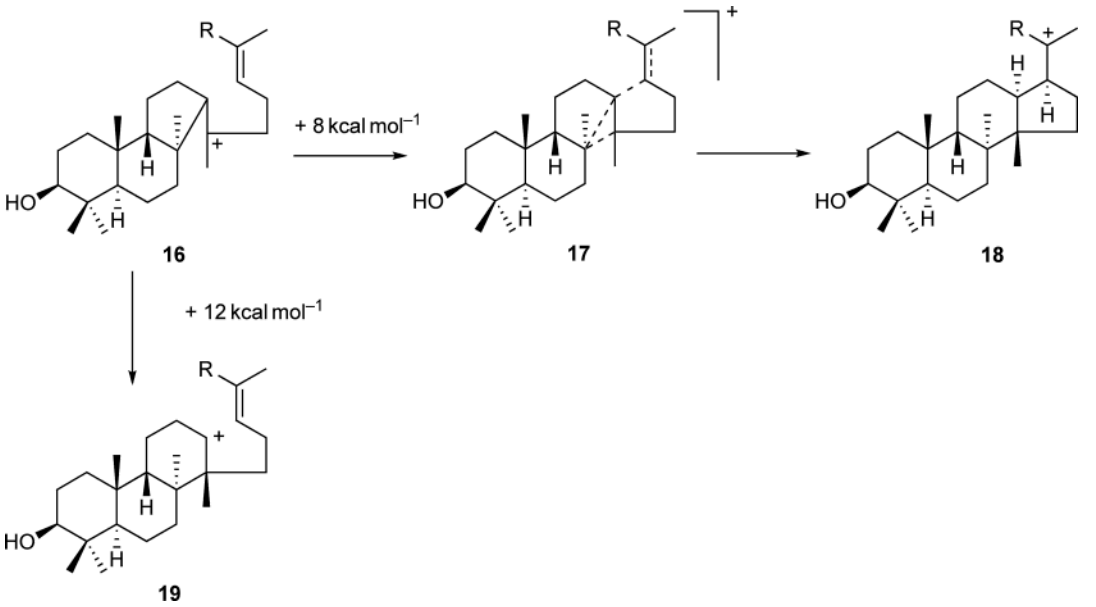
1.19.2 The Reaction Mechanism

The remarkable cyclization of squalene and oxidosqualene has fascinated chemists for more than half a century.[19–26] According to the biogenetic isoprene rule proposed by Echenmoser, Ruzicka, Jeger, and Arigoni, the cyclization of all *trans* squalene takes place through a well-defined sequence of pre-*chair* and pre-*boat* conformations.[20] The transformation was proposed to proceed according to the rules of *anti*-periplanar cationic 1,2-addition, 1,2-rearrangement, and 1,2-elimination.[20] For entropic reasons and on the basis of the experimental evidence, van Tamelen proposed that the enzyme-templated sequential carbon–carbon bond formation reactions proceed through a series of discrete, conformationally rigid, partially cyclized carbocationic intermediates.[27] Indeed, monocyclic, bicyclic, tricyclic, and tetracyclic skeletons corresponding to all possible cationic intermediates of a possible stepwise cyclization process have been isolated from natural sources or obtained as products of many of the site-directed mutants and the enzyme reactions with non-physiological substrates. Further, it has been reported that *A. acidocaldarius* SHC enzyme reaction *in vitro* afforded several minor hydrocarbons (**10–15**) as cyclization by-products of squalene, supporting the presence of tetracyclic and pentacyclic intermediate carbocations.[28]



10 **11**: $\Delta^{12(13)}$ **12**: $\Delta^{13(17)}$ **13**: 17β-H **14**: 17α-H **15**

During the enzymatic cyclization of squalene and oxidosqualene, the six-membered C-ring is formed through a thermodynamically unfavorable 6.6.6-fused tricyclic secondary cation (**19**) in a formally anti-Markovnikov addition. After the formation of a 6.6 bicyclic A/B-ring system, the six-membered C-ring is thought to be generated by initial ring closure to a 6.6.5-fused tricyclic Markovnikov tertiary cation with a



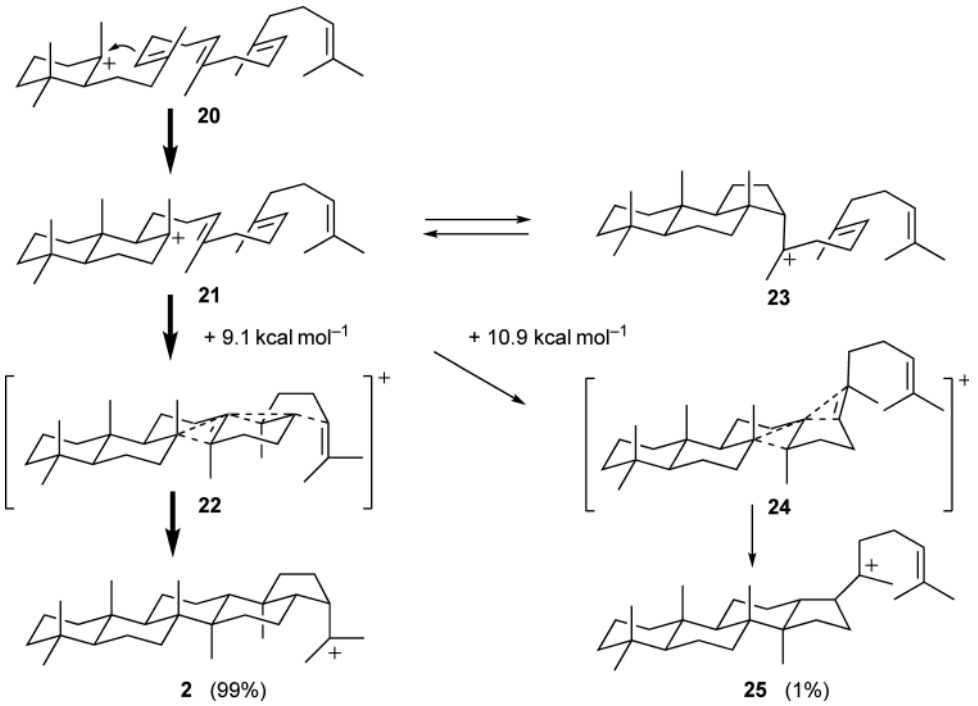
Scheme 2 Proposed mechanism for concerted C-ring expansion and D-ring formation by Hess.

five-membered C-ring (**16**), which is then followed by ring expansion to a six-membered C-ring to generate the 6.6.6-fused tricyclic anti-Markovnikov secondary cation (**19**).[13] However, it is still a matter for debate how does the C-ring construction overcome the energy barrier required to expand the tertiary cation to the thermodynamically less stable secondary carbocation.[29] Previously, Johnson proposed that axial delivery of negative point charges by the enzyme could stabilize the developing cationic centers as ion pairs, and that such a charge delivery is important in enhancing the rate and efficiency of the overall cyclization process.[30,31] On the other hand, Jorgensen suggested that this barrier may be lowered by selective replacement of nucleophilic groups from the protein backbone or side chains including aromatic amino acid residues through cation–π interactions as proposed by Dougherty.[17,32] Indeed, the X-ray crystal structure of the enzyme revealed that the active site cavity is lined with aromatic amino acid residues, which could play a crucial role in the sequential ring-forming reaction.[15]

On the basis of computational studies, Hess proposed that conversion of the 6.6.5-fused tricyclic Markovnikov tertiary cation **16** to **19** might be instead involved in the expansion of the C-ring in concert with the formation of five-membered D-ring in **18**, through transition state structure **17** (**Scheme 2**).[29,33] In addition, he also pointed out that a concerted A/B-ring closure is likely to occur in the cyclization of squalene to hopene.[34] On the other hand, Rajamani and Gao proposed that the 6.6 bicyclic tertiary cation (**21**) is the only reaction intermediate with a significant lifetime, and then formation of rings C, D, and E takes place through a single transition state to produce a 6.6.6.6.5-fused pentacyclic intermediate cation (**2**) (99%) along with a 6.6.6.5-fused tetracyclic tertiary cation by-product (**25**) (1%) (**Scheme 3**).[35] However, this proposal seems to be in contradiction to a recent structure-based proposal by Schluz that a 6.6.6.5-fused tetracyclic intermediate tertiary cation is stabilized by the active site aromatic residues and should have relatively long lifetime.[36]

1.19.3 The Enzyme Structure

In the early 1990s, Poralla and coworkers first reported enzyme purification, cloning, sequencing, and heterologous expression of SHC from a thermoacidophilic bacteria *A. acidocaldarius*.[37,38] The enzyme is a membrane-associated 71 524 Da protein with 631 amino acids, sharing 25% amino acid sequence identity with rat liver OSLC,[39] which is a 83 321 Da protein with 733 amino acids. The enzyme shows its catalytic optimum at 60 °C



Scheme 3 Proposed mechanism for the formation of hopene by Rajamani and Gao.

and pH 6.0. The X-ray crystal structure of *A. acidocaldarius* SHC was first solved by Schulz and coworkers at 2.9 Å resolution in 1997[15] and later refined to 2.0 Å resolution in 1999.[40] The crystal structures revealed that the homodimeric monotopic membrane-bound enzyme has an dumbbell-shaped (α/α) barrel domains connected by long loops, which together enclose a large central cavity of 1200 Å^3 (**Figure 1**). Importantly, recently reported X-ray crystal structure of human OSLC also showed a similar three-dimensional overall structure with identical topology to that observed in *A. acidocaldarius* SHC despite the low sequence identity.[41,42] The active site cavity of OSLC appears to be a little smaller than that of SHC, which has to accommodate the additional E-ring. Interestingly, *A. acidocaldarius* SHC and human OSLC contain eight and six repeats, respectively, of a highly conserved β-strand-turn motif rich in aromatic amino acids ([K/R][G/A]XX[F/Y/W][L/I/V]XXXQXXXGXW).[43] The repetitive QW motif (Q is glutamine and W is tryptophan) is thought to stabilize the enzyme structure by connecting surface α-helices during the highly exergonic cyclization reaction.[15]

The active site of *A. acidocaldarius* SHC is located in the large central cavity, consisting of an extended hydrophobic section lined with conserved aromatic amino acid residues (**Figure 1**).[15,40] As discussed, the π-electrons of these residues would stabilize the partially cyclized cationic intermediate species during the sequential ring-forming reaction. The top and bottom sections of the cavity are formed by polar hydrogen networks around Asp376 and Glu45, which are thought to play crucial roles in the initial protonation and in the final deprotonation reaction, respectively. Thus, Asp376 is hydrogen bonded to His451 and functions as the general acid that donates a proton to the terminal double bond of the substrate to initiate the cyclization cascade. After the protonation, Asp376 is most likely to be stabilized by the positively charged His451 that in turn is electrostatically stabilized by the solvent-accessible Glu454.[36] On the other hand, Glu45 at the bottom of the active site cavity is thought to be important in the termination of the reaction cascade. It has been proposed that the water molecules are polarized by residues in the hydrogen-bonding network around Glu45, and this polarized water molecule eliminates proton from the terminal methyl (Me-30) group or attacks the cationic center of the hopanyl C-22 carbocation.

Another recently solved crystal structure of *A. acidocaldarius* SHC complexed with the inhibitor 2-azasqualene at 2.1 Å resolution clearly demonstrated that the substrate analogue binds with a conformation very close

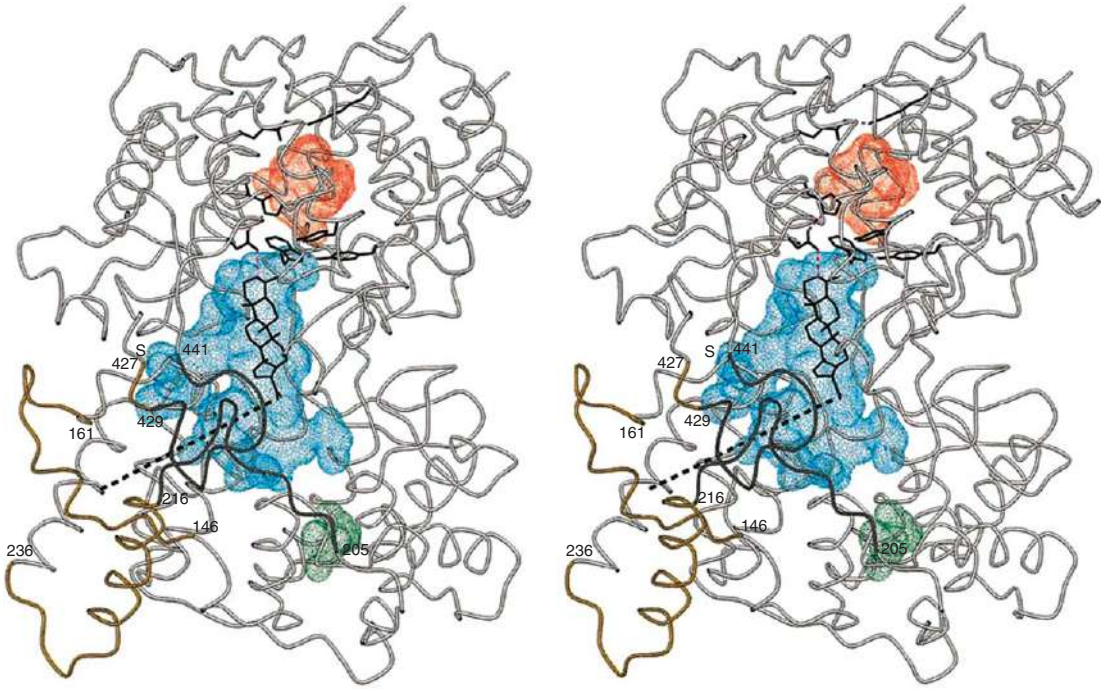


Figure 1 Stereoview of the Cα chain trace of *A. acidocaldarius* SHC together with the three largest cavities within the structure: the active center cavity (blue), the upper cavity (red), and the lower cavity (green). The figure contains a modeled hopene molecule in the active center cavity and the residues separating it from the upper cavity. The chain of the membrane-binding region is yellow. The mobile loops that presumably open up for product release are dark. The entrance channel is indicated by a broken line. Reprinted with permission from K. U. Wendt; A. Lenhart; G. E. Schulz, *J. Mol. Biol.* **1999**, *286*, 175. Copyright 1999 Academic Press.

to that required for the *chair* conformations of hopene rings A through D by forming a salt bridge to Asp376 (**Figure 2**).[36] This suggested that the bound squalene indeed undergoes only small conformational changes during the formation of rings A through D. On the basis of the crystal structure, Schluz proposed that the cyclization cascade up to the 6.6.6.5-fused tetracyclic intermediate cation proceeds almost concomitantly without significant conformational changes.[36] Thus, after the initial acid-catalyzed protonation of the terminal double bond, a Markovnikov-type cation-olefin addition first produces the 6.6-fused bicyclic tertiary cation, stabilized by the π-electrons and dipoles of aromatic residues including Phe365, Tyr420, Trp489, and Tyr609. Then, a six-membered C-ring is formed directly in concert with the formation of a Markovnikov-type five-membered D-ring, generating a 6.6.6.5-fused tetracyclic intermediate tertiary cation, stabilized by Phe601, Phe605, and Trp169 through cation–π interactions. The cyclization cascade is followed by D-ring expansion and Markovnikov E-ring closure, yielding the 6.6.6.6.5-fused hopanyl C-22 tertiary cation with the pentacyclic ring system, which is likely to be stabilized by the conserved Phe605 π-electrons.[40] Finally, as described, the water molecule polarized by residues in the hydrogen-bonding network around Glu45 eliminates proton from the methyl group or attacks the cationic center of the hopanyl carbocation to terminate the reaction.

On the other hand, substrate uptake and product release are thought to occur through a nonpolar channel connecting the active site cavity to the membrane-immersed region of the enzyme and thus to the membrane interior (**Figure 1**).[36] However, in the crystal structures, this channel appears to be too narrow for the bulky hopene products. Based on the observation that the wall between channel and dimer surface is rather mobile, it has been postulated that the channel cross section can be appreciably enlarged by melting and displacing the mobile peptide. This seems likely since the thermoacidophilic enzyme reaction is optimal at 60 °C, and the cyclization reaction itself is highly exergonic (about 200 kJ mol^{-1}).[36]

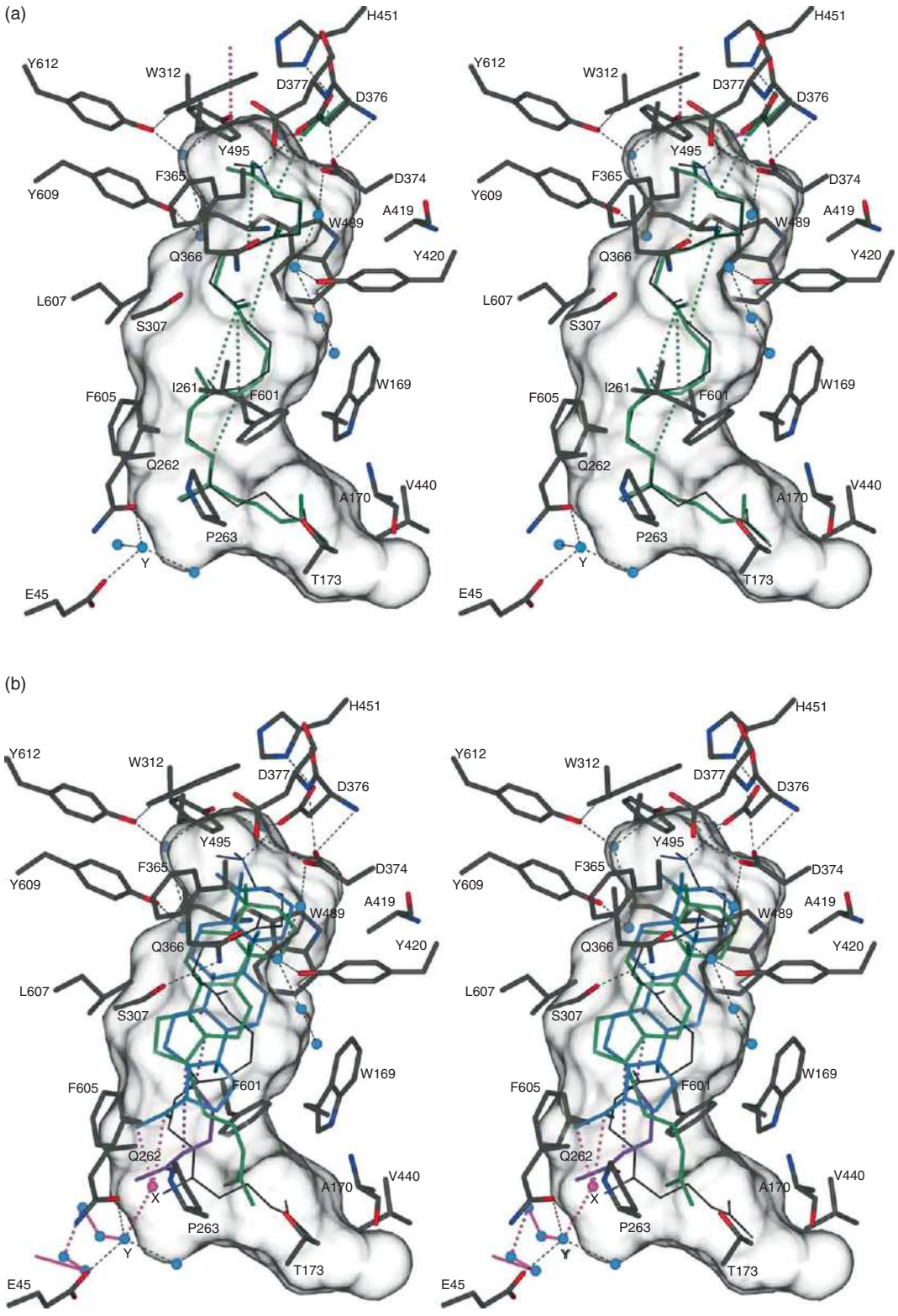
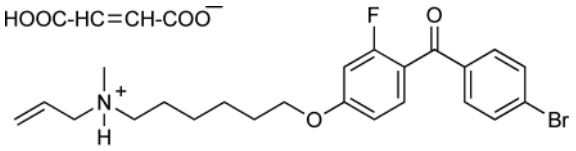


Figure 2 Stereoview of the active site of *A. acidocaldarius* SHC. (a) Squalene model (green) indicating the first four carbocation additions to double bonds (green dots). The observed 2-azasqualene structure (thin dark lines) is given for reference. (b) Models of the tetracycle (green) and the hopenyl cation (blue). Reprinted with permission from D. J. Reinert; G. Balliano; G. E. Schulz, *Chem. Biol.* **2004**, *11*, 121. Copyright 2004 Elsevier Science Ltd.



26

Recently, crystal structures of *A. acidocaldarius* SHC[44] and human OSLC[41] complexed with the potential anticholesteremic drug Ro48-8071 (**26**) have been reported by the Schulz group. The benzophenone-containing Ro48-8071 is a potent inhibitor of both the bacterial SHC ($IC_{50} = 9.0\ nmol\ l^{-1}$)[45] and the vertebrate OSLC ($IC_{50} = 6.5\ nmol\ l^{-1}$).[46] Further, Ro48-8071 blocked cholesterol biosynthesis *in vivo* by selective inhibition of OSLC in HepG2 cells when administered orally in hamsters, squirrel monkeys, and minipigs.[47,48] The crystal structure at 2.8 Å resolution demonstrated that Ro48-8071 is bound in the central cavity of the active site of SHC and extends into the channel that connects the cavity with the membrane (**Figure 3**).[44] The binding site is largely identical with the expected binding mode of squalene but differs from a previously reported model based on photoaffinity labeling experiment.[49] The inhibitor position appears to be dominated by the interactions of its protonated amino group, with the amino acid residues catalyzing the protonation of squalene. Further stabilization is achieved by cation–π and CH–π interactions. In comparison with the hopene model, however, the carbonyl carbon of the inhibitor is 2.5 Å distant from the final carbocation position of the E-ring. This suggested that during the cyclization reaction the partially cyclized product may be shifted along the main axis of the active site cavity so that the final carbocation position may move farther away from the initiation site.[44] The knowledge of the inhibitor-binding mode in the enzymes may help develop more potent and efficient enzyme inhibitors of SHC and OSLC.

1.19.4 Site-Directed Mutagenesis

1.19.4.1 Residues for Initiation and Termination

As described, the conserved Asp376 in *A. acidocaldarius* SHC is thought to be the general acid that initiates the cyclization cascade by protonating the terminal double bond. Earlier site-directed mutagenesis studies by Poralla and coworkers established that the active site residue Asp376 of the conserved DDTAVV motif of

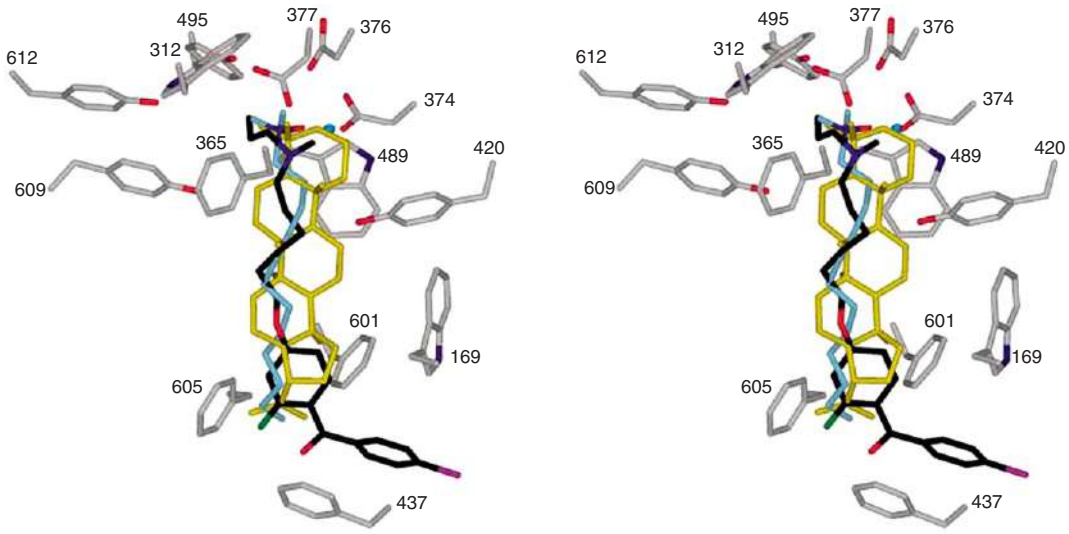


Figure 3 Stereoview of the binding mode of Ro48-8071 at *A. acidocaldarius* SHC. Superposition of Ro48-8071 (dark blue), a modeled hopene molecule (yellow), and the detergent LDAO (sky blue) as bound to SHC. Reprinted with permission from A. Lenhart; W. A. Weihofen; A. E. Pleschke; G. E. Schulz, *Chem. Biol.* **2002**, *9*, 639. Copyright 2002 Elsevier Science Ltd.

A. acidocaldarius SHC is essential for the catalytic activity.[50] It was demonstrated that, with the exception of D376E mutant, all other substitutions of Asp376, D376Q, D376G, and D376R resulted in almost or complete loss of enzyme activity. Kinetic analysis revealed that the specific activity of the D376E mutant was reduced to 10%, accompanied by a significant decrease in the apparent V_{max}, whereas the apparent K_m remained unchanged. On the other hand, Dang and Prestwich reported *A. acidocaldarius* SHC mutant in which the DDTAVV motif is changed to DCTAEA, as in the case of the corresponding conserved catalytic region of eukaryotic OSLC, by a triple mutation of D377C/V380E/V381A.[51] Remarkably, the triple mutant no longer accepted squalene as a substrate and completely lost the hopene-forming activity; however, the mutant still accepted both enantiomers of 2,3-oxidosqualene to produce a mixture of 3-hydroxyhopene as in the case of wild-type SHC.[1]

Glu45 at the bottom of the active site cavity has been postulated to be the crucial residue for the termination of the polycyclization reaction. Prestwich and coworkers reported that the site-directed mutants of Glu45, E45A, and E45D showed reduced activity, whereas E45Q showed slightly increased activity, and E45K was inactive.[51] A normal yield of pentacyclic products was obtained; however, the ratio of hopene to hopanol was significantly changed in the less active mutants. The decrease of cyclization activity for E45D and E45A, and the loss of the activity for E45K, confirmed a role for Glu45 in the cyclization process. The result is consistent with the proposal that a hydrogen-bonding network of Gln262:Glu45:Glu93:Arg127 involving also a water molecule may stabilize or quench the C-22 carbocation. The enhanced activity of E45Q is not as readily explained. Although Glu45 is important in stabilizing the terminal carbocation, it may not be the only residue involved in the termination of the reaction.

1.19.4.2 Residues for Cation Stabilization

Following the pioneering work by Poralla, Prestwich, and others, Hoshino and coworkers carried out extensive site-directed mutagenesis studies on *A. acidocaldarius* SHC, which includes point mutants of conserved active site aromatic residues Trp169,[52] Trp312,[52] Phe365,[53,54] Tyr420,[55–57] Trp489,[52] Phe601,[55,58] Phe605,[59,60] Tyr609,[54,61,62] and Tyr612[54] as well as other active site residues Ile261,[63] Gln262,[64] Pro263,[64] Asp377,[54] Gly600,[65] and Leu607.[57] Since the results were excellently summarized in a review article,[8] only a brief illustration of the structures of the unnatural triterpenes (**27–55**) produced by the site-directed mutants (**Figure 4**) is provided in this chapter. In general, almost all of the mutants exhibited altered product specificities, and in most of the cases, the point mutations resulted in the early truncation of the polycyclization cascade, resulting in the formation of monocyclic, bicyclic, tricyclic, and tetracyclic by-products. It was thus demonstrated that subtle changes in the catalytic environment significantly affect the stereoelectronical course of the cyclization reaction, thereby leading to the generation of the structurally diverse unnatural novel triterpene scaffolds. In some cases, the aberrant cyclization products whose stereochemistry is opposite to that of the normal cyclization intermediates were obtained. This also suggested that the stereochemical destiny during the polycyclization cascade is directed by the steric bulk size of the active site residues.

As described, the π-electrons of the conserved aromatic residues of Phe, Tyr, or Trp lining the catalytic site are thought to be important for the stabilization of the partially cyclized cationic intermediates during the sequential ring-forming reaction. However, the repeated observation of partially cyclized products arising from many of the site-directed mutants reflects the general problem of separating specific cation–π interactions from simple steric effects that may arise from the displacement of the forming product in the active site or from subtle changes in the active site structure.[6] To provide solid experimental evidence for the cation–π interactions, Hoshino and coworkers recently constructed *A. acidocaldarius* SHC mutants in which the active site aromatic residues Phe365 and Phe605 are substituted with unnatural amino acids: mono-, di-, or trifluorophenylalanines.[59] Since fluorine atom has an extremely higher electronegativity but a similar van der Waals radius compared to that of a hydrogen atom, these unnatural amino acids are thought to reduce cation–π binding energies without causing a structural alteration of the cyclase enzyme.[59] As anticipated, kinetic analyses demonstrated that the enzyme activities of the fluorophenylalanine mutants are inversely proportional to the number of the fluorine atoms on the aromatic ring and correlate with the cation–π binding energies of the ring moiety. Importantly, no serious structural perturbation was observed for the mutants even at high temperature. It was thus confirmed that the π-electron density of the residues 365 and 605 is indeed crucial for the sequential ring-forming reactions. On the other hand, the

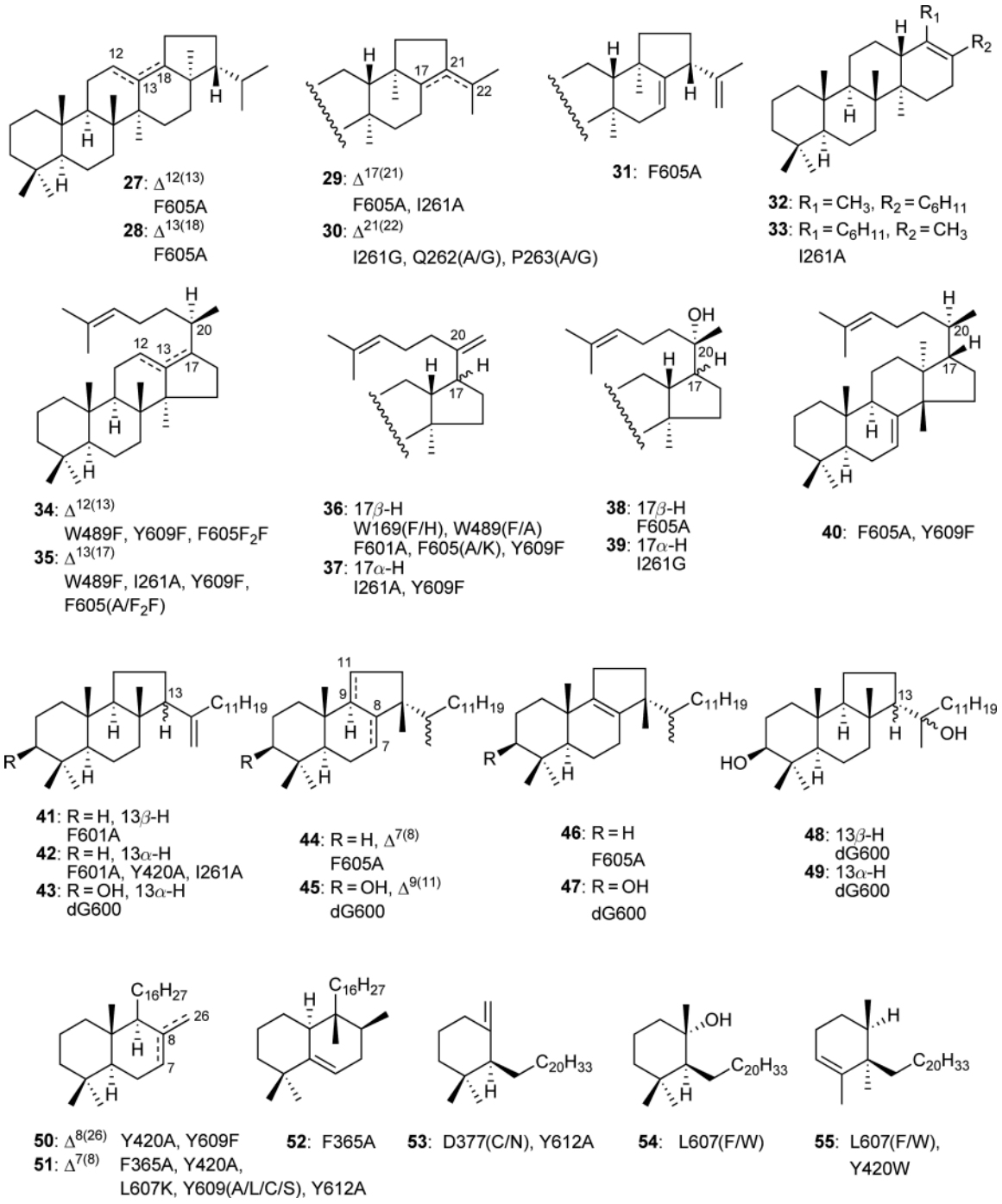


Figure 4 Structures of unnatural cyclic triterpenes produced by mutants of *A. acidocaldarius* SHC. dG600; deletion mutant of Gly600, $F605F_2F$; a point mutant in which Phe605 is substituted with difluorophenylalanine.

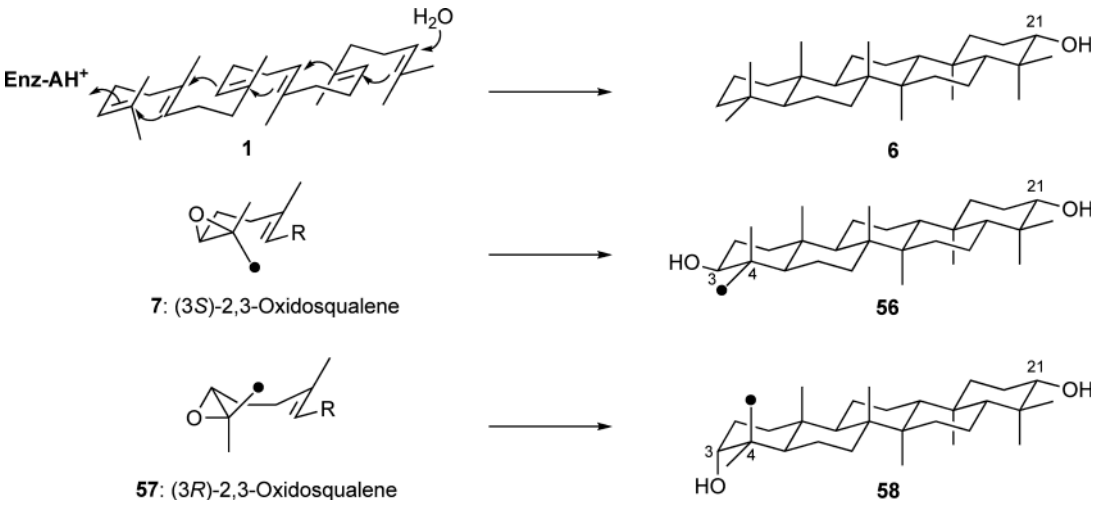
fluorophenylalanine mutants of Phe605 produced Markovnikov 6.6.6.5-fused tetracyclic by-products (20*R*)-17-*epi*-dammar-12(13)-ene (**34**) and (20*R*)-13(17)-dammarene (**35**) (**Figure 4**), in addition to hopene and hopanol.[59] This result also confirmed that the π-electron density of Phe605 is important for the anti-Markovnikov D-ring expansion and E-ring formation.

1.19.5 Substrate Recognition and Catalytic Potential

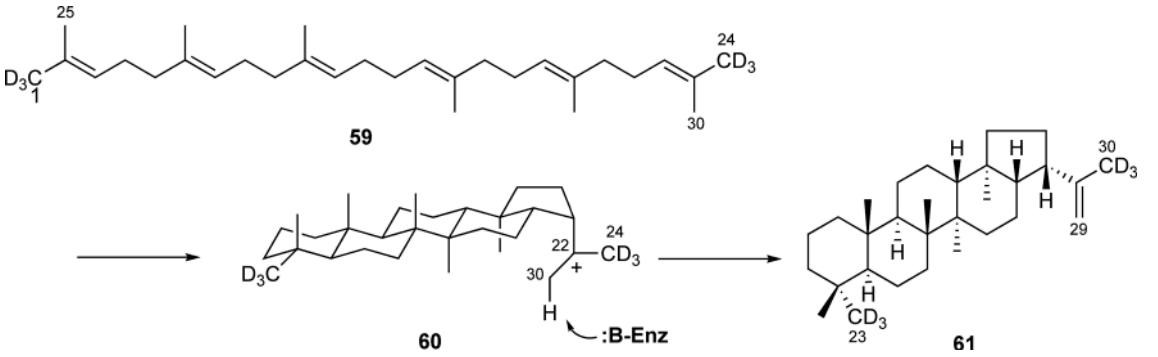
The substrate specificities of bacterial SCs have been probed with a variety of substrate analogues. Interestingly, the enzymes, catalyzing a stereochemically and mechanistically simpler process than eukaryotic OSCs, exhibit broad substrate tolerance and remarkable catalytic plasticity.[1,4,7] The bacterial SCs normally catalyzing the formation of a pentacyclic ring system without carbon skeletal rearrangement can also accept a variety of nonphysiological substrates and efficiently perform apparently complex cyclization reactions to produce structurally distinct unnatural novel polyprenoid scaffolds.[7] The catalytic potential and versatility are of great interest from the point of view of the molecular evolution of the cyclase enzymes. On the other hand, the formation of novel and unexpected products by the bacterial SCs also suggested that the active sites are simply permissive and therefore do not intervene completely in the folding and cyclization of substrates, which they do not normally encounter. Notably, no flexible motion of the folded conformation seems to be allowed inside the reaction cavity once the cyclization cascade has started. The role of the enzyme is most likely to involve the initiation of cyclization and prevents alternative modes of cyclization reactions.

1.19.5.1 Squalene and Oxidosqualene

Earlier studies by Ourisson, Rohmer, and coworkers demonstrated that *T. pyriformis* STC, normally catalyzing the formation of tetrahymanol (**6**) from squalene, also cyclized (3*R*)-oxidosqualene (**57**) into gammacerane-3α,21α-diol (**58**), whereas (3*S*)-oxidosqualene (**7**) was cyclized into gammacerane-3β,21α-diol (**56**) (**Scheme 4**).[66] Moreover, the enzyme reaction with (1,1,1,24,24,24-$^2H_6$)2,3-oxidosqualene, in which the 1-C$^2H_3$ and the 3-H were in *cis* position, demonstrated that during the cyclizations of both enantiomers of 2,3-oxidosqualene, the labeled methyl group completely maintained its stereochemical integrity. Thus, in gammacerane-3β,21α-diol, the equatorial methyl group at C-4α was labeled, whereas in gammacerane-3α,21α-diol, the axial methyl group at C-4β was labeled. This suggested that (3*S*)-oxidosqualene is cyclized in all pre-*chair* conformations, whereas its (3*R*)-enantiomer is cyclized in a pre-*boat* conformation for ring A (**Scheme 4**). This conformational versatility is characteristic of bacterial SCs, but has been lost in the eukaryotic OSCs. In a similar manner, SHC of *A. acidocaldarius* and *Acetobacter pasteurianum* also showed the same lack of substrate specificity and yielded the 3β-hydroxyhopene and 3α-hydroxyhopene from (3*S*) and (3*R*)-2,3-oxidosqualene, respectively.[67,68] On the other hand, cyclization of (1,1,1,24,24,24-$^2H_6$)squalene (**59**) into (23,23,23,30,30,30-$^2H_6$)hop-22(29)-ene (**61**) confirmed that the final proton elimination takes place regiospecifically from the *Z*-methyl (Me-30) group during the biosynthesis of hopene (**Scheme 5**).[69,70]



Scheme 4 Enzymatic cyclization of (3*S*)- and (3*R*)-2,3-oxidosqualene by *T. pyriforms* STC.



Scheme 5 Enzymatic cyclization of (1,1,1,24,24,24-$^2H_6$)squalene by *A. acidocaldarius* SHC.

1.19.5.2 Dihydrosqualene

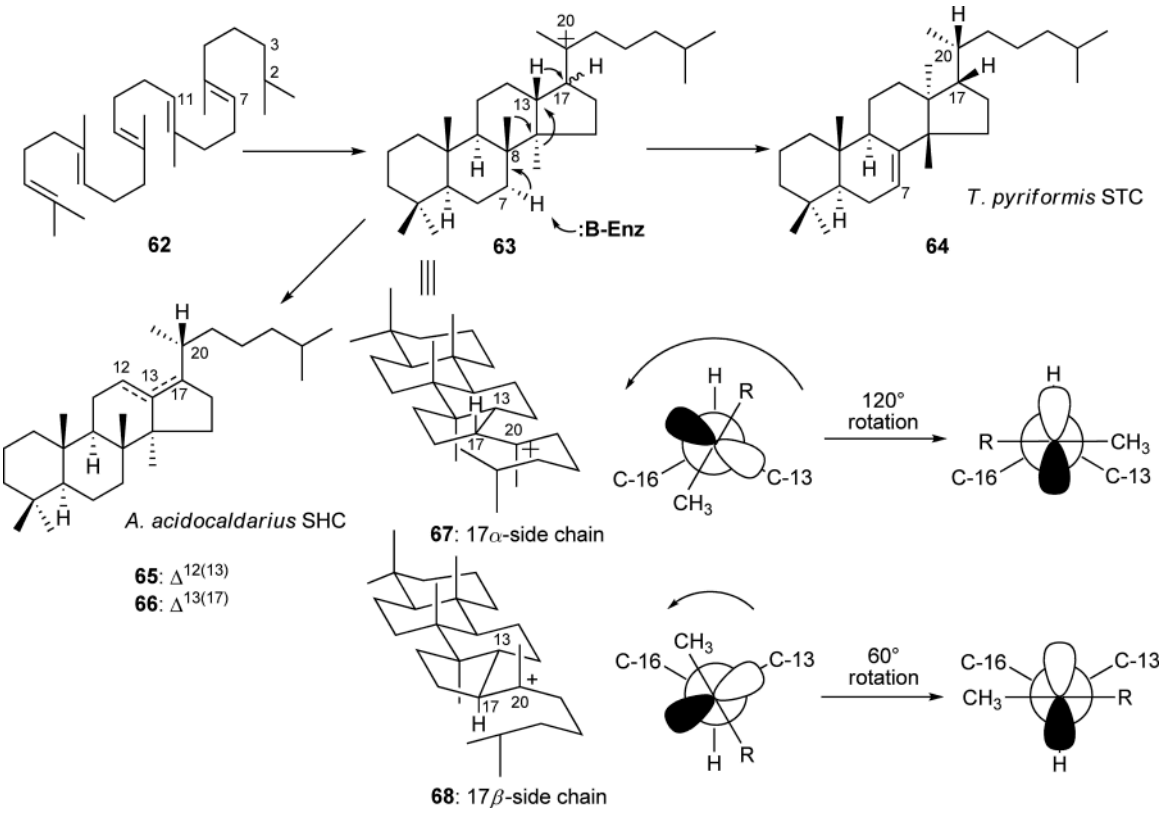
2,3-Dihydrosqualene (**62**) is a molecule lacking one of the terminal double bond of squalenes; therefore, it is no longer possible to produce the pentacyclic ring system. It has been reported that *T. pyriformis* STC accepted **62** and efficiently catalyzed the cyclization into a tetracyclic triterpene, euph-7-ene (**64**), with unexpected skeleton and backbone rearrangement, whereas *A. acidocaldarius* SHC cyclized 2,3-dihydrosqualene into a 1:1 mixture of tetracyclic (20*R*)-dammar-12-ene (**65**) and (20*R*)-dammar-13(17)-ene (**66**) (**Scheme 6**).[67,71] In both cases, the enzyme reactions produced tetracyclic ring system with a five-membered D-ring, and the products had only the 20*R* configuration. The formation of the thermodynamically favored five-membered D-ring instead of a six-membered D-ring is apparently only induced in the cyclization process by the lack of participation of the terminal π-electrons.

In the absence of the terminal double bond, the cyclization first produces a 6.6.6.5-fused tetracyclic intermediate carbocation having either a 17α- or a 17β-side chain; cyclization reaction folded in all-*chair* conformation produces a 17-*epi*-dammaranyl cation (**67**) with a 17α-side chain, whereas in *chair–chair–chair–boat* conformation yields a dammaranyl cation (**68**) with a 17β-side chain (**Scheme 6**). The final olefinic reaction products are subsequently obtained by a backbone rearrangement and the hydride shifts followed by elimination of proton to yield **64** in the case of *T. pyriformis* STC, or **65** and **66** in the case of *A. acidocaldarius* SHC, all of them with a 20*R* configuration. The stereochemistry of the cyclization and the rearrangement reactions are controlled by the enzymes, and the formation of different cyclization products by the two enzymes would reflect differences in the geometry of the active sites.

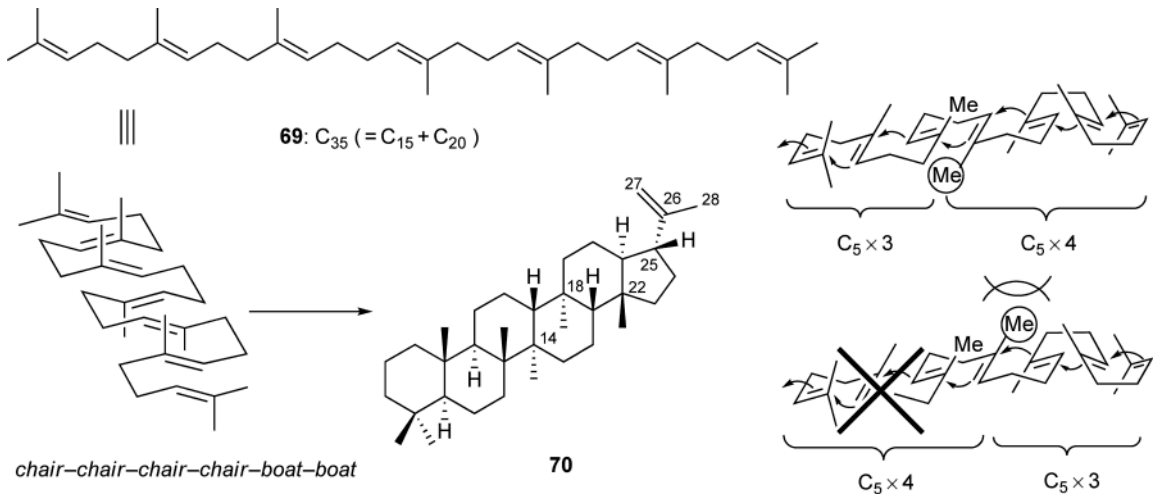
Here the formation of the dammaranyl cation with the 17β-side chain is thought to be less likely as it implies a folding of the precursor different from that adopted for squalene cyclization, although, in this case, the C-20*R* configuration can be attained through a least motion pathway involving only a small (<60°) rotation about the C-17–C-20 axis (**Scheme 6**).[67] Further, a 1,3-diaxial interaction between the two methyl groups in pre-D-ring is released and the subsequent rearrangement reaction can be explained in all antiparallel manner and in the least motion. Importantly, the differences between the two pathways are similar to those discussed for the formation of lanosterol, which also has the 20*R* configuration, from a protosterol cationic intermediate with a 17α-side chain as proposed by Cornforth[72] or with a 17β-side chain as confirmed by Corey.[12] Recently, Arigoni and coworkers repeated the experiments using [7-^{2}H]- and [11-^{2}H]-labeled isotopomers of **62** and confirmed that the formation of **64** by *T. pyriformis* STC involves two consecutive 1,2-hydride shifts, instead of a putative 1,3-hydride shift from C-13 to C-20, to achieve the 20*R* configuration.[73]

1.19.5.3 Analogues with Various Chain Lengths

It was remarkable that a synthetic C_{35} analogue (**69**) in which a C_{15} farnesyl unit is connected in a head-to-head fashion to a C_{20} geranylgeranyl unit was enzymatically cyclized into a novel hexacyclic polyprenoid (**70**) with a 6.6.6.6.6.5-fused ring system, as a single product in 10% yield, by *A. acidocaldarius* SHC (**Scheme 7**).[74] This is the first demonstration of the remarkable catalytic potential of the squalene (C_{30}) cyclizing the enzyme to generate

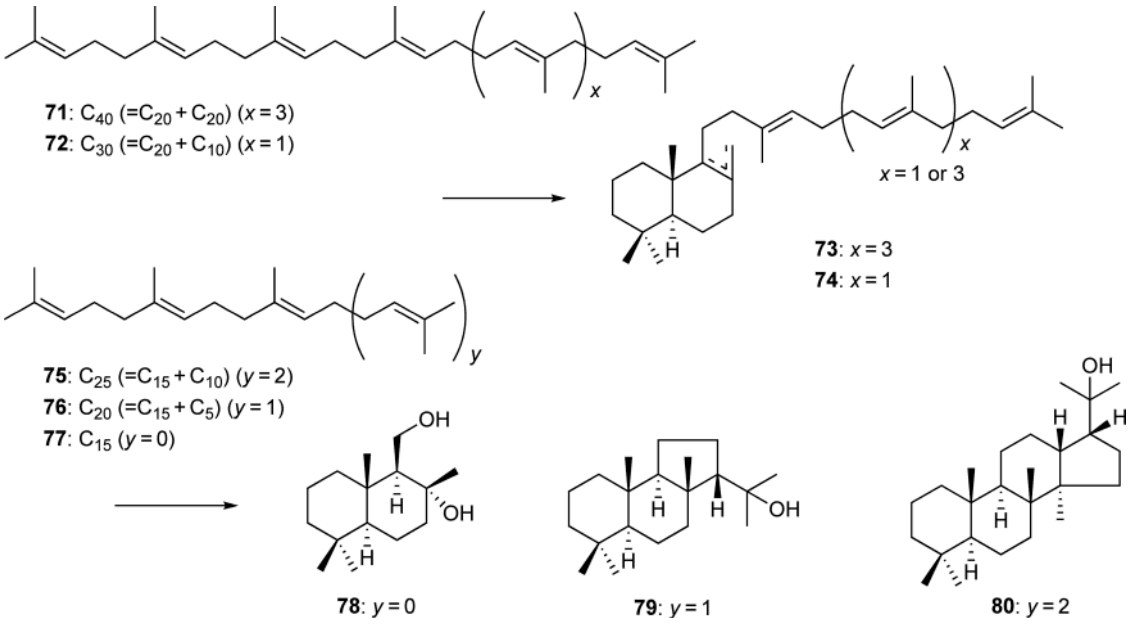


Scheme 6 Enzymatic cyclization of 2,3-dihydrosqualene by *T. pyriforms* STC and *A. acidocaldarius* SHC.



Scheme 7 Enzymatic cyclization of a C_{35} analogue by *A. acidocaldarius* SHC.

the supranatural hexacyclic polyprenoid scaffold. Interestingly, the cyclization of the C_{35} analogue is initiated from the C_{15} farnesyl end, and not by a proton attack on the terminal double bond of the C_{20} geranylgeranyl unit (**Scheme 7**). This suggested that α-orientation of the *pro*-C14 methyl group is crucial for the binding and cyclization of the substrate by *A. acidocaldarius* SHC, as in the case of the formation of hopene from squalene (C_{15} unit × 2). The alternative β-orientation of the *pro*-C14 methyl group may interact repulsively with the



Scheme 8 Enzymatic cyclization of C_{15}–C_{40} analogues by *A. acidocaldarius* SHC.

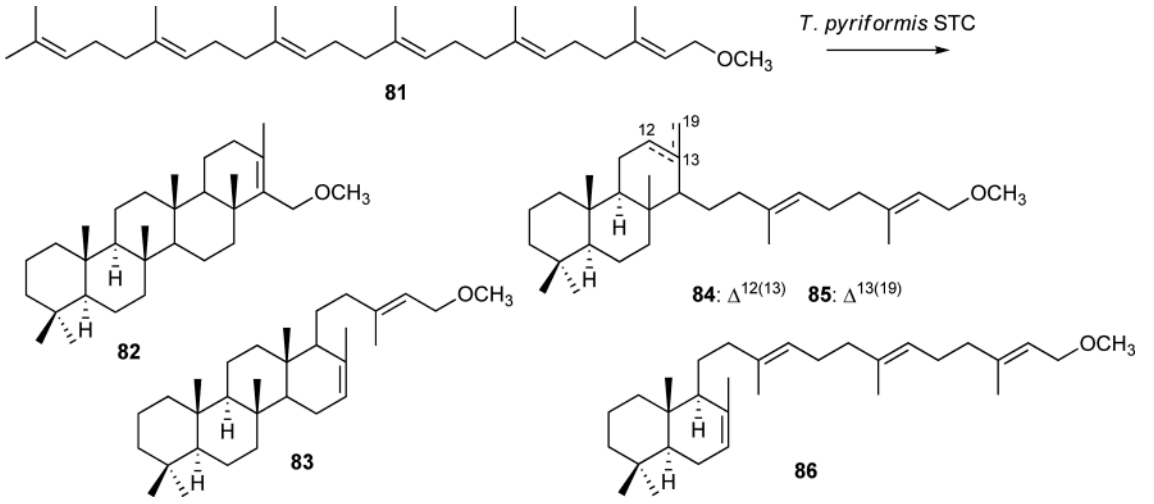
pro-C10 methyl or a nearest neighbor from the substrate-binding pocket. Moreover, the C_{35} analogue is thought to be folded in *chair–chair–chair–chair–boat–boat* conformation to achieve the stereochemistry of the α-axial orientation of the methyl group at C-18 and the isopropenyl group at C-25, as well as the β-axial orientation of the methyl group at C-22 (**Scheme 7**). Importantly, the polycyclization reaction proceeds without carbon skeletal rearrangement and is terminated by a specific proton elimination at C-27. The stereochemistry of the cyclization reaction is rigidly controlled by the enzyme even for the nonphysiological C_{35} substrate with the additional C_5 isoprene unit.

On the other hand, *A. acidocaldarius* SHC accepted a C_{40} analogue (**71**) in which two C_{20} geranylgeranyl units are connected in a head-to-head fashion (C_{20} unit + C_{20} unit) and an unnatural C_{30} analogue (**72**) in which a C_{20} geranylgeranyl unit is connected to a C_{10} geranyl in a head-to-head fashion (C_{20} unit + C_{10} unit).[75] However, the enzyme reaction did not yield highly cyclized products, just produced bicyclic polyprenoids. In contrast, a C_{25} analogue (**75**) (C_{15} unit + C_{10} unit) and a C_{20} analogue (**76**) (C_{15} unit + C_5 unit) were efficiently cyclized into a novel tetracyclic and a tricyclic alcohol, respectively (**Scheme 8**).[76,77] These results reconfirmed that the presence of the C_{15} farnesyl unit and the α-orientation of the *pro*-C14 methyl group are crucial for the formation of the highly cyclized product by the *A. acidocaldarius* SHC. Finally, it has been reported that the bacterial enzyme also cyclized farnesol (C_{15}), but not accepted geraniol (C_{10}) as a substrate.[77]

In comparison, earlier studies by Renoux and Rohmer[78] demonstrated that a cell-free system from *T. pyriformis* cyclized all-*trans* hexaprenyl methyl ether (**81**) to a series of polycyclic skeletons (**82–86**) corresponding to all cationic intermediates of a possible stepwise all-Markovnikov cyclization process (**Scheme 9**). In the substrate, the isoprene units are linked head-to-tail, whereas squalene is a symmetrical molecule with a central head-to-head linkage. It was anticipated that the enzyme should also encounter difficulty in cyclizing the C-ring since the methyl groups of hexaprenyl ether are on the β-side instead on the α-side; however, in contrast to *A. acidocaldarius* SHC, *T. pyriformis* STC efficiently yielded the highly polycyclic products, again suggesting structural differences in the active site geometry between the two enzymes.

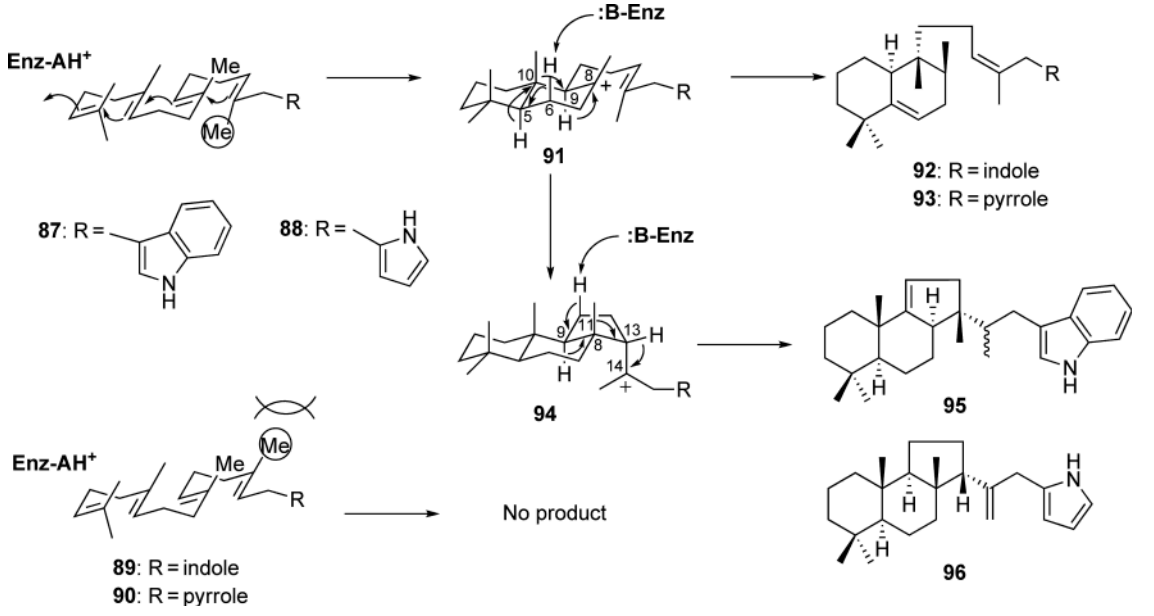
1.19.5.4 Heteroaromatic Ring-Containing Analogues

The crystal structure of *A. acidocaldarius* SHC suggested that the hydrophobic active site cavity (~1200 $Å^3$) lined with aromatic residues has enough space to accept bulky substrate analogues. Indole-containing substrate

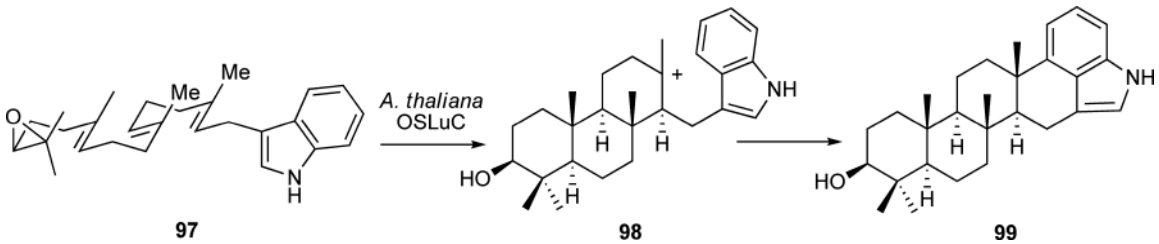


Scheme 9 Enzymatic cyclization of all-*trans* hexaprenyl methyl ether by *T. pyriforms* STC.

analogues, 3-(farnesyldimethylallyl)indole (**87**) and 3-(geranylgeranyl)indole (**89**), in which a C_{20} isoprene unit is connected to indole, were synthesized and subjected to the enzymatic cyclization by *A. acidocaldarius* SHC (**Scheme 10**).[79] The π–π interactions between the aromatic ring and the active site aromatic residues were anticipated for the substrate analogues, which could possibly affect the stereoelectronical course of the enzyme reactions. Indeed, 3-(farnesyldimethylallyl)indole (**87**) is efficiently converted to a 1:2 mixture (total yield 8%) of a bicyclic (**92**) and a tricyclic (**95**) unnatural novel polyprenoid scaffold (**Scheme 10**). The cyclization reaction was initiated in a *chair–chair–chair* conformation and apparently interrupted at the bicyclic or tricyclic stage due to the stereoelectronic effect of the bulky, electron-rich indole moiety. As a result, from the bicyclic tertiary cation (**91**), a backbone rearrangement (H-9$\alpha \rightarrow 8\alpha$, CH_3-10$\beta \rightarrow 9\beta$, H-5$\alpha \rightarrow 10\alpha$) with the elimination of H-6β produced the minor product **92**, whereas from the 6.6.5-fused tricyclic Markovnikov cation (**94**), a backbone rearrangement (H-13$\beta \rightarrow 14$, CH_3-8$\beta \rightarrow 13\beta$, H-9$\alpha \rightarrow 8\alpha$) with the elimination of H-11β yielded



Scheme 10 Enzymatic cyclization of heteroaromatic ring-containing analogues by *A. acidocaldarius* SHC.



Scheme 11 Enzymatic cyclization of 3-(ω-oxidogeranylgeranyl)indole by *A. thaliana* OSLuC.

the major product **95** (**Scheme 10**). The stereochemistry of the polycyclization of the unnatural bulky substrate analogue was rigidly controlled by the enzyme in a regio- and stereospecific manner.

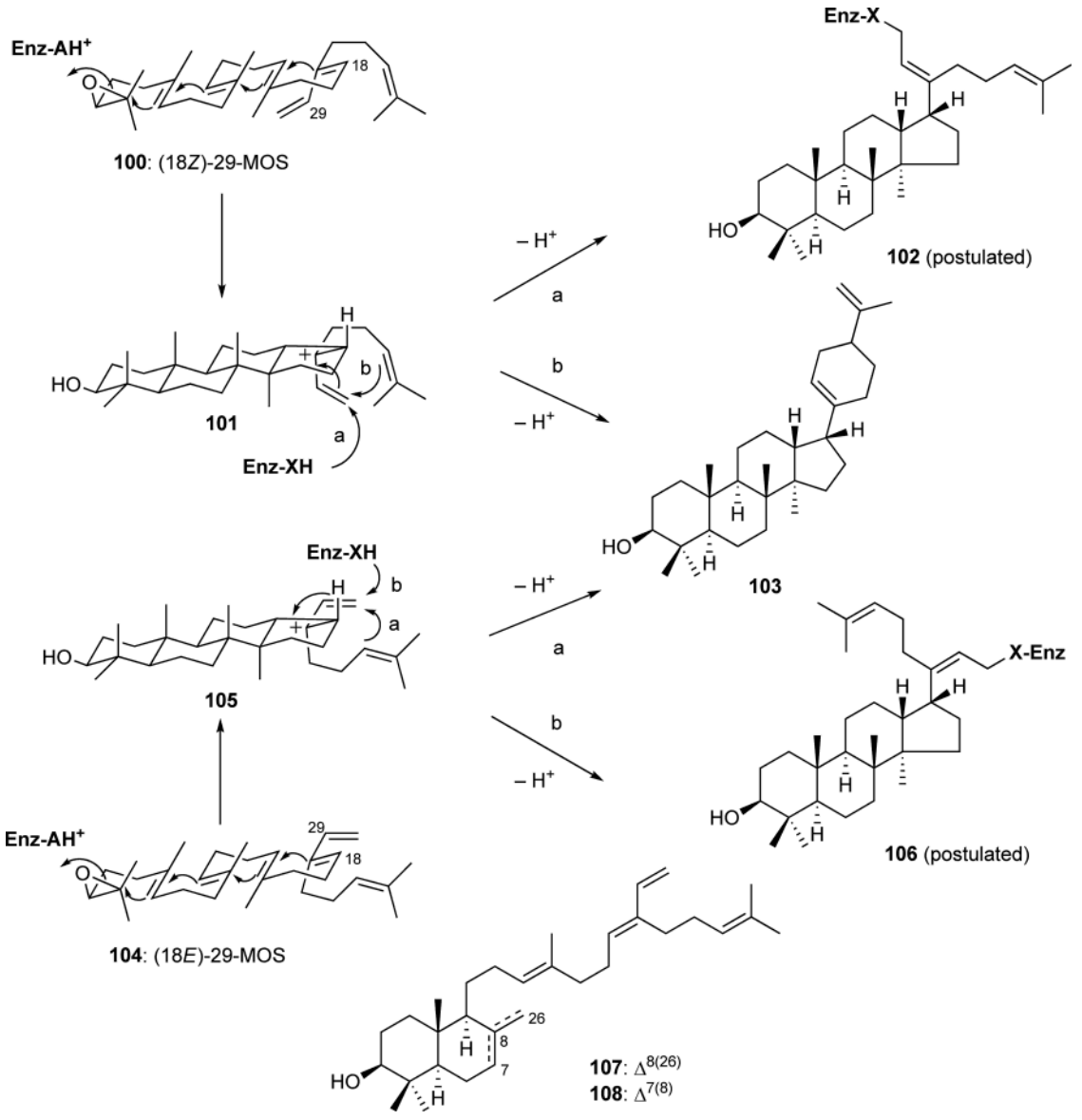
In contrast, *A. acidocaldarius* SHC did not accept 3-(geranylgeranyl)indole (**89**) as a substrate.[79] As discussed above, the bacterial SHC is sensitive to the β-orientation of the *pro*-C14 methyl group and fails to bind the geranylgeranyl substrate. Interestingly, substrate **89** is a putative precursor for natural indole diterpenes; it has been reported that plant oxidosqualene:lupeol cyclase (OSLuC) from *Arabidopsis thaliana* cyclized its epoxide, 3-(ω-oxidogeranylgeranyl)indole (**97**), into petromindole (**99**), a hexacyclic indole diterpene with a 6.6.6.6.6.5-fused ring system (**Scheme 11**).[80] The polycyclization reaction has been proposed to proceed in *chair–chair–chair* conformation to produce a 6.6.6 tricyclic tertiary Markovnikov cationic intermediate (**98**), which is followed by cyclization to indole at C-4′.

Similar results have also been obtained with the enzyme reactions of pyrrole-containing analogues, 2-(farnesyldimethylallyl)pyrrole (**88**) and 2-(geranylgeranyl)pyrrole (**90**).[81] Thus, *A. acidocaldarius* SHC did not accept 2-(geranylgeranyl)pyrrole (**90**) as a substrate, but cyclized 2-(farnesyldimethylallyl)pyrrole (**88**) into a 1:10 mixture (total yield 1%) of a bicyclic (**93**) and a tricyclic (**96**) unnatural novel polyprenoids (**Scheme 10**). Notably, the less bulky pyrrole analogue **88** was not a good substrate as the indole analogue **87**. The yield of the pyrrole products was fivefold less than that of the indole products, suggesting that the bulky, π-electron-rich indole moiety fit better into the active site of *A. acidocaldarius* SHC.

1.19.5.5 Methylidene-Extended Analogues

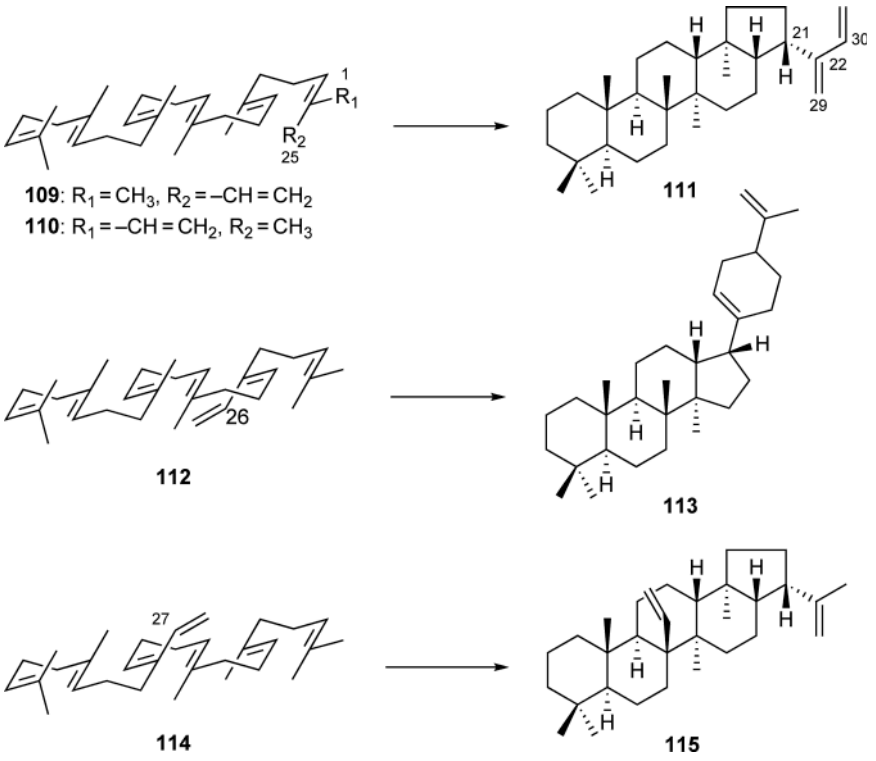
18(*Z*)-(3*S*)-29-Methylidene-2,3-oxidosqualene ((18*Z*)-29-MOS) (**100**) is the first mechanism-based irreversible inhibitor of vertebrate OSLC developed by Prestwich and coworkers.[39,82,83] Interestingly, *A. acidocaldarius* SHC accepted (18*Z*)-29-MOS as a substrate and cyclized into an unnatural C_{31} dammarene–limonene hybrid with a 6.6.6.5+6 ring system (**103**) (**Scheme 12**).[84] Further, as in the case of vertebrate OSLCs, (18*Z*)-29-MOS also functioned as an effective time-dependent irreversible inactivator of *A. acidocaldarius* SHC ($IC_{50} = 1.2\ \mu mol\ l^{-1}$, $K_I = 2.1\ \mu mol\ l^{-1}$, $k_{inact} = 0.06\ min^{-1}$).[84] On the other hand, (18*E*)-29-MOS (**104**), the Δ^{18} regioisomer, has also been reported to function as a mechanism-based irreversible inhibitor of *A. acidocaldarius* SHC ($IC_{50} = 4.5\ \mu mol\ l^{-1}$, $K_I = 7.4\ \mu mol\ l^{-1}$, $k_{inact} = 0.048\ min^{-1}$).[85] Moreover, the geometrical isomer (18*E*)-(3*S*)29-MOS was also cyclized into the same C_{31} dammarene derivative (**103**), while two bicyclic compounds (**107** and **108**) were obtained as major products, indicating that the Δ^{18} geometry significantly affects the folding and extent of the cyclization reaction (**Scheme 12**).[85] It is likely that both Δ^{18} stereoisomers interrupt the sequential ring-forming reaction at the tetracyclic 17-*epi*-dammarene C-20 cations **101** and **105**. The methylidene-extended allylcations could subsequently be further cyclized to generate the same C_{31} dammarene derivative **103** or could be trapped by an active site nucleophile resulting in covalent bond formation and concomitant irreversible inactivation of the enzyme. It is worth noting that the different orientation of 29-methylidene group in the allylcations could result in trapping by different nucleophilic residues in the active site of *A. acidocaldarius* SHC.

A series of methylidene-extended squalenes have also been synthesized and tested for the enzyme reaction. First, *A. acidocaldarius* SHC accepted both 1-methylidenesqualene (**109**) and 25-methylidenesqualene (**110**) and efficiently cyclized into 30-methylidene-hop-22(29)-ene (**111**) (35 and 38% yield, respectively) (**Scheme 13**).[69] In both cases, the cyclization first produces a pentacyclic hopanyl C-22 cation, and the



Scheme 12 Proposed mechanism for irreversible enzyme inhibition and cyclization of (18*Z*)- and (18*E*)-29-MOS by *A. acidocaldarius* SHC.

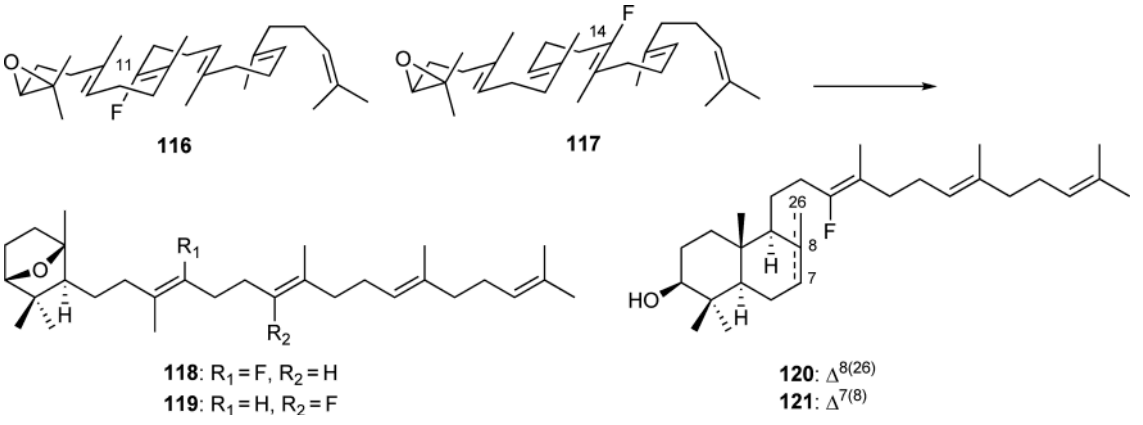
subsequent proton abstraction from the terminal methyl group yields **111** as a single product. Because of significant lifetime of the methylidene-extended intermediate cation, a rotation of the isobutenyl side chain around the C-21–C-22 bond took place prior to the final proton elimination. In contrast, as described, during the formation of hopene, the final proton abstraction takes place regiospecifically from the *Z*-methyl (Me-30) group (**Scheme 5**).[69] On the other hand, as in the case of 29-MOSs, *A. acidocaldarius* SHC cyclized 26-methylidenesqualene (**112**) into a C_{31} 17-*epi*-dammarene derivative (**113**) with the 6.6.6.5+6 ring system (10% yield), whereas 27-methylidenesqualene (**114**) was converted into 26-methylidene-hop-22(29)-ene (**115**) (3% yield) (**Scheme 13**).[86] Interestingly, the methylidene-extended squalene analogues were found to be the poor inhibitors of *A. acidocaldarius* SHC ($IC_{50} = \sim 100\,\mu mol\,l^{-1}$), except 27-methylidenesqualene ($IC_{50} = 5\,\mu mol\,l^{-1}$).[86]



Scheme 13 Enzymatic cyclization of methylidene-extended substrate analogues by *A. acidocaldarius* SHC.

1.19.5.6 Fluorine and Sulfur Analogues

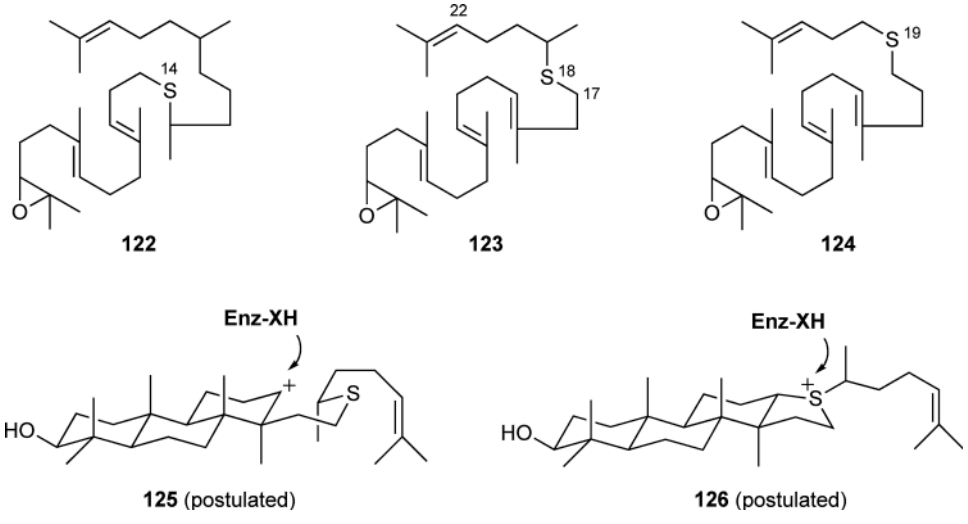
In nonenzymatic polyene cyclization reactions, Johnson demonstrated that a fluorine atom could act as a cation-stabilizing auxiliary that served to both enhance the cyclization reaction and control the regiochemistry of the product.[87–90] In order to test the effect of fluorine atom substitution on the enzymatic sequential ring-forming reaction, (3*S*)11-fluoro-2,3-oxidosqualene (11-FOS) (**116**) and (3*S*)14-fluoro-2,3-oxidosqualene (14-FOS) (**117**) were synthesized and tested for the enzyme reaction.[91,92] First, *A. acidocaldarius* SHC converted 14-FOS into a mixture of mono-carbocyclic product with a bridged ether **119** (33% yield) and a pair of bicyclic alcohols **120** and **121** (60% yield), while 11-FOS only afforded a similar monocarbocyclic product **118** (**Scheme 14**). Thus, the presence of the fluorine atom interrupted the cyclization reaction, which was initiated



Scheme 14 Enzymatic cyclization of 11-FOS and 14-FOS by *A. acidocaldarius* SHC.

by oxirane ring opening, at the mono- or bicyclic carbocationic intermediate stage, and intramolecular trapping by the 3β-OH led to the formation of the bridged bicyclic ethers. The partial cyclization is thought to be attributed to the strong electron-withdrawing effect of the fluorine atom at an incipient cationic site or to the distortion of the substrate folding by the slightly larger fluorine atom substitution.

On the other hand, Oehlschlager and coworkers synthesized a number of sulfur-containing oxidosqualene analogues in which a sulfur atom replaced carbons C-5, C-6, C-8, C-9, C-10, C-11, C-13, C-14, C-15, C-16, C-18, C-19, and C-20, and reported that many of them were potent enzyme inhibitors of vertebrate OSLCs.[93–97] When tested with *A. acidocaldarius* SHC, the sulfur analogues S-6, S-10, S-14, S-18, and S-19 also showed potent inhibition at submicromolar level. Enzyme kinetic analysis revealed that S-14 (**122**) ($K_I = 109\ \text{nmol l}^{-1}$, $k_{inact} = 0.058\ \text{min}^{-1}$) and S-19 (**124**) ($K_I = 83\ \text{nmol l}^{-1}$, $k_{inact} = 0.054\ \text{min}^{-1}$) were time-dependent inhibitors of SHC, whereas S-18 (**123**) exhibited the most potent inhibition and showed time-dependent irreversible inhibition ($K_I = 31\ \text{nmol l}^{-1}$, $k_{inact} = 0.071\ \text{min}^{-1}$).[98] The enzyme inhibition may be attributed to the higher nucleophilicity of the sulfur atom compared to nitrogen and oxygen. Further, the more flexible and longer C–S bond in comparison with the C–C double bond may serve to bring the sulfur closer to the active site nucleophilic residues important in cationic charge stabilization. Affinity-labeling experiments, using either [17-^{3}H]-S-18 or [22-^{3}H]-S-18, clearly demonstrated irreversible covalent modification of *A. acidocaldarius* SHC.[98] These results suggested that the presence of sulfur at C-18 can interrupt the cyclization and that an intermediate partially cyclized cation (**125** or **126**) may be captured by a nucleophilic residue of the SHC active site.

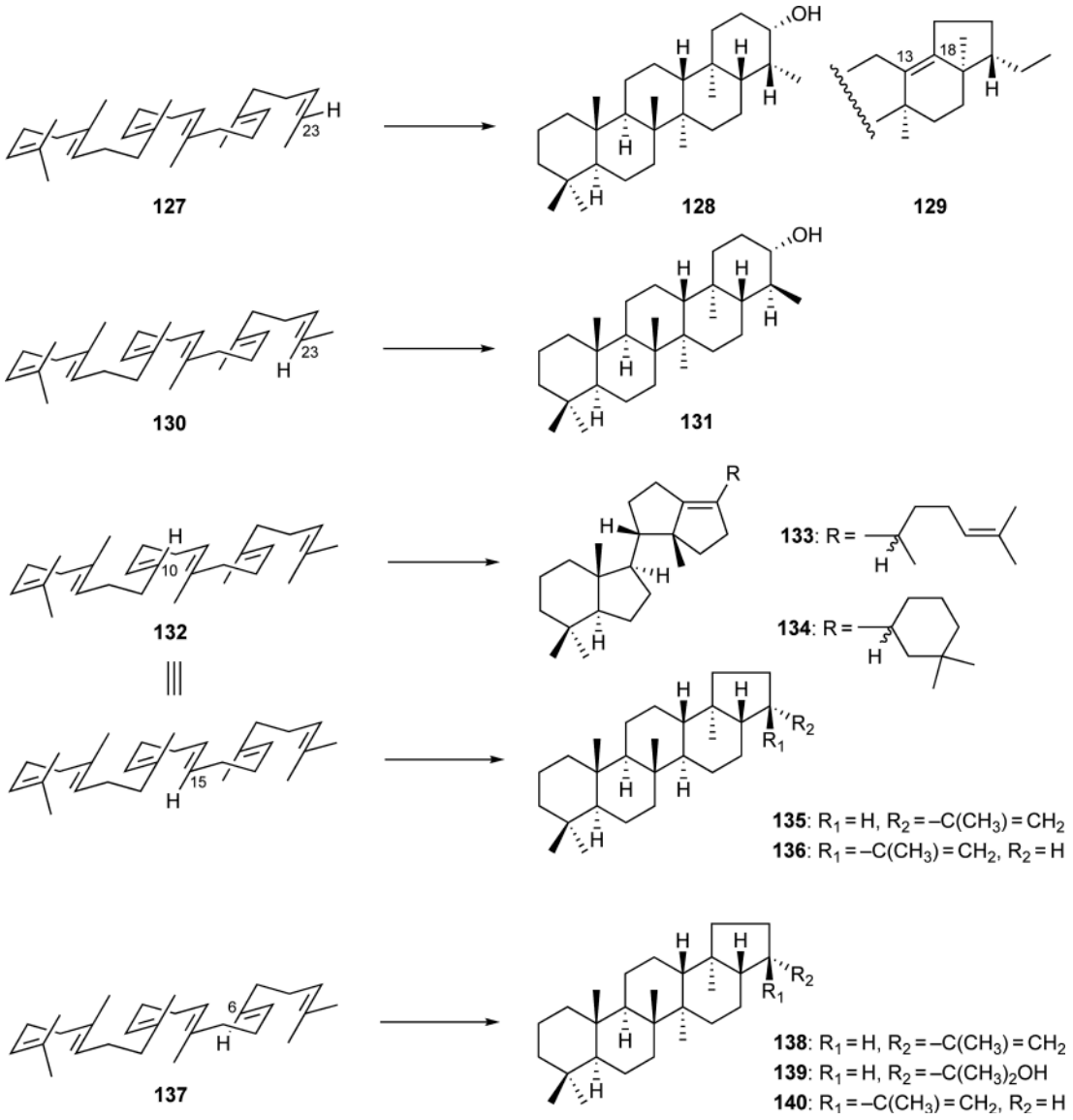


122 **123** **124**

125 (postulated) **126** (postulated)

1.19.5.7 Desmethylsqualenes and Other Analogues

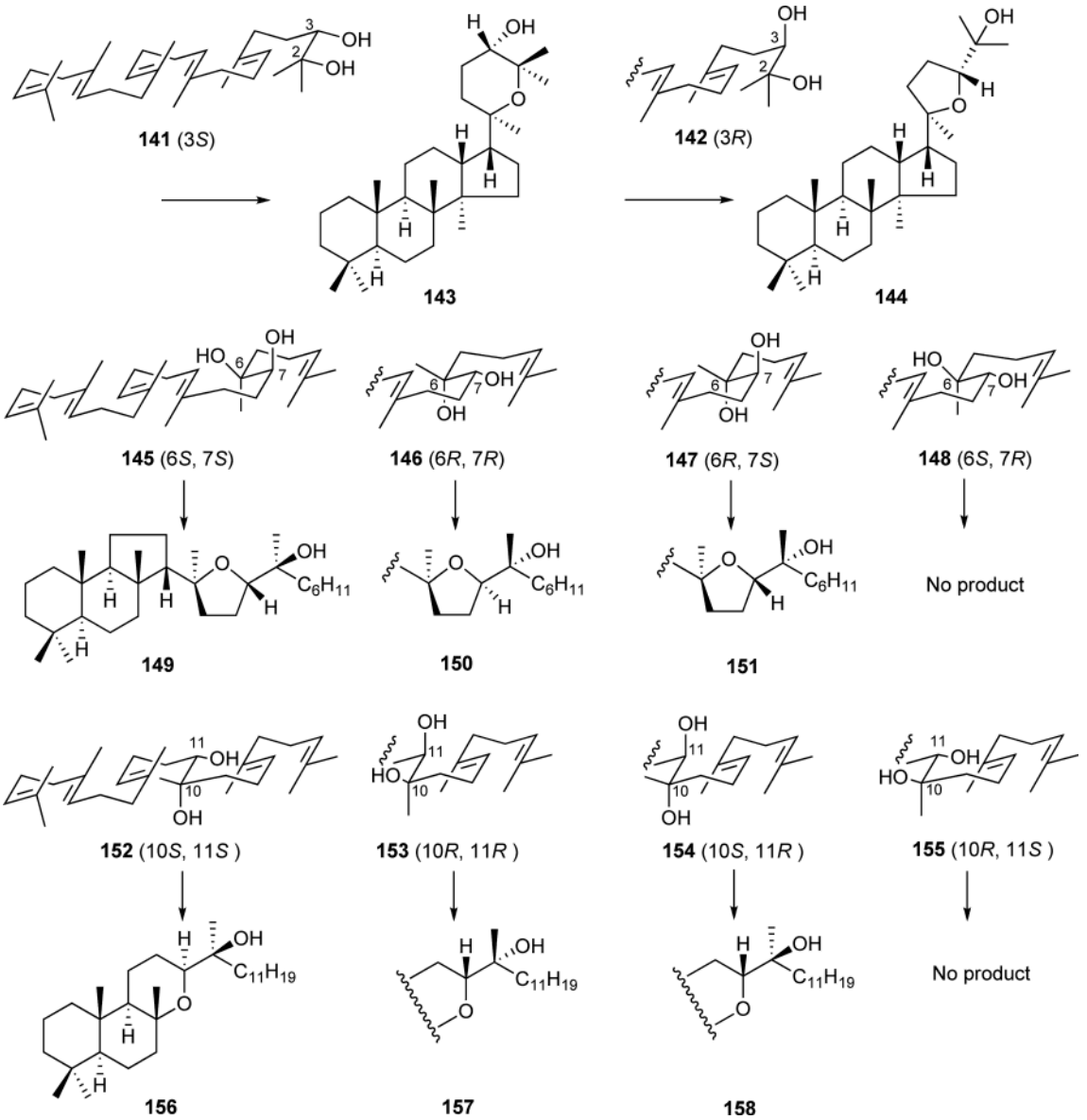
Hoshino and coworkers reported the enzymatic cyclization of a series of desmethylsqualenes by recombinant *A. acidocaldarius* SHC. 23-Desmethylsqualenes **127** and **130**, lacking one of the terminal methyl groups, were cyclized into gammacerane derivatives **128** and **131** along with a neohopane derivative **129** (**Scheme 15**), suggesting that the two geminal methyl groups of squalene are critical to the formation of the five-membered E-ring of hopene.[99] These results also suggested that the methyl groups are essential for the initiation of the polycyclization reaction. On the other hand, incubation of 10-desmethylsqualene (=15-desmethylsqualene) (**132**) with *A. acidocaldarius* SHC afforded two unprecedented polyprenoid scaffolds with a 6.5 + 5.5 tetracyclic (**133**) (46%) and a 6.5 + 5.5 + 6 pentacyclic ring system (**134**) (10%), accompanied by two novel hopane derivatives (**135** and **136**) (20 and 24%, respectively) (**Scheme 15**).[100] The former two products were generated by cyclization of C(10)-desmethylsqualene, whereas the latter derivatives were from C(15)-desmethylsqualene. This result suggested that the methyl group at C(10) is essential for the correct folding of squalene during the biosynthesis of hopene. This confirmed again that the presence of the C_{15} farnesyl unit and the α-orientation of the *pro*-C14 methyl group are crucial for the polycyclization reaction by *A. acidocaldarius*



Scheme 15 Enzymatic cyclization of desmethylsqualenes by *A. acidocaldarius* SHC.

SHC. On the other hand, 6-desmethylsqualene (**137**) just yielded a mixture of hopane derivatives through the normal cyclization pathway (**Scheme 15**).[101]

Finally, the Hoshino group also investigated the enzymatic cyclizations of a series of squalene diols by *A. acidocaldarius* SHC. First, the enzyme reactions of (3*R*)-2,3-squalenediol (**141**) and (3*S*)-2,3-squalenediol (**142**) afforded novel ether ring-containing dammaranes (**143** and **144**) derived from 6.6.6.5-fused tetracyclic tertiary cations (**Scheme 16**).[102] It has been reported that 2,3:22,23-dioxidosqualenes were also cyclized into the similar epoxydammaranes by *A. acidocaldarius* SHC.[102] On the other hand, *threo*- and *erythro*-diols at the 6,7-position (**145–148**) and 10,11-position (**152–155**) produced ether ring-containing polycyclic products derived from 6.6.5-fused tricyclic, 6.6-fused bicyclic, and monocyclic Markovnikov tertiary cations (**Scheme 16**).[103] The observed stereochemical course of the cyclization reactions suggested that the *chair* folding conformation of squalene diols is tightly constricted by the enzyme, and the substrate and product specificities are dominantly directed by the least motion of the nucleophilic hydroxyl group toward the intermediate carbocations.



Scheme 16 Enzymatic cyclization of squalenediols by *A. acidocaldarius* SHC.

1.19.6 Conclusions

In the past several years, there have been remarkable advances in the study of bacterial SC enzymes. The crystallographic and structure-based mutagenesis studies as well as utilization of synthetic active site-targeted probes have begun to reveal intimate structural details of the catalytic mechanism of the enzyme-templated sequential carbon–carbon bonds forming reactions. On the other hand, the relaxed substrate specificity and catalytic plasticity of the bacterial SCs are remarkable from the point of view of the engineered biosynthesis. The enzymes accept a wide variety of nonphysiological substrates and efficiently perform sequential ring-forming reactions to produce a series of unnatural cyclic triterpenes (see Chapter 1.18). Manipulation of the enzyme reactions by the newly designed active site probes and the rationally engineered mutant enzymes would thus lead to further production of chemically and structurally divergent supranatural novel polyprenoid scaffolds.

Acknowledgments

Financial support has been provided by the PRESTO program from Japan Science and Technology Agency, and Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

Abbreviations

11-FOS	(3*S*)11-fluoro-2,3-oxidosqualene
14-FOS	(3*S*)14-fluoro-2,3-oxidosqualene
(18*Z*)-29-MOS	18(*Z*)-(3*S*)-29-methylidene-2,3-oxidosqualene
OSC	oxidosqualene cyclase
OSLC	oxidosqualene:lanosterol cyclase
OSLuC	oxidosqualene:lupeol cyclase
SC	squalene cyclase
SHC	squalene:hopene cyclase
STC	squalene:tetrahymanol cyclase

References

1. G. Ourisson; M. Rohmer; K. Poralla, *Annu. Rev. Microbiol.* **1987**, *41*, 301–333.
2. M. Rohmer; P. Bouvier; G. Ourisson, *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 847–851.
3. G. Ourisson; M. Rohmer, *Acc. Chem. Res.* **1992**, *25*, 403–408.
4. I. Abe; M. Rohmer; G. D. Prestwich, *Chem. Rev.* **1993**, *93*, 2189–2206.
5. K. Poralla, Cycloartenol and Other Triterpene Cyclases. In *Comprehensive Natural Products Chemistry*; D. H. R. Barton, K. Nakanishi, Eds.; Pergamon: Oxford, 1999; Vol. 2, Chapter 11, pp 299–320.
6. K. U. Wendt; G. E. Schulz; E. J. Corey; D. R. Liu, *Angew. Chem. Int. Ed.* **2000**, *39*, 2812–2833.
7. I. Abe, *Nat. Prod. Rep.* **2007**, *24*, 1311–1331.
8. T. Hoshino; T. Sato, *Chem. Commun.* **2002**, 291–301.
9. E. Caspi, *Acc. Chem. Res.* **1980**, *13*, 97–104.
10. I. Abe; G. D. Prestwich, Squalene Epoxidase and Oxidosqualene: Lanosterol Cyclase–Key Enzymes in Cholesterol Biosynthesis. In *Comprehensive Natural Products Chemistry*; D. H. R. Barton, K. Nakanishi, Eds.; Pergamon: Oxford, 1999; Vol. 2, Chapter 10, pp 267–298.
11. E. J. Corey; S. C. Virgil; S. Sarshar, *J. Am. Chem. Soc.* **1991**, *113*, 8171–8172.
12. E. J. Corey; S. C. Virgil, *J. Am. Chem. Soc.* **1991**, *113*, 4025–4026.
13. E. J. Corey; S. C. Virgil; H. Cheng; C. H. Baker; S. P. T. Matsuda; V. Singh; S. Sarshar, *J. Am. Chem. Soc.* **1995**, *117*, 11819–11820.
14. G. Ourisson, *Pure Appl. Chem.* **1989**, *61*, 345–348.
15. K. U. Wendt; K. Poralla; G. E. Schulz, *Science* **1997**, *277*, 1811–1815.
16. R. Xu; G. C. Fazio; S. P. Matsuda, *Phytochemistry* **2004**, *65*, 261–291.
17. D. A. Dougherty, *Science* **1996**, *271*, 163–168.
18. J. C. Ma; D. A. Dougherty, *Chem. Rev.* **1997**, *97*, 1303–1324.
19. R. B. Woodward; K. Bloch, *J. Am. Chem. Soc.* **1953**, *75*, 2023–2024.
20. A. Eschenmoser; L. Ruzicka; O. Jeger; D. Arigoni, *Helv. Chim. Acta* **1955**, *38*, 1890–1904.
21. G. Stork; A. W. Burgstahler, *J. Am. Chem. Soc.* **1955**, *77*, 5068–5077.
22. J. W. Cornforth; R. H. Cornforth; C. Donninger; G. Popják; Y. Shimizu; S. Ichii; E. Forchielli; E. Caspi, *J. Am. Chem. Soc.* **1965**, *87*, 3224–3228.
23. E. J. Corey; W. E. Russey; P. R. Ortiz de Montellano, *J. Am. Chem. Soc.* **1966**, *88*, 4750–4751.
24. E. E. van Tamelen; J. D. Willett; R. B. Clayton; K. E. Lord, *J. Am. Chem. Soc.* **1966**, *88*, 4752–4754.
25. J. D. Willett; K. B. Sharpless; K. E. Lord; E. E. van Tamelen; R. B. Clayton, *J. Biol. Chem.* **1967**, *242*, 4182–4191.
26. R. A. Yoder; J. N. Johnston, *Chem. Rev.* **2005**, *105*, 4730–4756.
27. E. E. van Tamelen, *J. Am. Chem. Soc.* **1982**, *104*, 6480–6481.
28. C. Pale-Grosdemange; C. Feil; M. Rohmer; K. Poralla, *Angew. Chem. Int. Ed.* **1998**, *37*, 2237–2240.
29. B. A. Hess, Jr., *J. Am. Chem. Soc.* **2002**, *124*, 10286–10287.
30. W. S. Johnson; S. J. Telfer; S. Cheng; U. Schubert, *J. Am. Chem. Soc.* **1987**, *109*, 2517–2518.
31. W. S. Johnson; S. D. Lindell; J. Steele, *J. Am. Chem. Soc.* **1987**, *109*, 5852–5853.
32. C. Jenson; W. L. Jorgensen, *J. Am. Chem. Soc.* **1997**, *119*, 10846–10854.
33. B. A. Hess, Jr., *Org. Lett.* **2003**, *5*, 165–167.
34. B. A. Hess, Jr.; L. Smentek, *Org. Lett.* **2004**, *6*, 1717–1720.

35. R. Rajamani; J. Gao, *J. Am. Chem. Soc.* **2003**, *125*, 12768–12781.
36. D. J. Reinert; G. Balliano; G. E. Schulz, *Chem. Biol.* **2004**, *11*, 121–126.
37. D. Ochs; C. H. Tappe; P. Gartner; R. Kellner; K. Poralla, *Eur. J. Biochem.* **1990**, *194*, 75–80.
38. D. Ochs; C. Kaletta; K. D. Entian; A. Beck-Sickinger; K. Poralla, *J. Bacteriol.* **1992**, *174*, 298–302.
39. I. Abe; G. D. Prestwich, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9274–9278.
40. K. U. Wendt; A. Lenhart; G. E. Schulz, *J. Mol. Biol.* **1999**, *286*, 175–187.
41. R. Thoma; T. Schulz-Gasch; B. D'Arcy; J. Benz; J. Aebi; H. Dehmlow; M. Hennig; M. Stihle; A. Ruf, *Nature* **2004**, *432*, 118–122.
42. A. Ruf; F. Muller; B. D'Arcy; M. Stihle; E. Kusznir; C. Handschin; O. H. Morand; R. Thoma, *Biochem. Biophys. Res. Commun.* **2004**, *315*, 247–254.
43. K. Poralla; A. Hewelt; G. D. Prestwich; I. Abe; I. Reipen; G. Sprenger, *Trends Biochem. Sci.* **1994**, *19*, 157–158.
44. A. Lenhart; W. A. Weihofen; A. E. Pleschke; G. E. Schulz, *Chem. Biol.* **2002**, *9*, 639–645.
45. I. Abe; Y. F. Zheng; G. D. Prestwich, *Biochemistry* **1998**, *37*, 5779–5784.
46. O. H. Morand; J. D. Aebi; H. Dehmlow; Y.-H. Ji; N. Gains; H. Lengsfeld; J. Himber, *J. Lipid Res.* **1997**, *38*, 373–390.
47. M. Mark; P. Müller; R. Mainer; B. Eisele, *J. Lipid Res.* **1996**, *37*, 148–158.
48. B. Eisele; R. Budzinski; P. Müller; R. Maier; M. Mark, *J. Lipid Res.* **1997**, *38*, 564–575.
49. T. Dang; I. Abe; Y. F. Zheng; G. D. Prestwich, *Chem. Biol.* **1999**, *6*, 333–341.
50. C. Feil; R. Süssmuth; G. Jung; K. Poralla, *Eur. J. Biochem.* **1996**, *242*, 51–55.
51. T. Dang; G. D. Prestwich, *Chem. Biol.* **2000**, *7*, 643–649.
52. T. Sato; T. Hoshino, *Biosci. Biotechnol. Biochem.* **1999**, *63*, 1171–1780.
53. T. Hoshino; T. Sato, *Chem. Commun.* **1999**, 2005–2006.
54. T. Sato; T. Hoshino, *Biosci. Biotechnol. Biochem.* **2001**, *65*, 2233–2242.
55. T. Merkofer; C. Pale-Grosdemange; K. U. Wendt; M. Rohmer; K. Poralla, *Tetrahedron Lett.* **1999**, *40*, 2121–2124.
56. C. Pale-Grosdemange; T. Merkofer; M. Rohmer; K. Poralla, *Tetrahedron Lett.* **1999**, *40*, 6009–6012.
57. T. Sato; S. Sasahara; T. Yamakami; T. Hoshino, *Biosci. Biotechnol. Biochem.* **2002**, *66*, 1660–1670.
58. T. Hoshino; M. Kouda; T. Abe; S. Ohashi, *Biosci. Biotechnol. Biochem.* **1999**, *63*, 2038–2041.
59. N. Morikubo; Y. Fukuda; K. Ohtake; N. Shinya; D. Kiga; K. Sakamoto; M. Asanuma; H. Hirota; S. Yokoyama; T. Hoshino, *J. Am. Chem. Soc.* **2006**, *128*, 13184–13194.
60. T. Hoshino; M. Kouda; T. Abe; T. Sato, *Chem. Commun.* **2000**, 1485–1486.
61. C. Full, *FEBS Lett.* **2001**, *509*, 361–364.
62. C. Full; K. Poralla, *FEMS Microbiol. Lett.* **2000**, *183*, 221–224.
63. T. Hoshino; T. Abe; M. Kouda, *Chem. Commun.* **2000**, 441–442.
64. T. Sato; M. Kouda; T. Hoshino, *Biosci. Biotechnol. Biochem.* **2004**, *68*, 728–738.
65. T. Hoshino; K. Shimizu; T. Sato, *Angew. Chem. Int. Ed.* **2004**, *43*, 6700–6703.
66. P. Bouvier; Y. Berger; M. Rohmer; G. Ourisson, *Eur. J. Biochem.* **1980**, *112*, 549–556.
67. I. Abe; M. Rohmer, *J. Chem. Soc. Perkin Trans. I* **1994**, 783–791.
68. M. Rohmer; C. Anding; G. Ourisson, *Eur. J. Biochem.* **1980**, *112*, 541–547.
69. H. Tanaka; H. Noguchi; I. Abe, *Org. Lett.* **2004**, *6*, 803–806.
70. T. Hoshino; S. Nakano; T. Kondo; T. Sato; A. Miyoshi, *Org. Biomol. Chem.* **2004**, *2*, 1456–1470.
71. I. Abe; M. Rohmer, *J. Chem. Soc. Chem. Commun.* **1991**, 902–903.
72. J. W. Cornforth, *Angew. Chem. Int. Ed. Engl.* **1968**, *7*, 903–911.
73. J. L. Giner; S. Rocchetti; S. Neunlist; M. Rohmer; D. Arigoni, *Chem. Commun.* **2005**, 3089–3091.
74. I. Abe; H. Tanaka; H. Noguchi, *J. Am. Chem. Soc.* **2002**, *124*, 14514–14515.
75. I. Abe; H. Tanaka; Y. Takahashi; W. Lou; H. Noguchi, Enzymatic Synthesis of Unnatural Novel Compounds on the Basis of Structure Function Analysis of Biosynthetic Enzymes. In *Proceedings of the 45th Symposium on the Chemistry of Natural Products*, Kyoto, Japan, 2003; pp 19–24.
76. H. Noma; H. Tanaka; H. Noguchi; M. Shibuya; Y. Ebizuka; I. Abe, *Tetrahedron Lett.* **2004**, *45*, 8299–8301.
77. T. Hoshino; Y. Kumai; I. Kudo; S. Nakano; S. Ohashi, *Org. Biomol. Chem.* **2004**, *2*, 2650–2657.
78. J.-M. Renoux; M. Rohmer, *Eur. J. Biochem.* **1986**, *155*, 125–132.
79. H. Tanaka; H. Noguchi; I. Abe, *Org. Lett.* **2005**, *7*, 5873–5876.
80. Q. Xiong; W. K. Wilson; S. P. Matsuda, *Angew. Chem. Int. Ed.* **2006**, *45*, 1–5.
81. H. Tanaka; H. Noma; H. Noguchi; I. Abe, *Tetrahedron Lett.* **2007**, *47*, 3085–3089.
82. X.-Y. Xiao; G. D. Prestwich, *J. Am. Chem. Soc.* **1991**, *113*, 9673–9674.
83. I. Abe; G. D. Prestwich, *J. Biol. Chem.* **1994**, *269*, 802–804.
84. I. Abe; T. Dang; Y.-F. Zheng; B. A. Madden; C. Feil; K. Poralla; G. D. Prestwich, *J. Am. Chem. Soc.* **1997**, *119*, 11333–11334.
85. Y. F. Zheng; I. Abe; G. D. Prestwich, *J. Org. Chem.* **1998**, *63*, 4872–4873.
86. H. Tanaka; H. Noguchi; I. Abe, *Tetrahedron Lett.* **2004**, *45*, 3093–3096.
87. W. S. Johnson; R. A. Buchanan; W. R. Bartlett; F. S. Tham; R. K. Kullnig, *J. Am. Chem. Soc.* **1993**, *115*, 504–515.
88. W. S. Johnson; B. Chenera; F. S. Tham; R. K. Kullnig, *J. Am. Chem. Soc.* **1993**, *115*, 493–497.
89. W. S. Johnson; V. R. Fletcher; B. Chenera; W. R. Bartlett; F. S. Tham; R. K. Kullnig, *J. Am. Chem. Soc.* **1993**, *115*, 497–504.
90. W. S. Johnson; M. S. Plummer; S. P. Reddy; W. R. Bartlett, *J. Am. Chem. Soc.* **1993**, *115*, 515–521.
91. B. Robustell; I. Abe; G. D. Prestwich, *Tetrahedron Lett.* **1998**, *39*, 957–960.
92. B. Robustell; I. Abe; G. D. Prestwich, *Tetrahedron Lett.* **1998**, *39*, 9385–9388.
93. Y. F. Zheng; D. S. Dodd; A. C. Oehlschlager; P. G. Hartman, *Tetrahedron* **1995**, *51*, 5255–5276.
94. Y. F. Zheng; A. C. Oehlschlager; N. H. Georgopapadakou; P. G. Hartman; P. Scheliga, *J. Am. Chem. Soc.* **1995**, *117*, 670–680.
95. Y. F. Zheng; A. C. Oehlschlager; P. G. Hartman, *J. Org. Chem.* **1994**, *59*, 5803–5809.
96. I. Abe; W. Liu; A. C. Oehlschlager; G. D. Prestwich, *J. Am. Chem. Soc.* **1996**, *118*, 9180–9181.
97. D. Stach; Y. F. Zheng; A. L. Perez; A. C. Oehlschlager; I. Abe; G. D. Prestwich; P. G. Hartman, *J. Med. Chem.* **1997**, *40*, 201–209.

98. Y. F. Zheng; I. Abe; G. D. Prestwich, *Biochemistry* **1998**, *37*, 5981–5987.
99. T. Hoshino; T. Kondo, *Chem. Commun.* **1999**, 731–732.
100. T. Hoshino; S. Ohashi, *Org. Lett.* **2002**, *4*, 2553–2556.
101. S. Nakano; S. Ohashi; T. Hoshino, *Org. Biomol. Chem.* **2004**, *2*, 2012–2022.
102. T. Hoshino; Y. Yonemura; T. Abe; Y. Sugino, *Org. Biomol. Chem.* **2007**, *5*, 792–801.
103. T. Abe; T. Hoshino, *Org. Biomol. Chem.* **2005**, *3*, 3127–3139.

Biographical Sketch

Ikuro Abe graduated from The University of Tokyo in 1984 and obtained his Ph.D. degree in 1989 from The University of Tokyo under the direction of Professor Yutaka Ebizuka. After a 2-year postdoctoral research with Professor Guy Ourisson at the CNRS Institut de Chimie des Substances Naturelles and Professor Michel Rohmer at the Ecole Nationale Supérieure de Chimie de Mulhouse (1989–91), he moved to the United States to work with Professor Glenn D. Prestwich at the State University of New York at Stony Brook (1991–96) and then at The University of Utah (1996–98) as a research assistant professor. In 1998, he returned to Japan at University of Shizuoka, School of Pharmaceutical Sciences, and was an investigator of PRESTO, Japan Science and Technology Agency (2005–09). In 2009, he moved back to The University of Tokyo as Professor of Natural Products Chemistry. His research interests involve exploring and engineering the biosynthesis of natural products.

1.20 Carotenoids

Norihiko Misawa, Kirin Holdings Co. Ltd., Nonoichi-machi, Ishikawa, Japan

© 2010 Elsevier Ltd. All rights reserved.

1.20.1	**Introduction**	733
1.20.2	**Biosynthesis of Basic Structures of Carotenoids**	734
1.20.3	**Biosynthesis of C_{30} Carotenoids**	736
1.20.4	**Biosynthesis of C_{40} Carotenoids**	736
1.20.4.1	Photosynthetic Bacteria	736
1.20.4.2	Nonphotosynthetic Bacteria – Genus *Pantoea*	737
1.20.4.3	Nonphotosynthetic Bacteria – Genera *Paracoccus* and *Brevundimonas*	739
1.20.4.4	Structure of Gene Clusters from Cyclic Carotenoid-producing Bacteria	742
1.20.4.5	Cyanobacteria	742
1.20.5	**Pathway Engineering for Useful Carotenoid Formation in *Escherichia coli***	745
1.20.5.1	Pathway Engineering for Increasing the Intracellular Concentration of Farnesyl Diphosphate	745
1.20.5.2	Pathway Engineering for the Production of Novel or Rare Carotenoids	746
1.20.6	**Pathway Engineering for Useful Carotenoid Formation in Higher Plants**	746
1.20.6.1	Biosynthesis of Carotenoids in Higher Plants	746
1.20.6.2	Pathway Engineering for Increasing Carotenoid Content in Higher Plants	749
1.20.6.3	Pathway Engineering for the Production of Ketocarotenoids in Higher Plants	749
1.20.7	**Conclusion**	750
References		750

1.20.1 Introduction

Carotenoids are natural pigments responsible for colors ranging from yellow through orange to red, and are biosynthesized in all photosynthetic organisms, including cyanobacteria, photosynthetic bacteria belonging to Alphaproteobacteria, algae, and higher plants, and also in some nonphotosynthetic bacteria, yeasts, and fungi. Photosynthetic organisms utilize carotenoids as essential components of the light-harvesting antenna complex in photosynthesis, and as a precursor of the hormone abscisic acid in higher plants. The pigments are also indispensable in animals. Since animals including fish and crustaceans, as well as mammals, cannot biosynthesize carotenoids, they must ingest the pigments from their diet and metabolize them for use as their essential nutritional elements, for example, as a precursor of vitamin A. Carotenoids are also considered to protect organism cells from oxidative damage caused by active oxygen species.[1–3] Several carotenoid pigments such as β-carotene, astaxanthin, and canthaxanthin have been used commercially as food colorants and as animal feed additives in poultry farming and aquaculture. More recently, some carotenoids including astaxanthin, β-cryptoxanthin, lutein, zeaxanthin, and lycopene have attracted great attention due to their beneficial effects on human health, for example, in preventing chronic diseases such as cancer,[4–6] cardiovascular disease,[7–9] and age-related macular degeneration,[10,11] and have been used commercially as nutraceuticals for pharmaceutical and cosmetic purposes. The global market for carotenoids was US$766 million in 2007 according to a BCC research report (code: FOD025C).

More than 750 structurally different carotenoids have been isolated from natural sources to date.[12] The pigments are classified based on their chemical composition as either carotenes (hydrocarbon carotenoids) or xanthophylls (oxygen-containing carotenoids). All of the carotenoids are biosynthesized from the hydrocarbon phytoene (or 4,4′-diapophytoene in a minor case). Phytoene and 4,4′-diapophytoene are generated from two molecules of geranylgeranyl diphosphate (pyrophosphate) (GGPP) and farnesyl diphosphate (FPP), respectively. These hydrocarbon precursors are desaturated (dehydrogenated) to form conjugated double

bonds, frequently cyclized, and subjected to a series of modifications to form a large variety of xanthophylls. Thus, most of the structural diversity of carotenoids is due to that of xanthophylls, which are synthesized from relatively small numbers of carotenes. It is interesting to contrast this with the structural diversity of other terpenes, sesquiterpenes, diterpenes, and triterpenes, in which a variety of structurally different cyclic terpenes are synthesized with distinct terpene cyclases from FPP, GGPP, and 2,3-oxidosqualene, respectively, which include the chemically active oxygen atom in their molecules, and further have various modification reactions. The structural characteristic of carotenoids seems to enable us, via effective pathway engineering (metabolic engineering) approaches, to generate a variety of new or rare xanthophylls as well as industrially important carotenoids by using combinations of genes encoding a series of modification enzymes such as oxygenases and glycosyl transferases. In this chapter, the biosynthesis of basic structures of carotenoids and typical examples of subsequent xanthophyll formation at the levels of reactions with enzymes as gene products are shown. Finally, a description of the pathway engineering approaches used with recombinant *Escherichia coli* or higher plants for the biosynthesis of useful, new, or naturally rare carotenoids, in terms of economic production, is given. The common name of the carotenoids is, and the IUPAC name is also given in parentheses.

1.20.2 Biosynthesis of Basic Structures of Carotenoids

The vast majority of carotenoids have a skeleton composed of 40 carbon atoms (C_{40}), which can be classified as tetraterpenes, and originate from phytoene ((15*Z*)-7,8,11,12,7′,8′,11′,12′-octahydro-ψ,ψ-carotene), the biosynthetic first hydrocarbon-carotenoid to exhibits no color (**Figure 1**).[12] Phytoene is synthesized through tail-to-tail condensation of two molecules of GGPP with the elimination of two diphosphate groups. This step is catalyzed by a phytoene synthase enzyme requiring adenosine triphosphate (ATP),[13] named CrtB in bacteria[14,15] or PSY in higher plants.[16,17] The CrtB and PSY proteins have amino acid sequences, suggesting that both enzymes evolved from a common ancestor. GGPP is formed from FPP and isopentenyl diphosphate (IPP) with GGPP synthase (that is, CrtE) in the case of carotenogenic (carotenoid-producing) bacteria (**Figure 1**).[15,18] Genes for CrtE (*crtE*) often exist in carotenoid gene clusters (see also **Figure 6**).[19,20]

Phytoene is typically converted into lycopene ((all-*E*)-ψ,ψ-carotene) by four double bond formation reactions at positions 7–8, 11–12, 7′–8′, and 11′–12′, and the 15-*cis*- to *trans*-isomerization reaction (**Figure 1**). This step occurs with three distinct enzymes in higher plants; that is, two dehydrogenated enzymes, phytoene desaturase (PDS)[21,22] and ζ-carotene desaturase (ZDS),[23] and one isomerase enzyme (carotene isomerase (CRTISO)), which is postulated to convert the generated prolycopene ((7*Z*,9*Z*,7′*Z*, 9′*Z*)-tetra-*cis*-lycopene) to the all-*trans* form.[24,25] Algae and cyanobacteria possess a similar three-component enzyme system to higher plants. The ZDS protein from the higher plant *Capsicum annuum* revealed 33–35% similarity to plant or cyanobacterial phytoene desaturases.[23] In the case of carotenogenic bacteria, the step from phytoene to lycopene is obviously performed with only one enzyme phytoene desaturase (CrtI), requiring FAD as a cofactor (**Figure 1**).[26,27] This CrtI protein shows significant amino acid homology with CRTISO, indicating that both enzymes originate from a common ancestor.[24] A primitive cyanobacterium *Gloeobacter violaceus* exceptionally possesses the CrtI enzyme instead of the plant-type enzymes.[28,29] One such enzyme system also seems to have been adopted in carotenogenic yeasts and fungi. In purple photosynthetic bacteria belonging to the genus *Rhodobacter* and its related genera, the phytoene is converted into neurosporene (7,8-dihydro-ψ,ψ-carotene) by three-step desaturation reactions as well as the isomerization reaction with a distinct CrtI, as shown in **Figure 1**.[30–32] Lycopene is cyclized to form the β end group (β-ionone ring, β-ring; see **Scheme 1**) in majority of photosynthetic organisms including cyanobacteria, algae, and higher plants, and also in some nonphotosynthetic bacteria, yeasts, and fungi. Lycopene is also cyclized to form on ε end group (α-ionone ring, ε-ring) in higher plants, green algae, and some cyanobacterial strains.[33] Generated dicyclic carotene and monocyclic carotenes, specifically carotenes with β-ring(s), β-carotene, and γ-carotene, as well as acyclic carotenes, lycopene, and neurosporene, are subjected to a series of various modification enzymes to synthesize the wide diversity of xanthophyll pigments.

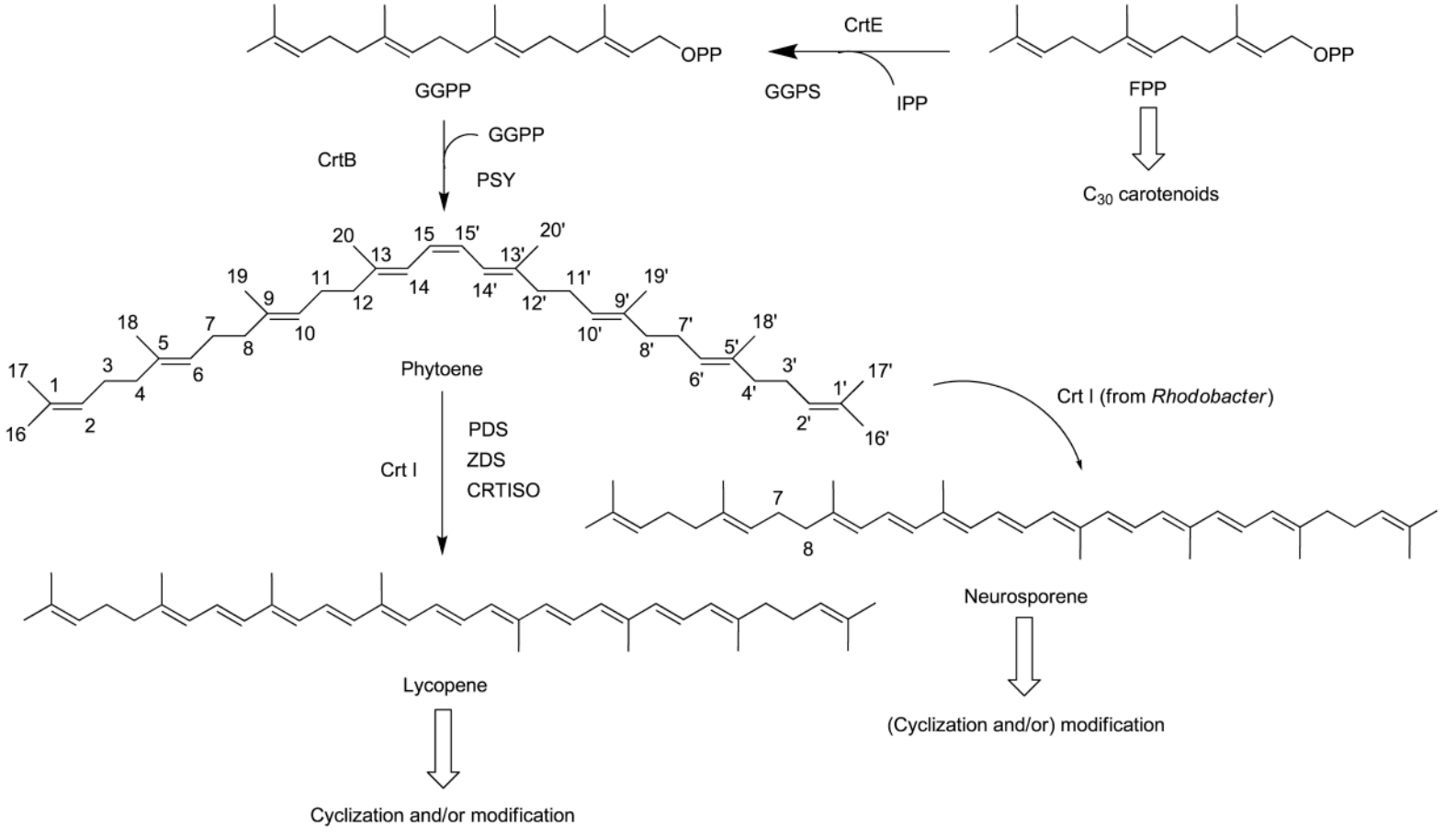
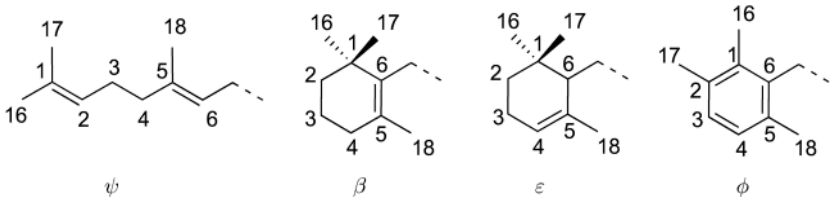


Figure 1 Biosynthesis of basic structures of carotenoids. The numbering scheme for the carotenoid skeleton is shown with phytoene. Abbreviations of bacterial and plant enzyme names are shown on the left (or above) and right (or below) of the arrows, respectively.



Scheme 1 Numbering scheme for dominant types of carotenoid end groups.

1.20.3 Biosynthesis of C_{30} Carotenoids

Some bacteria such as *Staphylococcus aureus* belonging to the class Bacilli, and methanotrophs, *Methylobacterium rhodium* (formerly *Pseudomonas rhodos*), and *Methylomonoas* sp. can produce unusual C_{30} carotenoids that possess a skeleton composed of 30 carbon atoms.[34–36] In order to synthesize the C_{30} carotenoids, two molecules of FPP instead of GGPP are condensed into 4,4′-diapophytoene ((15*Z*)-7,8,11,12,7′,8′,11′,12′-octahydro-4,4′-diapo-ψ,ψ-carotene)[12] with 4,4′-diapophytoene synthase (CrtM) whose amino acid sequence is similar to that of CrtB, as first shown by Wieland *et al.*[37] (**Figure 2**). This colorless hydrocarbon is desaturated to 4,4′-diaponeurosporene ((all-*E*)-7,8-dihydro-4,4′-diapo-ψ,ψ-carotene) and further to 4,4′-diapolycopene (4,4′-diapo-ψ,ψ-carotene), through the sequential desaturation and 15-*cis*- to *trans*-isomerization reactions catalyzed by 4,4′-diapophytoene desaturase (CrtN), which is a homologous protein with CrtI (**Figure 2**).[36,37] 4,4′-Diaponeurosporene or 4,4′-diapolycopene are oxidized without cyclization in such C_{30} carotenoid-synthesizing bacteria. In *Methylomonoas* sp. strain 16a (ATCC PTA-2402), 4,4′-diapolycopene has been shown to be oxidized to the corresponding terminal aldehyde with CrtNb and further to the acid with an aldehyde dehydrogenase (Ald) (**Figure 2**).[36] The generated dioic acid has been further esterified with β-D-glucose in this bacterium, while a gene for the reaction has not yet been identified.[34,36] *M. rhodium* is thought to have a similar carotenoid metabolism and genes. It is also thought that in *S. aureus* not 4,4′-diapolycopene but 4,4′-diaponeurosporene is oxidized to form 4,4′-diaponeurospoeren-4-oic acid by way of the corresponding monoaldehyde.[36] Recently, Shindo *et al.*[38,39] showed that marine bacteria *Rubritalea squalenifaciens*, belonging to Verrucomicrobia, which is a recently discovered bacterial phylum, and *Planococcus maritimus*, belonging to the same class (Bacilli) as *S. aureus*, produced new types of C_{30} carotenoids, diapolycopenedioic acid, xylosyl ester, and methyl 1-β-D-glucosyloxy-3,4-didehydro-1,2-dihydro-8′-apo-ψ-caroten-8′-oate, respectively. C_{30} carotenoids may not be 'unusual carotenoids' in the world of carotenogenic bacteria.

1.20.4 Biosynthesis of C_{40} Carotenoids

1.20.4.1 Photosynthetic Bacteria

Purple photosynthetic bacteria of the genera *Rhodobacter, R. capsulatus*, and *R. sphaeroides*, which belong to the class Alphaproteobacteria, were the target of a molecular approach to elucidating carotenoid biosynthesis, since they had been found to possess a carotenoid biosynthesis gene cluster. The first DNA sequence of a carotenoid biosynthesis gene cluster was reported in *R. capsulatus* in 1989.[14] In the genus, the neurosporene which is generated with the distinct CrtI[30,40] is hydrated to 1-hydroxyneurosporene with neurosporene 1,2-hydratase (CrtC) without its terminal cyclization, and converted to spheroidene (1-methoxy-3,4-didehydro-1,2,7′,8′-tetrahydro-ψ,ψ-carotene), a major xanthophyll, by the action of 1-hydroxyneurosporene 3,4-desaturase (CrtD) and 1-hydroxyneurosporene methylase (CrtF) (**Figure 3**).[14] Spheroidenone and hydroxyspheroidene (a minor component) are also detected in the *Rhodobacter* strains,[12] which are considered to be synthesized with spheroidene ketolase (CrtA) and with CrtC, respectively, as shown in **Figure 3**.

Another purple photosynthetic bacterium, *Rubrivivax gelatinosus*, synthesizes not only spheroidene and hydroxyspheroidene but also spirilloxanthin (1,1′-dimethoxy-3,4,3′4′-tetradehydro-1,2,1′,2′-tetrahydro-ψ,ψ-carotene) (**Figure 3**).[41] Spirilloxanthin has also been isolated from other purple phototrophic bacteria such as

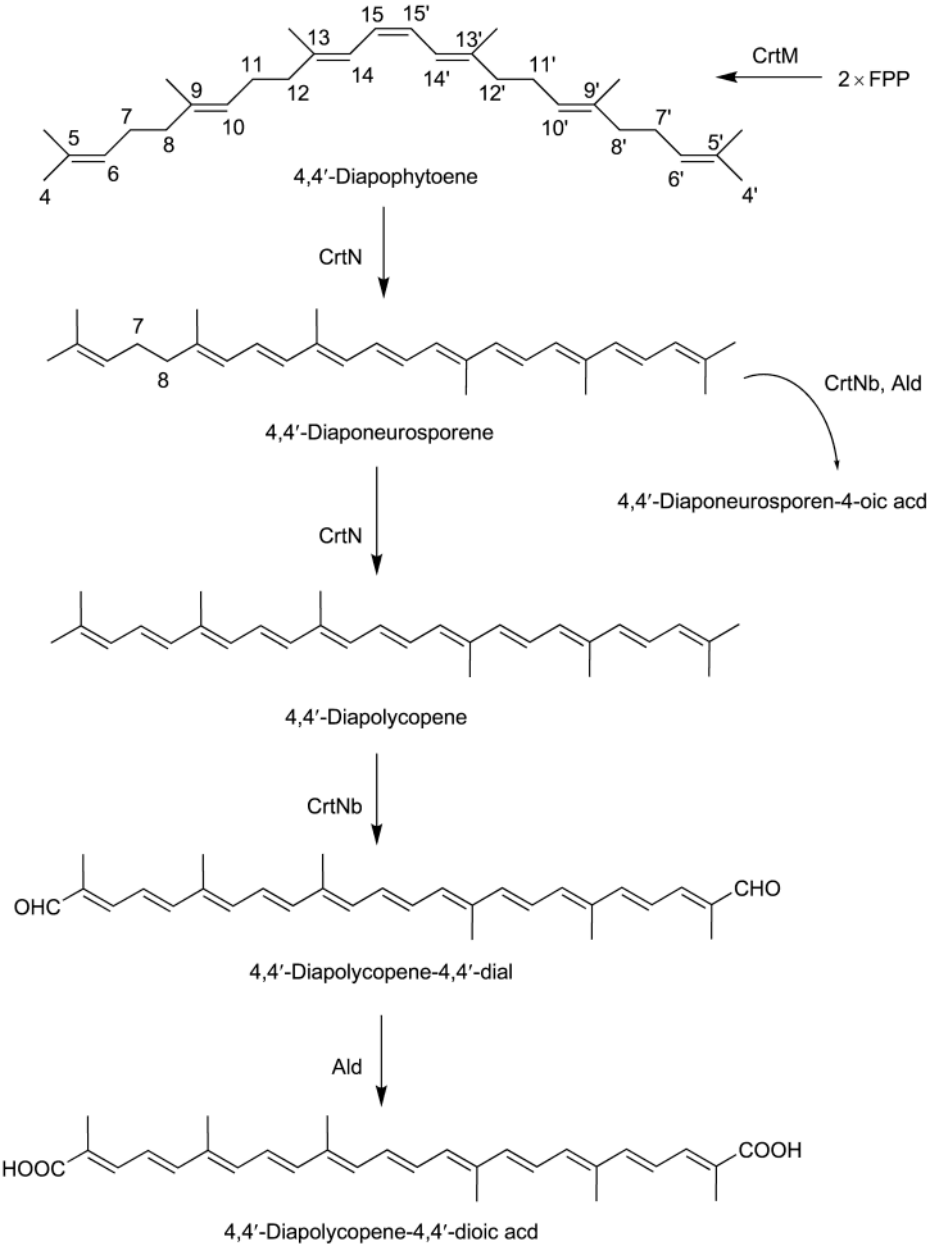


Figure 2 Biosynthesis of untypical carotenoids with the C_{30} skeleton in bacteria such as *Staphylococcus aureus* and *Methylomonoas* sp.

Rhodospirillum rubrum and *Rhodomicrobium vannielii*.[12,42] The *crtI* gene product isolated from *R. gelatinosus* has been shown to catalyze the formation of lycopene as well as neurosporene, indicating that the *R. gelantinosus* CrtI is bifunctional.[41] Based on this finding, the spirilloxanthin biosynthetic pathway has been proposed as shown in **Figure 3**.

1.20.4.2 Nonphotosynthetic Bacteria – Genus *Pantoea*

A carotenoid biosynthesis gene cluster similar to *Rhodobacter* has been found in the nonphotosynthetic bacteria *Pantoea ananatis* (reclassified from *Erwinia uredovora*)[26] and *Pantoea agglomerans* (*Erwinia herbicola*),[43]

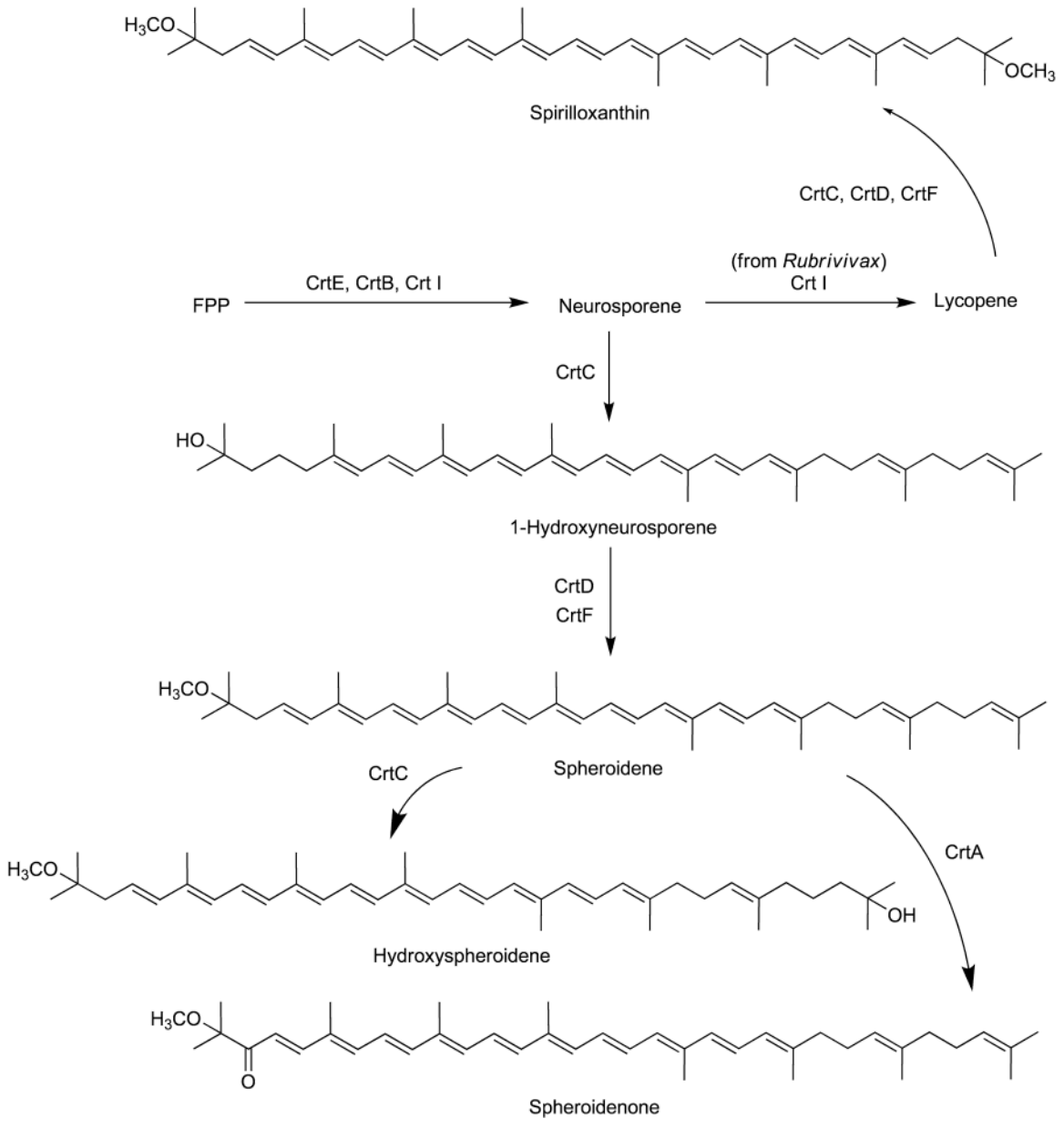


Figure 3 Biosynthesis of acyclic carotenoids in the photosynthetic bacteria *Rhodobacter capsulatus, Rhodobacter sphaeroides*, and *Rubrivivax gelatinosus*. The CrtI enzyme from *Rubrivivax gelatinosus* is bifunctional, producing both neurosporene and lycopene.

which belong to the family Enterobacteriaceae of class Gammaproteobacteria (the same family as *E. coli*). Functional analysis in *E. coli* of the individual genes in the *P. ananatis* carotenogenic gene cluster with a determined DNA sequence revealed that *E. coli* contains six genes, required for synthesizing cyclic carotenoids with the β-ionone rings, for example, β-carotene and zeaxanthin ((3*R*,3′*R*)-β,β-carotene-3,3′-diol), which are also found in higher plants.[26] This gene cluster included three new genes named *crtX*, *crtY*, and *crtZ*.[26] It has been found in the *Pantoea* strains that lycopene is cyclized to β-carotene with lycopene β-cyclase (CrtY), and then hydroxylated to zeaxanthin by β-carotene 3,3′-hydroxylase (CrtZ),[26,44] as shown in **Figure 4**. Zeaxanthin is further converted to its diglucoside with glucosyltransferase (CrtX) (**Figure 4**).[26,45,46] A thermophilic bacterium *Thermus thermophilus* HB27, which accumulates fatty acid esters

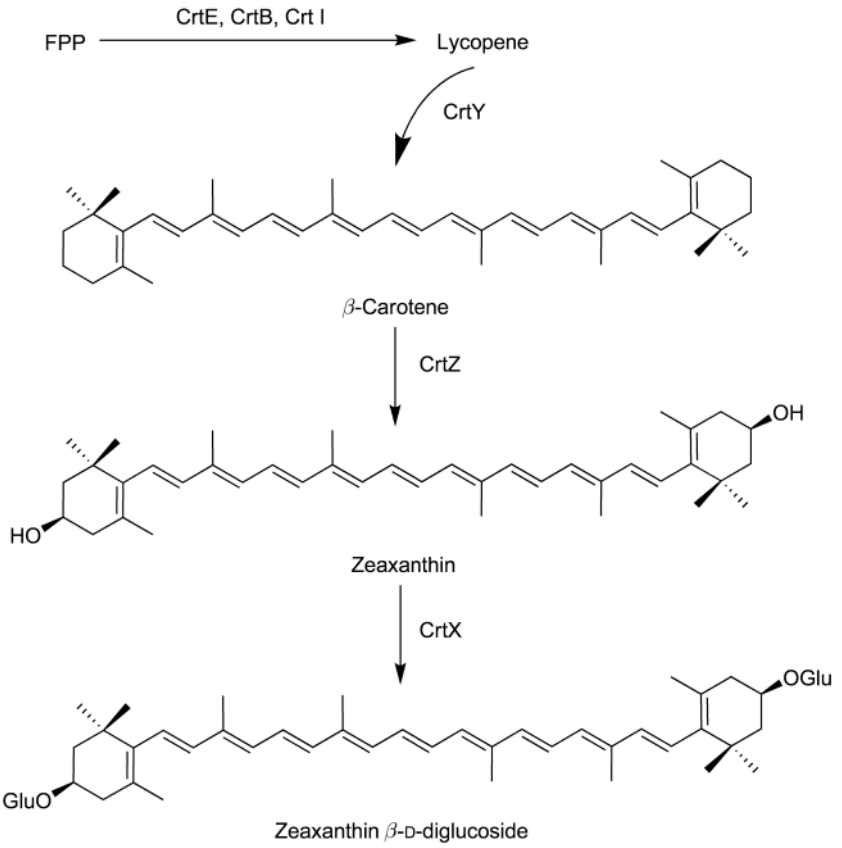


Figure 4 Biosynthesis of cyclic carotenoids with the β-rings in nonphototrophic bacteria, genus *Pantoea*. Glu, β-D-glucoside.

of zeaxanthin β-diglucoside in the membrane,[47] was shown to possess cytochrome P-450 (CYP175A1), whose structures are different from that of CrtZ, as a β-carotene 3,3′-hydroxylase.[48]

1.20.4.3 Nonphotosynthetic Bacteria – Genera *Paracoccus* and *Brevundimonas*

Some nonphotosynthetic bacteria that belong to the class Alphaproteobacteria, for example, genera *Paracoccus, Brevundimonas*, and *Bradyrhizobium*, have been shown to produce carotenoids including the 4-ketolated β-ionone rings (ketocarotenoids), for example, astaxanthin ((3*S*,3′*S*)-3,3′-dihydroxy-β,β-carotene-4,4′-dione) and canthaxanthin (β,β-carotene-4,4′-dione), which are commercially important carotenoids.[19,49,50] Several gene clusters encoding a series of carotenoid biosynthesis enzymes have been isolated from these bacteria and functionally analyzed. A gene cluster mediating ketocarotenoid formation was first isolated from a marine bacterium *Paracoccus* sp. strain N81106 (NBRC101723; formerly named *Agrobacterium aurantiacum*).[51] Functional analysis in *E. coli* of the individual genes in this cluster has shown the presence of a gene (named *crtW*) that encodes β-carotene (carotenoid) 4,4′-ketolase (4,4′-oxygenase), which in turn is responsible for the synthesis of canthaxanthin from β-carotene (**Figure 5**).[52] The *crtW* gene has also been found in carotenogenic gene clusters isolated from other *Paracoccus* strains,[52,53] *Brevundimonas* strains,[20,54] and a *Bradyrhizobium* strain.[19] It has been interestingly revealed that CrtW is able to introduce a keto group not only into the β-ionone rings but also into the 3(3′)-hydroxylated β-ionone rings at the 4(4′) positions.[51,55–58]

The *crtZ* gene frequently exists in carotenoid biosynthesis gene clusters derived from the bacteria that produce cyclic carotenoids (**Figure 6**).[19,20,51,53,54] CrtZ enzymes not only from *Paracoccus* and *Brevundimonas* strains but also from the *Pantoea* strains have been shown to accept canthaxanthin as well as β-carotene as

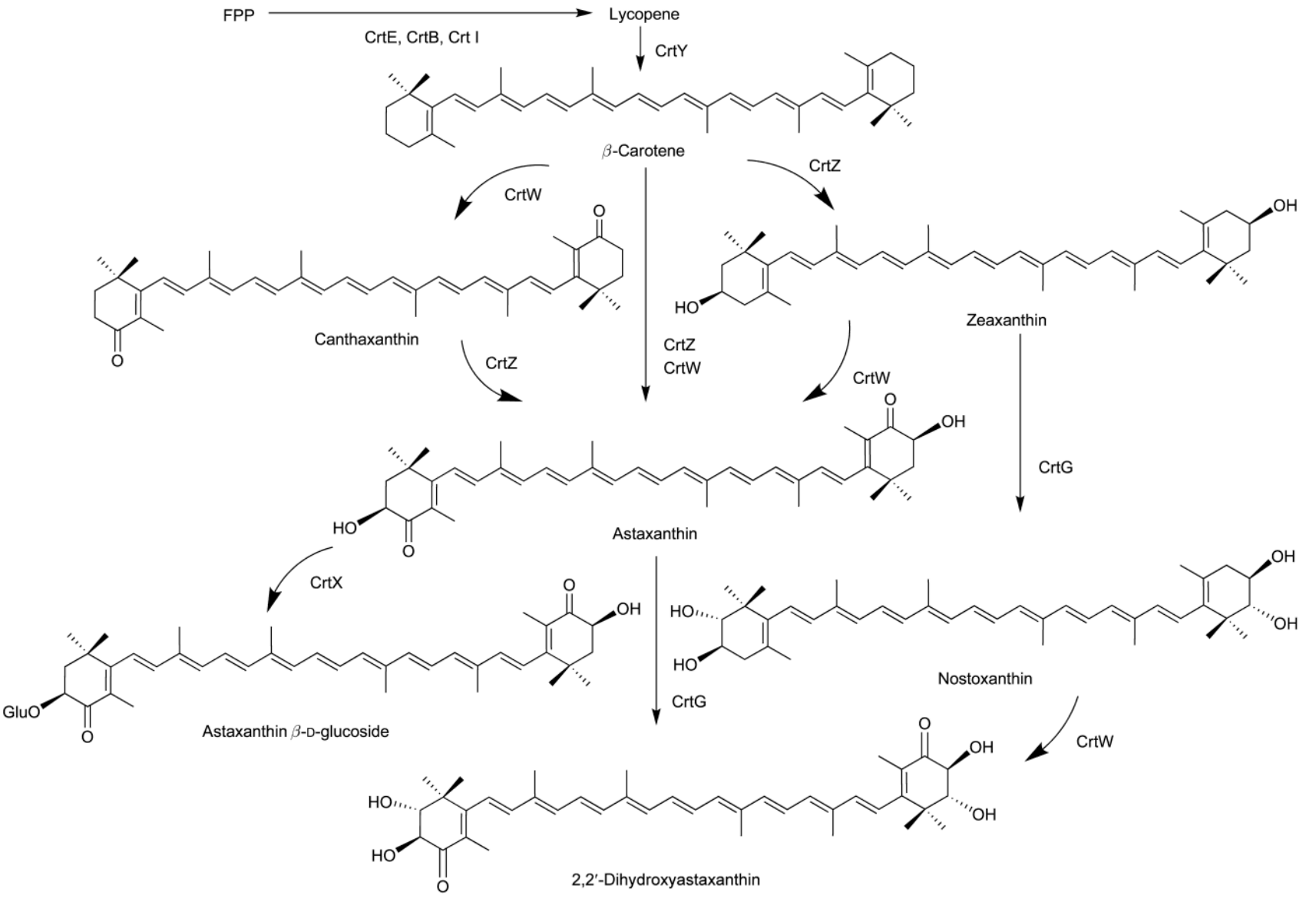


Figure 5 Biosynthesis of carotenoids in nonphototrophic bacteria, genera *Paracoccus* and *Brevundimonas*.

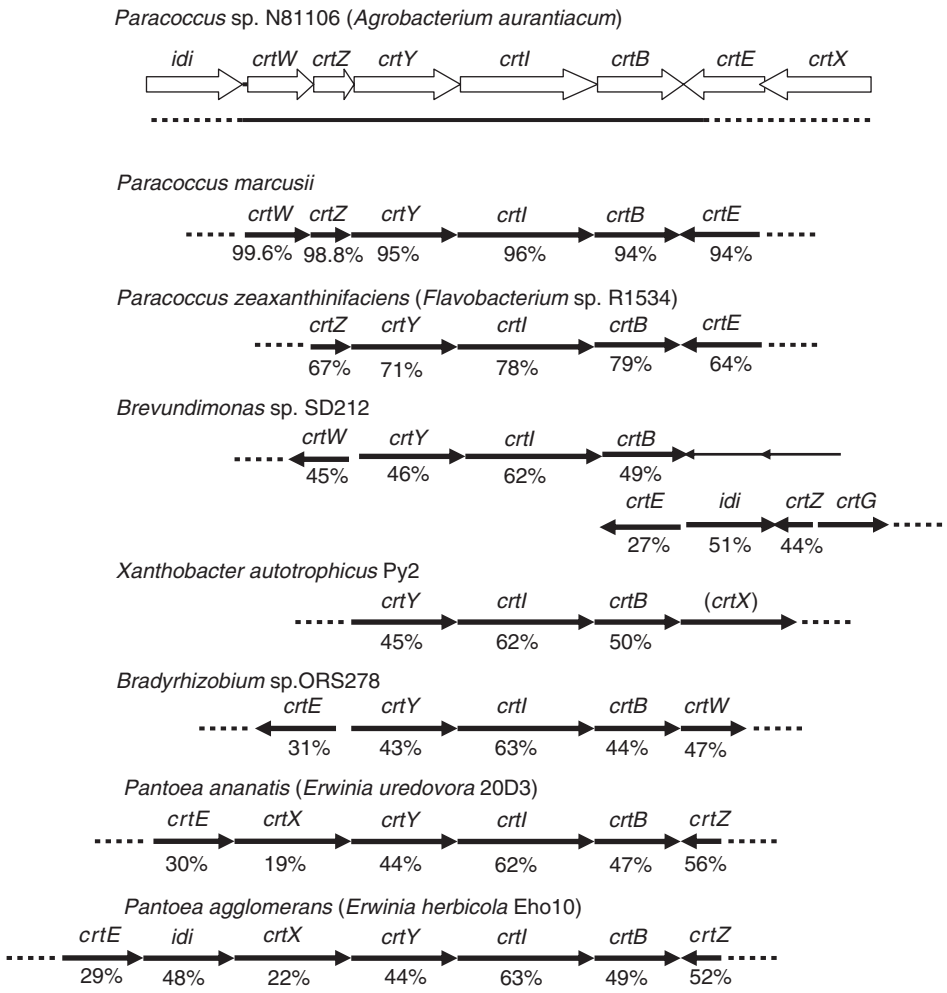


Figure 6 Organization of a complete carotenoid biosynthesis gene cluster from *Paracoccus* sp. N81106 and comparison of other clusters for the biosynthesis of cyclic carotenoids. The thick line shows the 5.4 kb fragment previously sequenced (accession no. D58420).[51] *Paracoccus marcusii* (Y15112), *Paracoccus zeaxanthinifaciens* (U62808),[62] *Brevundimonas* sp. SD212 (AB181388),[20] *Xanthobacter autotrophicus* Py2 (AF408848),[63] *Bradyrihizobium* sp. ORS278 (AF218415),[19] *Pantoea ananatis* (formerly classified as *Erwinia uredovora* 20D3; D90087),[26] and *Pantoea agglomerans* (formerly classified as *Erwinia herbicola* Eho10; M87280).[44] Arrows represent the orientation of genes. The values shown below the genes indicate the percentage of amino acid identity compared to the corresponding gene products of *Paracoccus* sp. N81106.[61] The zeaxanthin-hyperproducing bacterium formerly classified as *Flavobacterium* sp. R1534 has recently been reported to be the mutant of *P. zeaxanthinifaciens* ATCC21588, as taxonomically reevaluated by Berry *et al*.[64]

substrates; that is, they introduce a hydroxyl group not only into the β-ionone rings but also into the 4(4′)-ketolated β-ionone rings at the 3(3′) positions.[55,56,59] By the actions of CrtW and CrtZ, astaxanthin is synthesized from β-carotene by way of various intermediates (**Figure 5**). *In vitro* analysis using crude enzyme extracts from the recombinant *E. coli* cells expressing the corresponding genes has shown that CrtW and CrtZ are very likely to be α-ketoglutarate (2-oxoglutarate)-dependent dioxygenases.[55] Yokoyama *et al*.[60] found that *Paracoccus* sp. N81106 can synthesize astaxanthin β-D-glucoside. **Figure 5** show its proposed biosynthetic pathway which results from the fact that the N81106 carotenogenic cluster has been shown to include *crtX* (**Figure 6**).[61]

More recently, the carotenogenic gene cluster in *Brevundimonas* sp. strain SD212 (NBRC 101024) or *Brevundimonas vesicularis* DC263 has been shown to have a new gene (named *crtG*) for carotenoid 2,2′-hydroxylase that is responsible for the conversion from zeaxanthin to nostoxanthin ((2*R*,3*R*,2′*R*,3′*R*)-β,β-

carotene-2,3,2′,3′-tetrol) (**Figure 5**).[20,54] These bacterial strains can produce 2,2′-dihydroxy-astaxanthin as a minor carotenoid component.[50] A feasible biosynthetic pathway is shown in **Figure 5**.

1.20.4.4 Structure of Gene Clusters from Cyclic Carotenoid-producing Bacteria

The structure of the *Paracoccus* sp. N81106 gene cluster for ketocarotenoid formation was discovered in 1995, as mentioned above, and contains the 5.4 kb *Bam*HI fragment that includes the *crtW, crtZ, crtY, crtI*, and *crtB* genes (**Figure 6**).[51] However, since this DNA fragment was found not to contain all of the carotenogenic genes, Maruyama *et al.*[61] determined the nucleotide sequence of the 10 074 bp *Hin*dIII–*Sac*II fragment (accession no. AB206672), which included the 5.4 kb *Bam*HI fragment and its 5′- and 3′-flanking regions. As a result, a sequence for the IPP isomerase (*idi*; type 2) gene[65] was found in the 5′-flanking region of the 5.4 kb fragment, which was oriented in the same direction as *crtW, crtZ, crtY, crtI*, and *crtB*, as shown in **Figure 6**. Sequences for *crtE* and *crtX* were also in the 3′-flanking region of the 5.4 kb fragment, and were oriented in the opposite direction to *crtW, crtZ, crtY, crtI*, and *crtB* (**Figure 6**). The net length of the cluster containing the eight genes from *idi* to *crtX* was 8.2 kb. The results strongly suggest that the *Paracoccus* sp. N81106 carotenoid biosynthesis gene cluster is composed of eight genes, *idi, crtW, crtZ, crtY, crtI, crtB, crtE*, and *crtX*, which mediate the synthesis of astaxanthin β-D-glucoside via astaxanthin from FPP, as shown in **Figure 5**. **Figure 6** also shows a comparison of the organization of the carotenoid biosynthesis gene clusters that have been isolated from several bacteria-producing cyclic carotenoids belonging to class Alphaproteobacteria or Gammaproteobacteria. Interestingly, the *crtY, crtI*, and *crtB* genes occurred in this order and were oriented in the same direction, in all of the cyclic carotenoid biosynthesis gene clusters. Sequences for a type 2 IPP isomerase (*idi*) gene[66] were found in several of the gene clusters, that is, in the *Brevundimonas* sp. SD212 cluster, (the *idi* gene of which existed between the *crtZ* and *crtE* genes[20]), in the *P. agglomerans* cluster (the gene of which existed between the *crtE* and *crtX* genes[44]) and the *Paracoccus* N81106 cluster (**Figure 6**). The expression of the *idi* gene present in the clusters is likely to push a stream from IPP to dimethylallyl pyrophosphate (DMAPP), which is metabolized to FPP via geranyl pyrophosphate (GPP) by the endogenous FPP synthase enzyme (IspA).[67]

1.20.4.5 Cyanobacteria

Cyanobacteria (blue-green algae), from which chloroplasts in higher plants and algae originated via endosymbiosis, have the ability to perform oxygenic photosynthesis. Genome sequences of 34 cyanobacterial species are now available (CyanoBase, http://bacteria.kazusa.or.jp). Carotenoid biosynthesis genes are ordinarily not clustered in cyanobacteria, unlike the other carotenogenic bacteria mentioned above. *Synechocystis* sp. PCC 6803, whose genome sequence was first revealed by Kazusa DNA Research Institute in 1996, is the cyanobacterium, whose carotenogenic genes are most assigned to the carotenoid biosynthetic pathway.[68–71] Phytoene, which is synthesized from GGPP with CrtB, is converted to lycopene with phytoene desaturase (CrtP), ζ-carotene desaturase (CrtQ), and carotene isomerase (CrtH), which correspond to PDS, ZDS, and CRTISO in higher plants, respectively (**Figure 7**).[70,71] Lycopene is cyclized to β-carotene via the monocyclic intermediate γ-carotene (β,ψ-carotene) with a cyclase that remains unclear in cyanobacteria. Takaichi *et al.*[72] showed that *Synechocystis* sp. PCC 6803 produces myxol ((3*R*,2′*S*)-3′,4′-didehydro-1′2′-dihydro-β,ψ-carotene-3,1′2′-triol) 2′-dimethylfucoside (36% of the total), β-carotene (26%), echinenone (β,β-caroten-4-one) (18%), and zeaxanthin (14%) as major carotenoids, besides minor components of deoxymyxol 2′-dimethylfucoside (1%) and 3′--hydroxyechinenone (4%); their structures are shown in **Figure 7**. This cyanobacterium has been shown to have a new gene: β-carotene ketolase (named *crtO*) which converts β-carotene to echinenone[68] and a β-carotene hydroxylase (named *crtR*) which converts β-carotene to zeaxanthin via β-cryptoxanthin ((3*R*)-β,β-caroten-3-ol).[69] CrtO and CrtR are structurally different enzymes from the corresponding proteins (CrtW and CrtZ) that are typically present in other carotenogenic bacteria. The function of the CrtO and CrtR enzymes can be assigned to the carotenoid biosynthetic pathway, as shown in **Figure 7**. This CrtR seems to catalyze the formation of myxol 2′-dimethylfucoside from deoxymyxol 2′-dimethylfucoside, in addition to its function as β-carotene hydroxylase.

Other cyanobacteria whose genome sequences have been elucidated, *Anabaena* (also known as *Nostoc*) sp. PCC 7120 and *Nostoc punctiforme* PCC 73102 (ATCC 29133), synthesize β-carotene (62 and 45%), echinenone (25 and 17%), canthaxanthin (1 and 13%), myxol 2′-fucoside (8 and 11%), and 4-ketomyxol 2′-fucoside (4 and 13%).[73]

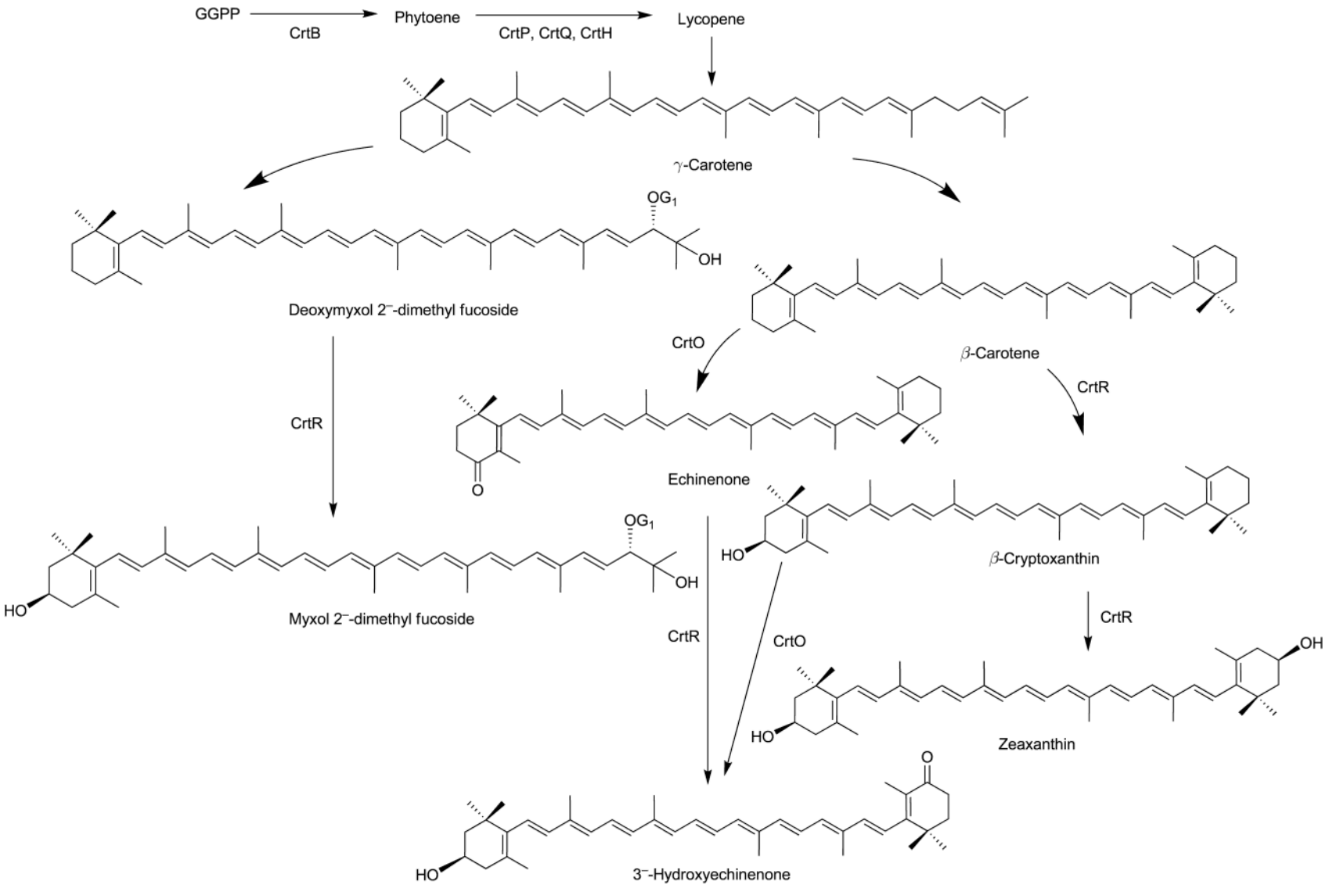


Figure 7 Biosynthesis of carotenoids in cyanobacterium *Synechocystis* sp. PCC 6803. G_1, 2,4-di-*O*-methyl-α-L-fucoside.

Anabaena sp. PCC 7120 carries both the *crtO* and *crtW* forms of the β-carotene ketolase gene,[74] whereas *N. punctiforme* PCC 73102 carries no *crtO* genes but two forms of the *crtW* β-carotene ketolase gene (*crtW38* and *crtW148*).[75] Both the cyanobacterial strains also possess CrtR. Based on results achieved so far,[58,74,76] we can assign the catalytic functions of the CrtO, CrtW, and CrtR enzymes to these cyanobacteria, as shown in **Figure 8**. Makino *et al.*[58] suggested that the CrtR enzymes from *Anabaena* sp. PCC 7120 and another *Anabaena* strain have

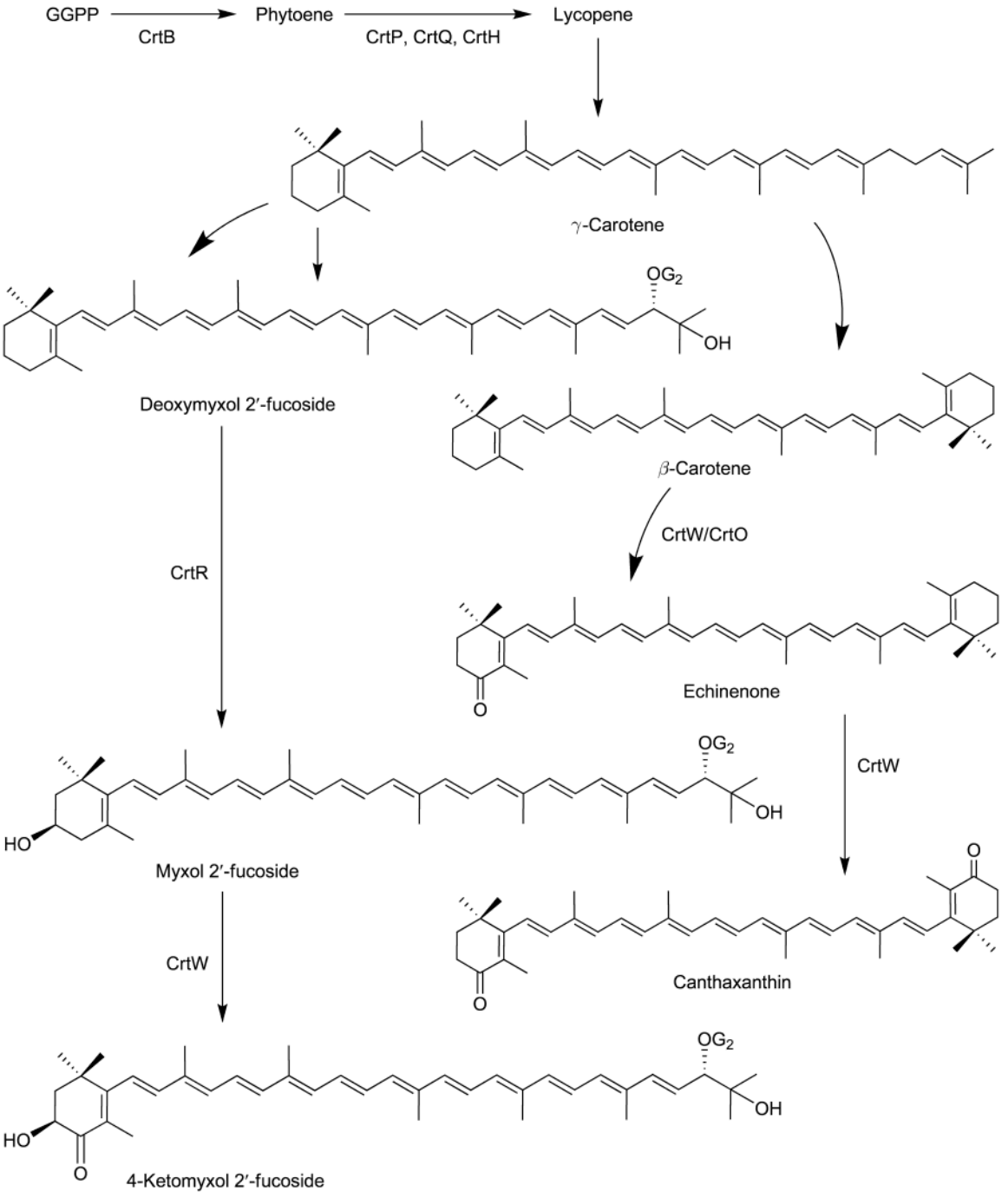


Figure 8 Biosynthesis of carotenoids in cyanobacteria *Anabaena* (*Nostoc*) sp. PCC 7120 and *Nostoc punctiforme* PCC 73102. Potential catalytic functions for the carotenoid ketolase and hydroxylase enzymes are indicated. The latter cyanobacterium carries two *crtW* genes (*crtW38* and *crtW148*); CrtW148 is very likely to work efficiently in the pathway. CrtO exists only in *Anabaena* sp. PCC 7120.

virtually no ability to accept β-carotene and canthaxanthin as their substrates. The CrtR enzymes from the *Anabaena* and *Nostoc* species are very likely to concentrate on converting deoxymyxol 2′-fucoside into myxol 2′-fucoside,[76] indicating that, in general, *crtR* codes for 3-deoxymyxol glycoside 3-hydroxylase rather than β-carotene 3,3′-hydroxylase. On the other hand, the *Anabaena* PCC 7120 CrtW and *N. punctiforme* PCC 73102 CrtW148 enzymes seem to be bifunctional enzymes, which have been postulated not only to catalyze the synthesis of 4-ketomyxol 2′-fucoside from myxol 2′-fucoside[74] but also to function as β-carotene 4,4′-ketolase.[58]

As mentioned, cyanobacteria typically synthesize myxol glycosides (also called myxoxanthophyll) and their 4-ketolated derivative in addition to β-carotene and its 3(3′)-hydroxylated and/or 4(4′)-ketolated derivatives. *Anabaena variabilis* ATCC 29413 has been shown to exceptionally have the free form of myxol without myxol glycosides.[77] Some marine bacteria that belong to the family Flavobacteriaceae have also been shown to produce the free form of myxol without its glycosides.[78,79] A lycopene β-monocyclase gene (named *crtYm*), which mediates the conversion from lycopene to γ-carotene, has been isolated from one of these bacteria.[78]

1.20.5 Pathway Engineering for Useful Carotenoid Formation in *Escherichia coli*

So far, the biosynthesis of the basic structures of carotenoids and typical examples of subsequent xanthophyll formation with carotenogenic bacteria or cyanobacteria, on the level of the reactions with enzymes as gene products, have been shown. The carotenoid biosynthesis genes of bacterial origin that have been described here can be functionally expressed as *E. coli*, the most convenient and amenable microorganism for biotechnological applications, enabling the use of the pathway engineering approach to synthesize a variety of new or minor carotenoids as well as industrially important carotenoids such as astaxanthin and lycopene. Since *E. coli*, like other bacteria, has a scanty amount of GGPP for carotenoid formation, we have to introduce in *E. coli* a series of biosynthesis genes from FPP to a desired carotenoid.[15,26] For example, when the *crtE* and *crtB* genes and the *crtE, crtB*, and *crtI* genes derived from *P. ananatis* are introduced into *E. coli* and expressed there, the resultant transformants have the ability to produce phytoene and lycopene as the predominant carotenoids, respectively (see **Figure 1**). It has been similarly reported that the expression in *E. coli* of the *crtE, crtB, crtI, crtY*, and *crtZ* genes from *P. ananatis* (for zeaxanthin formation) (**Figure 4**) and the *crtW* gene from *Paracoccus* sp. N81106 (**Figure 5**) resulted in the synthesis of astaxanthin (50% of the total) and its intermediates.[51] The levels of carotenoids produced with these recombinant *E. coli* cells are approximately 0.2–1 mg g^{-1} dry weight.

1.20.5.1 Pathway Engineering for Increasing the Intracellular Concentration of Farnesyl Diphosphate

When carotenoid biosynthesis genes starting from FPP are introduced in *E. coli*, the amount of carotenoids produced with the recombinant *E. coli* cells is far from that required for practical use; it is difficult to exceed 1 mg g^{-1} dry weight. In order to overcome this problem, pathway engineering approaches have been performed to increase the intracellular concentration of FPP. Two types of Idi enzymes, type 1 (eukaryote) and type 2 (prokaryote),[65] are known to date. When the *idi* (type 1) gene derived from green alga *Haematococcus pluvialis* or yeast *Saccharomyces cerevisiae* was introduced into *E. coli* and expressed there, the FPP content was estimated to increase by several times; for example, *E. coli* expressing the *P. ananatis crtE, crtB, crtI*, and *crtY* genes with the *H. pluvialis idi* gene produced 1.3 mg g^{-1} dry weight of β-carotene.[80] Similarly, the expression of the type 2 *idi* gene in *E. coli* was shown to be effective to increase FPP content.[81] When the intrinsic 1-deoxy-D-xylulose 5-phosphate synthase (*dxs*) or 1-deoxy-D-xylulose 5-phosphate reductoisomerase (*dxr*) gene in the nonmevalonate (2-*C*-methyl-D-erythritol-4-phosphate (MEP)) pathway was overexpressed in *E. coli*, the resultant transformant was shown to increase FPP amounts (1.6 mg g^{-1} dry weight of carotenoids at the highest level).[82] Yuan *et al.*[83] have shown that the simultaneous overexpression of *dxs, idi*, and other genes resulted in producing higher levels (6 mg g^{-1} dry weight) of β-carotene. On the other hand, the introduction of heterologous mevalonate pathway genes in *E. coli* has been described as improving the productivity of carotenoids or sesquiterpenes that are synthesized from FPP.[81,84–89] These studies showed that the production levels of the desired isoprenoids are significantly increased with addition of exogenous substrates such as D-mevalonolactone (D-mevalonate) in the culture medium. Harada *et al.*[81] produced 12.5 mg g^{-1} dry weight

of lycopene in *E. coli* by expressing the *P. ananatis crtE, crtB*, and *crtI* genes using lithium acetoacetate, which is a cheaper substrate than D-mevalonolactone, by simultaneously expressing the rat acetoacetate-CoA ligase gene and the *S. cerevisiae* type 1 *idi* genes as well as the *Streptomyces* sp. CL190 mevalonate pathway genes, including type 2 *idi*.[65,84]

1.20.5.2 Pathway Engineering for the Production of Novel or Rare Carotenoids

All the carotenoids described in **Figures 2–5** can nowadays feasibly be produced in *E. coli* using pathway engineering techniques. Following, pathway engineering in *E. coli* for the production of novel carotenoids as well as naturally rare or minor carotenoids that have not been described above are shown.

Escherichia coli expressing the *crtE, crtB*, and *crtI* genes from *P. ananatis* or *P. agglomerans* synthesizes lycopene.[44,51] By introducing a combination of the genes for acyclic carotenoid-modifying enzymes from *R. capsulatus* or *R. sphaeroides* (**Figure 3**) into this transformant, it is feasible to produce a rare carotenoid that has been modified without cyclization, for example, the expression of the *R. capsulatus crtC* gene in the lycopene-producing *E. coli* resulted in the formation of 1,1′-dihydroxylycopene (1,2,1′,2′-tetrahydro-ψ,ψ-carotene-1,1′-diol), which is a minor carotenoid found in a photosynthetic bacteria (**Figure 9**).[90] Similarly, *E. coli* expressing the *crtE, crtB, crtI*, and *crtY* genes from *P. ananatis* or *P. agglomerans* synthesizes β-carotene. It is possible to desaturate the β-rings to form isorenieratene (ϕ,ϕ-carotene) with a ϕ end group (an aromatic end group) by introducing a β-carotene desaturase gene (named *crtU*) into the β-carotene-accumulating *E. coli*, which is carried in actinobacteria such as *Streptomyces griseus* and *Brevibacterium linens* (**Figure 9**).[12,91] It is also feasible to produce various β-carotene derivatives by using various combinations of *crtW, crtZ, crtX*, and *crtG*. For example, the expression in the β-carotene-accumulating *E. coli* of the *Paracoccus* sp. N81106 *crtW* and *Brevundimonas* sp. SD212 *crtG* genes, and the *P. ananatis crtZ, crtX*, and the *crtG* resulted in the formation of the novel carotenoids 2,2′-dihydroxycanthaxanthin ((2*S*2′*S*)-2,2′-dihydroxy-β, β-carotene-4,4′-dione)[20] and caloxanthin ((2*R*,3*R*,3′*R*)-β, β-carotene-2,3,3′-triol) β-D-glucoside (H. Harada, K. Shindo, N. Misawa, in preparation), respectively (**Figure 9**). Takaichi *et al.*[92] previously showed that *E. coli* carrying the *crtE, crtB*, and *crtY* genes from *P. ananatis* in addition to the *crtI* gene from *R. capsulatus* was able to cyclize both of the end groups of neurosporene to synthesize a new carotenoid, 7,8-dihydro-β-carotene (7,8-dihydro-β, β-carotene), which was hydroxylated to form parasiloxanthin ((3*R*,3′*R*)-7,8-dihydro-β, β-carotene-3,3′-diol), previous a rare carotenoid isolated from the fin and skin of the Japanese common catfish (**Figure 9**).[93]

Schmidt-Dannert *et al.*[94] modified the catalytic function of CrtI by *in vitro* evolution to further desaturate lycopene to the fully conjugated acyclic carotenoid 3,4,3′4′-tetradehydrolycopene by way of 3,4-didehydrolycopene, which was converted to torulene (3′,4′-didehydro-β,ψ-carotene) with CrtY. The torulene generated was shown to be modified to the corresponding xanthophylls with CrtU, CrtO, CrtZ, or CrtZ and CrtX.[95] Umeno and Arnold[96] showed that the expression of the *P. ananatis crtE* and *S. aureus crtM* genes in *E. coli* resulted in the formation of C_{35} carotenes by the condensation of FPP and GGPP, which was subjected to desaturations with CrtN or CrtI.

1.20.6 Pathway Engineering for Useful Carotenoid Formation in Higher Plants

1.20.6.1 Biosynthesis of Carotenoids in Higher Plants

Figure 10 shows the biosynthetic pathway of carotenoids in the leaves of higher plants on the level of reactions with enzymes as gene products. In higher plants, lycopene is competitively cyclized to form β-carotene with lycopene β-cyclase (LCY-b), which is a homolog of CrtY, or α-carotene with lycopene ε-cyclase (LCY-e) and LCY-b. β-Carotene is usually one of the dominant carotenoids in plants, and can be metabolized to zeaxanthin by the action of a β-carotene 3,3′-hydroxylase, BHY[97] or presumably BHY has substantial sequence homology to the bacterial corresponding CYP97A.[98,99] enzyme CrtZ.[97] Zeaxanthin is metabolized to violaxanthin via antheraxanthin, and to neoxanthin, as shown in **Figure 10**. The ε-ring of α-carotene is hydroxylated to form lutein with ε-carotene hydroxylase (EHY; CYP97C).[99,100] Family CYP97 containing CYP97A and CYP97C appears to be related to CYP175A1 from the thermophilic bacterium *T. thermophilus*.[98]

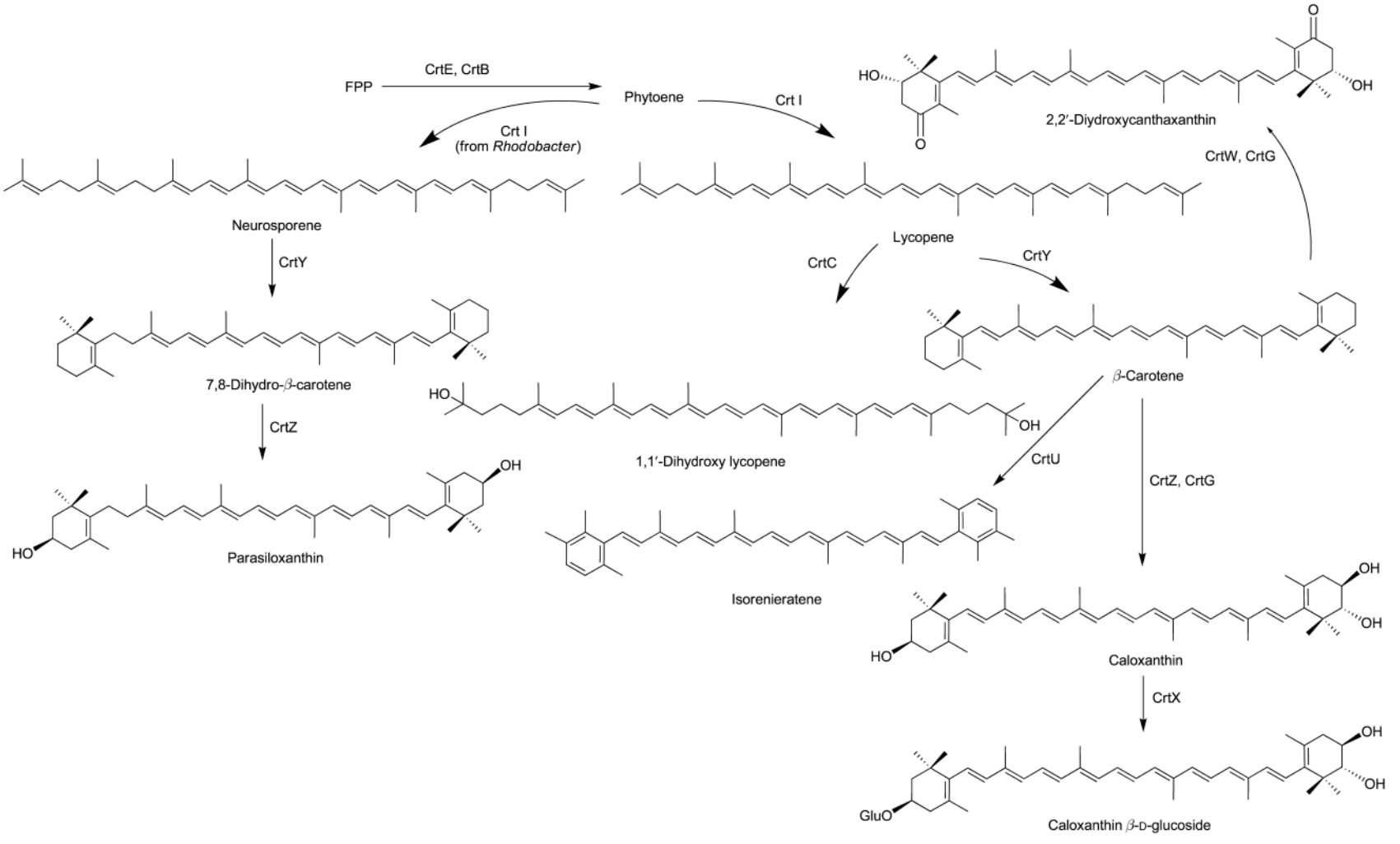


Figure 9 Biosynthesis of new or naturally rare carotenoids with recombinant *Escherichia coli* cells, in which various combinations of bacterial carotenoid biosynthesis genes are introduced.

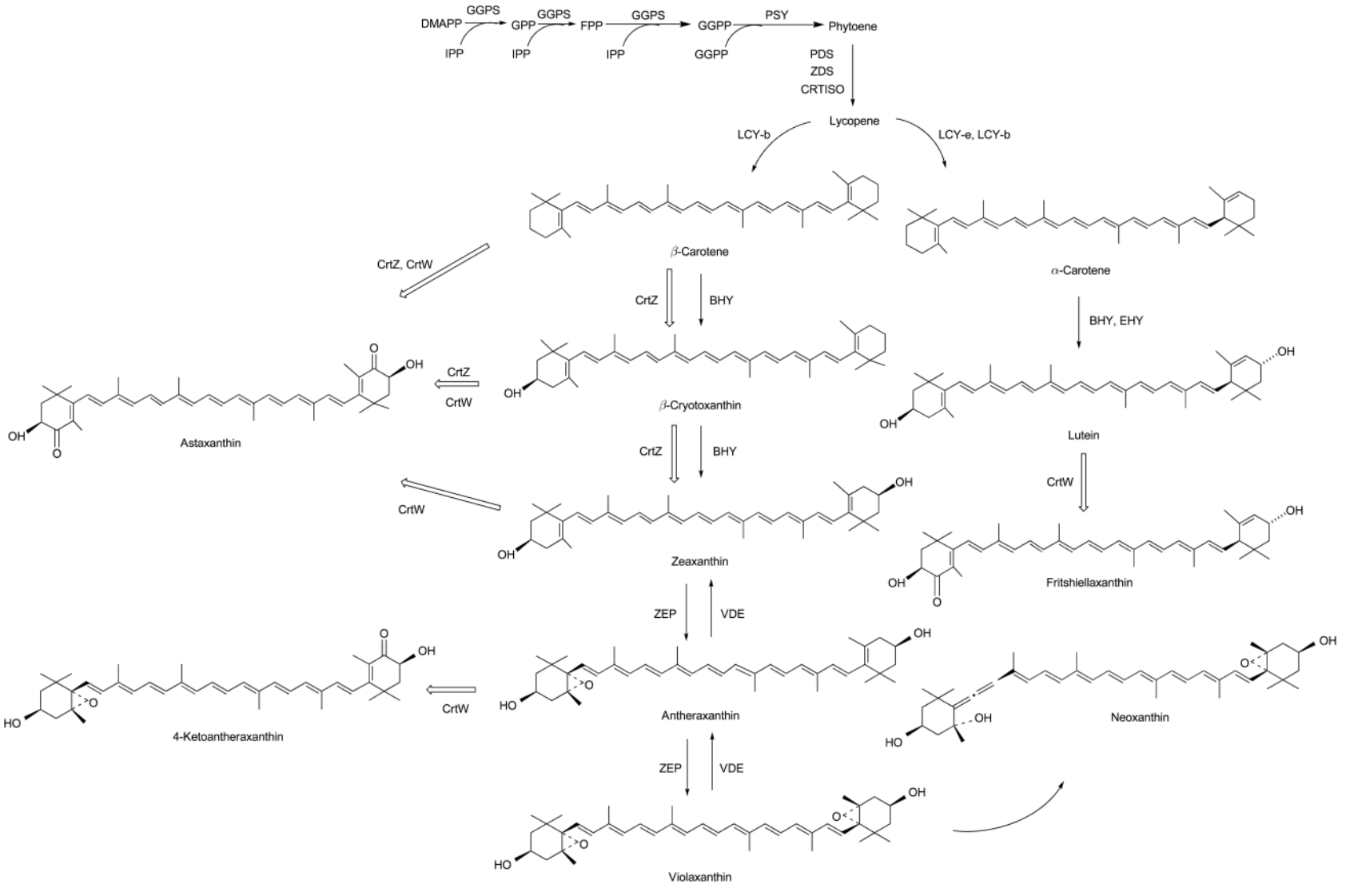


Figure 10 Biosynthesis of carotenoids in higher plants, and catalytic functions of CrtW and CrtZ introduced and expressed in tobacco leaves. Structures of the carotenoids generated as the result of the CrtZ and CrtW expression in tobacco leaves are also shown in the figure. GGPS, GGPP synthase; PSY, phytoene synthase; PDS, phytoene desaturase; ZDS, ζ-carotene desaturase; CRTISO, carotene isomerase; LCY-b, lycopene β-cyclase; LCY-e, lycopene ε-cyclase; BHY, β-carotene hydroxylase; ZEP, zeaxanthin epoxidase; VDE, violaxanthin de-epoxidase; EHY, ε-carotene hydroxylase.

1.20.6.2 Pathway Engineering for Increasing Carotenoid Content in Higher Plants

In order to increase carotenoid content and/or produce a desired carotenoid, pathway engineering approaches have been performed using higher plants as hosts.[101,102] Carotenoid biosynthesis (*crt*) genes that originate from carotenogenic bacteria such as the *Pantoea, Paracoccus*, and *Brevundimonas* strains have frequently been introduced in higher plants.[103–111] Since plastids containing chloroplasts and chromoplast are the location of carotenoid formation in plants, a transit peptide is needed to target a bacterial gene product into plant plastids. The *P. ananatis crtI* gene, which has been fused at the N-terminus to the transit peptide sequence of the RuBisCO small subunit from pea plants,[112] was first introduced into tobacco (*Nicotiana tabacum*) under the control of the cauliflower mosaic virus (CaMV) 35S promoter.[103] The resultant transgenic tobacco plants changed carotenoid composition in the leaves and showed resistance to bleaching herbicides such as norflurazon and diflufenican, which are inhibitors of the plant PDS enzyme.[103,104] The *crtI* gene has also been functionally expressed in tomato fruit to increase the β-carotene content (to three times that of a nontransformed control).[106] The step from GGPP to phytoene has been shown to be very important in increasing carotenoid levels, that is, the expression of the phytoene synthase gene (*crtB*) from *P. ananatis* has been found to elevate total carotenoid content in various plant tissues such as rapeseed,[105] tomato fruit,[107] potato tuber,[113] and flax seed.[110] Remarkably, Shewmaker *et al.*[105] introduced the *crtB* gene into oilseed rape plants (canola; *Brassica napus*) under the control of the seed-specific napin promoter, and consequently observed a dramatic increase in carotenoid level in the mature seeds; that is, the total amount of carotenoids, which contained β-carotene, α-carotene, and lutein, was elevated up to 50-fold (1.6 mg g^{-1} fresh weight at the most). It has also recently been shown that flax seed (linseed, *Linum usitatissimum*) expressing the *crtB* gene produced an up to 19-fold (160 μg g^{-1} fresh weight) increase in carotenoids.[110]

1.20.6.3 Pathway Engineering for the Production of Ketocarotenoids in Higher Plants

Biotechnological approaches for producing useful ketocarotenoids such as astaxanthin and canthaxanthin, whose global markets were US$240 and US$110 million, respectively, in 2007 according to a BCC research report (code: FOD025C), have been carried out using higher plants as hosts.[108,109,111,113–118] Since higher plants do not possess a β-carotene ketolase gene (except for *Adonis* plants),[119] at the least its introduction into plants and subsequent expression are necessary to confer the ability to produce ketocarotenoids. A β-carotene ketolase gene is initially introduced into model plants such as tobacco (*N. tabacum*) and *Arabidopsis* (*Arabidopsis thaliana*), and ketocarotenoid formation has been shown to be feasible in plant tissues: Mann *et al.*[114] introduced the green algal *H. pluvialis bkt1* gene (a homolog of *crtW*)[120] into tobacco plants under the control of the tomato *pds* promoter. As a result, a transgenic tobacco produced ketocarotenoids (84% of total carotenoids) including 84 μg g^{-1} fresh weight of astaxanthin in the nectary tissue, although they were not accumulated by this transgenic plant in the leaf tissue. Stålberg *et al.*[115] introduced the *H. pluvialis bkt2* gene (a paralog of *bkt1*)[121] into *Arabidopsis* plants under the regulation of the *B. napus napA* promoter, and found the formation of ketocarotenoids such as canthaxanthin, adonirubin, and fritshiellaxanthin (4-ketolutein) (23% of total pigments) in the seeds. As for practical plants, there have been attempts to express several ketolase genes in several crops, that is, tomato,[108] potato,[113,116] carrot,[122] and rapeseed.[111] Specifically, Jayaraj *et al.*[122] expressed the *bkt1* gene in carrot roots using the double CaMV 35S promoter, and found that the transgenic roots were able to accumulate large amounts of ketocarotenoids (236 μg g^{-1} fresh weight; 68% of total carotenoids) including 91.6 μg g^{-1} fresh weight of astaxanthin. Recently, Hasunuma *et al.*[118] introduced chemically synthetic genes with the codon usage of rape plants encoding the CrtW and CrtZ proteins of *Brevundimonas* sp. SD212, which are a combination of the most efficient enzymes for synthesizing astaxanthin from β-carotene,[57] into the chloroplasts of tobacco plants by plastid transformation. It has consequently been revealed that the transformed plants can produce the largest amount of astaxanthin found so far as the predominant carotenoid in the leaves, without any significant damage to the plants: that is, astaxanthin levels reached 5.44 mg g^{-1} dry weight (74% of total carotenoids), which could correspond to 0.54 mg g^{-1} fresh weight based on 90% moisture level in the leaves. Surprisingly, almost all of the carotenoids generated in the leaves were ketocarotenoids (99%), which contained the ketolated products of lutein and antheraxanthin, fritshiellaxanthin (4-ketolutein), and 4-ketoantheraxanthin, respectively, in addition to astaxanthin.[123] The pathway for synthesizing these ketocarotenoids is shown in **Figure 10**. Moreover, the total carotenoid amount in the transplastomic tobacco was

2.1-fold higher than that of wild-type tobacco,[118] suggesting upregulation for carotenoid biosynthesis, probably due to the stream from the plant's original carotenoids to the ketocarotenoid formation. It is also likely that the multiple expression of the *crtB* as well as the *crtW* (or probably *bkt1*) and *crtZ* genes in a desired plant tissue is effective for producing huge amounts of ketocarotenoids including astaxanthin together with other commercially useful carotenoids.[111]

1.20.7 Conclusion

Over the past 20 years, biosynthesised genes for basic structures of carotenoids and subsequent diverse xanthophylls have been isolated from carotenogenic organisms, specifically bacteria and higher plants, and the carotenoid biosynthesis has been elucidated on the level of reactions with enzymes as gene products. Using the isolated biosynthesised genes, pathway engineering approaches have been performed in *E. coli* (and other microbes) and higher plants, using them as the hosts for the production of useful, new, or naturally rare carotenoids aiming to find ways to produce them efficiently. Industrial production of such carotenoids with recombinant bacteria or higher plants, or commercial cultivation of transgenic crops enriched in the desired carotenoids, should be realized in the near future by further efforts of biotechnologists and relevant workers to harvest the fruits of the great progress made in this field over the last 20 years.

Acknowledgment

Part of our work was supported by the New Energy and Industrial Technology Development Organization (NEDO), Japan.

Nomenclature

astaxanthin	3,3′-dihydroxy-β,β-carotene-4,4′-dione
canthaxanthin	β,β-carotene-4,4′-dione
lutein	β,ε-carotene-3,3′-diol
lycopene	ψ,ψ-carotene
zeaxanthin	β,β-carotene-3,3′-diol
α-carotene	β,ε-carotene
β-carotene	β,β-carotene
β-cryptoxanthin	β,β-caroten-3-ol
γ-carotene	β,ψ-carotene

References

1. W. Miki, *Pure Appl. Chem.* **1991**, *63*, 141–146.
2. H. Tatsuzawa; T. Maruyama; N. Misawa; K. Fujimori; M. Nakano, *FEBS Lett.* **2000**, *484*, 280–284.
3. E. Camera; A. Matrofrancesco; C. Fabbri; F. Daubrawa; M. Picardo; H. Sies; W. Stahl, *Exp. Dermatol.* **2009**, *18*, 222–231.
4. H. Nishino; M. Murakoshi; T. Ii; M. Takemura; M. Kuchide; M. Kanazawa; X. Y. Mou; S. Wada; M. Masuda; Y. Ohsaka; S. Yogosawa; Y. Satomi; K. Jinno, *Cancer Metastasis Rev.* **2002**, *21*, 257–264.
5. T. M. Vogt; S. T. Mayne; B. I. Graubard; C. A. Swanson; A. L. Sowell; J. B. Schoenber; G. M. Swanson; R. S. Greenberg; R. N. Hoover; R. B. Haynes; R. G. Ziegler, *Am. J. Epidemiol.* **2002**, *155*, 1023–1032.
6. S. Mannisto; S. A. Smith-Warner; D. Spiegelman; D. Albanes; K. Anderson; P. A. van den Brandt; J. R. Cerhan; G. Colditz; D. Feskanich; J. L. Freudenheim; E. Giovannucci; R. A. Gohldbohm; S. Graham; A. B. Miller; T. E. Rohan; J. Virtamo; W. C. Willett; D. J. Hunter, *Cancer Epidemiol. Biomarkers Prev.* **2004**, *13*, 40–48.
7. T. Iwamoto; K. Hosoda; R. Hirano; H. Kurata; A. Matsumoto; W. Miki; M. Kamiyama; H. Itakura; S. Yamamoto; K. Kondo, *J. Atheroscler. Thromb.* **2000**, *7*, 216–222.
8. F. J. Pashkow; D. G. Watumull; C. L. Campbell, *Am. J. Cardiol.* **2008**, *101*, 58D–68D.

9. M. Sugiura; M. Nakamura; K. Ogawa; Y. Ikoma; H. Matsumoto; F. Ando; H. Shimokata; M. Yano, *Br. J. Nutr.* **2008**, *100*, 1297–1306.
10. M. Mozaffarieh; S. Sacu; A. Wedrich, *Nutr. J.* **2003**, *2*, 20.
11. R. L. Roberts; J. Green; B. Jewis, *Clin. Dermatol.* **2009**, *27*, 195–201.
12. G. Britton; S. Liaaen-Jensen; H. Pfander, Eds., *Carotenoid Handbook*; Birkhauser Verlag: Basel, 2004.
13. U. Neudert; I. M. Martines-Ferez; P. D. Fraser; G. Sandmann, *Biochim. Biophys. Acta* **1998**, *1392*, 51–58.
14. G. A. Armstrong; M. Alberti; F. Leach; J. E. Hearst, *Mol. Gen. Genet.* **1989**, *216*, 254–268.
15. G. Sandmann; N. Misawa, *FEMS Microbiol. Lett.* **1992**, *90*, 253–258.
16. J. Ray; C. Bird; M. Maunders; D. Grierson; W. Schuch, *Nucleic Acids Res.* **1987**, *15*, 10587.
17. N. Misawa; M. R. Truesdale; G. Sandmann; P. D. Fraser; C. Bird; W. Schuch; P. M. Bramley, *J. Biochem.* **1994**, *116*, 980–985.
18. S. K. Math; J. E. Hearst; C. D. Poulter, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 6761–6764.
19. L. Hannibal; J. Lorquin; N. A. D'Ortoli; N. Garcia; C. Chaintreuil; C. Masson-Boivin; B. Dreyfus; E. Giraud, *J. Bacteriol.* **2000**, *182*, 3850–3853.
20. Y. Nishida; K. Adachi; H. Kasai; Y. Shizuri; K. Shindo; A. Sawabe; S. Komemushi; W. Miki; N. Misawa, *Appl. Environ. Microbiol.* **2005**, *71*, 4286–4296.
21. G. E. Bartley; P. V. Viitanen; I. Pecker; D. Chamovitz; J. Hirschberg; P. A. Scolnik, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 6532–6536.
22. I. Pecker; D. Chamovitz; H. Linden; G. Sandmann; J. Hirschberg, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4962–4966.
23. M. Albrecht; A. Klein; P. Hugueney; G. Sandmann; M. Kuntz, *FEBS Lett.* **1995**, *372*, 199–202.
24. H. Park; S. S. Kreunen; A. J. Cuttriss; D. DellaPenna; B. J. Pogson, *Plant Cell* **2002**, *14*, 321–332.
25. T. Isaacson; G. Ronen; D. Zamir; J. Hirschberg, *Plant Cell* **2002**, *14*, 333–342.
26. N. Misawa; M. Nakagawa; K. Kobayashi; S. Yamano; Y. Izawa; K. Nakamura; K. Harashima, *J. Bacteriol.* **1990**, *172*, 6704–6712.
27. P. D. Fraser; N. Misawa; H. Linden; S. Yamano; K. Kobayashi; G. Sandmann, *J. Biol. Chem.* **1992**, *267*, 19891–19895.
28. T. Tsuchiya; S. Takaichi; N. Misawa; T. Maoka; H. Miyashita; M. Mimuro, *FEBS Lett.* **2005**, *579*, 2125–2129.
29. S. Steiger; Y. Jackisch; G. Sandmann, *Arch. Microbiol.* **2005**, *184*, 207–214.
30. G. E. Bartley; P. A. Scolnik, *J. Biol. Chem.* **1989**, *264*, 13109–13113.
31. H. Linden; N. Misawa; D. Chamovitz; I. Pecker; J. Hirschberg; G. Sandmann, *Z. Naturforsch. C* **1991**, *46*, 1045–1051.
32. A. Raisig; G. Bartley; P. Scolnik; G. Sandmann, *J. Biochem.* **1996**, *119*, 559–564.
33. P. Stickforth; S. Steiger; W. R. Hess; G. Sandmann, *Arch. Microbiol.* **2003**, *179*, 409–415.
34. H. Kleinig; R. Schmitt; W. Meister; G. Englert; H. Thommen, *Z. Naturforcsch. C* **1979**, *34*, 181–185.
35. J. H. Marshall; G. J. Wilmoth, *J. Bacteriol.* **1981**, *147*, 900–913.
36. L. Tao; A. Schenzle; J. M. Odom; Q. Cheng, *Appl. Environ. Microbiol.* **2005**, *71*, 3294–3301.
37. B. Wieland; C. Feil; E. Gloria-Maercker; G. Thumm; M. Lechner; J. M. Bravo; K. Poralla; F. Gotz, *J. Bacteriol.* **1994**, *176*, 7719–7726.
38. K. Shindo; E. Asagi; A. Sano; Y. Hotta; N. Minemura; K. Mikami; E. Tamesada; N. Misawa; T. Maoka, *J. Antibiot.* **2008**, *61*, 185–191.
39. K. Shindo; M. Endo; Y. Miyake; K. Wakasugi; D. Morritt; P. M. Bramley; P. D. Fraser; H. Kasai; N. Misawa, *J. Antibiot.* **2008**, *61*, 729–735.
40. H. P. Lang; R. J. Cogdell; A. T. Gardiner; C. N. Hunter, *J. Bacteriol.* **1994**, *176*, 3859–3869.
41. J. Harada; K. V. P. Nagashima; S. Takaichi; N. Misawa; K. Matsuura; K. Shimada, *Plant Cell Physiol.* **2001**, *42*, 1112–1118.
42. T. W. Goodwin; H. G. Osman, *Biochem. J.* **1954**, *56*, 222–230.
43. B. S. Hundle; P. Beyer; H. Kleinig; G. Englert; J. E. Hearst, *Photochem. Photobiol.* **1991**, *54*, 89–93.
44. B. S. Hundle; M. Alberti; V. Nievelstein; P. Beyer; H. Kleinig; G. A. Armstrong; D. H. Burke; J. E. Hearst, *Mol. Gen. Genet.* **1994**, *245*, 406–416.
45. M. Nakagawa; N. Misawa, *Agric. Biol. Chem.* **1991**, *55*, 2147–2148.
46. B. S. Hundle; D. A. O'Brien; M. Alberti; P. Beyer; J. E. Hearst, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9321–9325.
47. A. Yokoyama; G. Sandmann; T. Hoshino; K. Adachi; M. Sakai; Y. Shizuri, *Tetrahedron Lett.* **1995**, *36*, 4901–4904.
48. F. Blasco; I. Kauffmann; R. D. Schmid, *Appl. Micrbiol. Biotechnol.* **2004**, *64*, 671–674.
49. A. Yokoyama; H. Izumida; W. Miki, *Biosci. Biotechnol. Biochem.* **1994**, *58*, 1842–1844.
50. A. Yokoyama; W. Miki; H. Izumida; Y. Shizuri, *Biosci. Biotechnol. Biochem.* **1996**, *60*, 200–203.
51. N. Misawa; Y. Satomi; K. Kondo; A. Yokoyama; S. Kajiwara; T. Saito; T. Ohtani; W. Miki, *J. Bacteriol.* **1995**, *177*, 6575–6584.
52. N. Misawa; S. Kajiwara; K. Kondo; A. Yokoyama; Y. Satomi; T. Saito; W. Miki; T. Ohtani, *Biochem. Biophys. Res. Commun.* **1995**, *209*, 867–876.
53. J. H. Lee; Y. T. Kim, *Gene* **2006**, *370*, 86–95.
54. L. Tao; P. E. Rouviere; Q. Cheng, *Gene* **2006**, *379*, 101–108.
55. P. D. Fraser; Y. Miura; N. Misawa, *J. Biol. Chem.* **1997**, *272*, 6128–6135.
56. P. D. Fraser; H. Shimada; N. Misawa, *Eur. J. Biochem.* **1998**, *252*, 229–236.
57. S. K. Choi; Y. Nishida; S. Matsuda; K. Adachi; H. Kasai; X. Peng; S. Komemushi; W. Miki; N. Misawa, *Mar. Biotechnol.* **2005**, *7*, 515–522.
58. T. Makino; H. Harada; H. Ikenaga; S. Matsuda; S. Takaichi; K. Shindo; G. Sandmann; T. Ogata; N. Misawa, *Plant Cell Physiol.* **2008**, *49*, 1867–1878.
59. S. K. Choi; S. Matsuda; T. Hoshino; X. Peng; N. Misawa, *Appl. Microbiol. Biotechnol.* **2006**, *72*, 1238–1246.
60. A. Yokoyama; K. Adachi; Y. Shizuri, *J. Nat. Prod.* **1995**, *58*, 1929–1933.
61. T. Maruyama; H. Kasai; S. K. Choi; A. K. Ramasamy; Y. Inomata; N. Misawa, *Carotenoid Sci.* **2007**, *11*, 50–55.
62. L. Pasamontes; D. Hug; M. Tessier; H. P. Hohmann; J. Schierle; A. P. van Loon, *Gene* **1997**, *185*, 35–41.
63. R. A. Larsen; M. M. Wilson; A. M. Guss; W. W. Metcalf, *Arch. Microbiol.* **2002**, *178*, 193–201.
64. A. Berry; D. Janssens; M. Humbelin; J. P. M. Jore; B. Hoste; I. Cleenwerck; M. Vancanneyt; W. Bretzel; A. F. Mayer; R. Lopez-Ulibarri; B. Shanmugam; J. Swings; L. Pasmontes, *Int. J. Syst. Evol. Microbiol.* **2003**, *53*, 231–238.
65. K. Kaneda; T. Kuzuyama; M. Takagi; Y. Hayakawa; H. Seto, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 932–937.

66. KuzuyamaT; Seto H, *Nat. Prod. Rep.* **2003**, *20*, 171–183.
67. S. Fujisaki; H. Hara; Y. Nishimura; K. Horiuchi; T. Nishino, *J. Biochem.* **1990**, *108*, 995–1000.
68. B. Fernandez-Gonzalez; G. Sandmann; A. Vioque, *J. Biol. Chem.* **1997**, *272*, 9728–9733.
69. K. Masamoto; N. Misawa; T. Kaneko; R. Kikuno; H. Toh, *Plant Cell Physiol.* **1998**, *39*, 560–564.
70. K. Masamoto; H. Wada; T. Kaneko; S. Takaichi, *Plant Cell Physiol.* **2001**, *42*, 1398–1402.
71. S. Takaichi; M. Mochimaru, *Cell. Mol. Life Sci.* **2007**, *64*, 2607–2619.
72. S. Takaichi; T. Maoka; K. Masamoto, *Plant Cell Physiol.* **2001**, *42*, 756–762.
73. S. Takaichi; M. Mochimaru; T. Maoka; H. Katoh, *Plant Cell Physiol.* **2005**, *46*, 497–504.
74. M. Mochimaru; H. Masukawa; S. Takaichi, *FEBS Lett.* **2005**, *579*, 6111–6114.
75. S. Steiger; G. Sandmann, *Biotechnol. Lett.* **2004**, *26*, 813–817.
76. M. Mochimaru; H. Masukawa; T. Maoka; H. E. Mohamed; W. F. J. Vermaas; S. Takaichi, *J. Bacteriol.* **2008**, *190*, 6726–6733.
77. S. Takaichi; M. Mochimaru; T. Maoka, *Plant Cell Physiol.* **2006**, *47*, 211–216.
78. M. Teramoto; S. Takaichi; Y. Inomata; H. Ikenaga; N. Misawa, *FEBS Lett.* **2003**, *545*, 120–126.
79. K. Shindo; K. Kikuta; A. Suzuki; A. Katsuta; H. Kasai; M. Yasumoto-Hirose; Y. Matsuo; N. Misawa; S. Takaichi, *Appl. Microbiol. Biotechnol.* **2007**, *74*, 1350–1357.
80. S. Kajiwara; P. D. Fraser; K. Kondo; N. Misawa, *Biochem. J.* **1997**, *324*, 421–426.
81. H. Harada; F. Yu; S. Okamoto; T. Kuzuyama; R. Utsumi; N. Misawa, *Appl. Microbiol. Biotechnol.* **2009**, *81*, 915–925.
82. M. Albrecht; N. Misawa; G. Sandmann, *Biotechnol. Lett.* **1999**, *21*, 791–795.
83. L. Z. Yuan; P. E. Rouviere; W. Suh, *Metab. Eng.* **2006**, *8*, 79–90.
84. K. Kakinuma; Y. Dekishima; Y. Matsushima; T. Eguchi; N. Misawa; M. Takagi; T. Kuzuyama; H. Seto, *J. Am. Chem. Soc.* **2001**, *123*, 1238–1239.
85. D. M. Martin; J. Gershenzon; J. Bohlmann, *Plant Physiol.* **2003**, *132*, 1586–1599.
86. R. V. Vadali; Y. Fu; G. N. Bennett; K. Y. San, *Biotechnol. Prog.* **2005**, *21*, 1558–1561.
87. J. D. Newman; J. Marshal; M. Chang; F. Nowroozi; E. Paradise; D. Pitera; K. L. Newman; J. D. Keasling, *Biotechnol. Bioeng.* **2006**, *95*, 684–691.
88. S. H. Yoon; Y. M. Lee; J. E. Kim; S. H. Lee; J. H. Lee; J. Y. Kim; K. H. Jung; Y. C. Shin; J. D. Keasling; S. W. Kim, *Biotechnol. Bioeng.* **2006**, *94*, 1025–1032.
89. S. H. Yoon; H. M. Park; J. E. Kim; S. H. Lee; M. S. Choi; J. Y. Kim; D. K. Oh; J. D. Keasling; S. W. Kim, *Biotechnol. Prog.* **2007**, *23*, 599–605.
90. M. Albrecht; S. Takaichi; N. Misawa; G. Schnurr; P. Boger; G. Sandmann, *J. Biotechonol.* **1997**, *58*, 177–185.
91. H. Krugel; P. Krubasik; K. Weber; H. P. Saluz; G. Sandmann, *Biochim. Biophys. Acta* **1999**, *1439*, 57–64.
92. S. Takaichi; G. Sandmann; G. Schnurr; Y. Satomi; A. Suzuki; N. Misawa, *Eur. J. Biochem.* **1996**, *241*, 291–296.
93. T. Matsuno; S. Nagata; K. Kitamura, *Tetrahedron Lett.* **1976**, *50*, 4601–4604.
94. C. Schmidt-Dannert; D. Umeno; F. H. Arnold, *Nat. Biotechnol.* **2000**, *18*, 750–753.
95. P. C. Lee; A. Z. R. Momen; B. N. Mijts; C. Schmidt-Dannert, *Chem. Biol.* **2003**, *10*, 453–462.
96. D. Umeno; F. H. Arnold, *Appl. Environ. Microbiol.* **2003**, *69*, 3573–3579.
97. Z. Sun; E. Gantt; F. X. Cunningham, Jr., *J. Biol. Chem.* **1996**, *271*, 24349–24352.
98. K. Inoue, *Trends Plant Sci.* **2004**, *9*, 515–517.
99. R. F. Quinlan; T. T. Jaradat; E. Wurtzel, *Arch. Biochem. Biophys.* **2007**, *458*, 146–157.
100. L. Tian; V. Musetti; J. Kim; M. Magallanes-Lundback; D. DellaPenna, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 402–407.
101. G. Giuliano; R. Tavazza; G. Diretto; P. Beyer; M. A. Taylor, *Trends Biotechnol.* **2008**, *26*, 139–145.
102. N. Misawa, *Plant Biotechnol.* **2009**, *26*, 93–99.
103. N. Misawa; S. Yamano; H. Linden; M. R. de Felipe; M. Lucas; H. Ikenaga; G. Sandmann, *Plant J.* **1993**, *4*, 833–840.
104. N. Misawa; K. Masamoto; T. Hori; T. Ohtani; P. Boger; G. Sandmann, *Plant J.* **1994**, *6*, 481–489.
105. C. K. Shewmaker; J. A. Sheehy; M. Daley; S. Colburn; D. Y. Ke, *Plant J.* **1999**, *20*, 401–412.
106. S. Römer; P. D. Fraser; J. W. Kiano; C. A. Shipton; N. Misawa; W. Schuch; P. M. Bramley, *Nat. Biotechnol.* **2000**, *18*, 666–669.
107. P. D. Fraser; S. Romer; C. A. Shipton; P. B. Mills; J. W. Kiano; N. Misawa; R. G. Drake; W. Schuch; P. M. Bramley, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1092–1097.
108. L. Ralley; E. M. Enfissi; N. Misawa; W. Schuch; P. M. Bramley; P. D. Fraser, *Plant J.* **2004**, *39*, 477–486.
109. S. Suzuki; M. Nishihara; T. Nakatsuka; N. Misawa; I. Ogiwara; S. Yamamura, *Plant Cell Rep.* **2007**, *26*, 951–959.
110. M. Fujisawa; M. Watanabe; S. K. Choi; M. Teramoto; K. Ohyama; N. Misawa, *J. Biosci. Bioeng.* **2008**, *105*, 636–641.
111. M. Fujisawa; E. Takita; H. Harada; N. Sakurai; H. Suzuki; K. Ohyama; D. Shibata; N. Misawa, *J. Exp. Bot.* **2009**, *60*, 1319–1332.
112. P. H. Schreier; E. A. Seftor; J. Schell; H. J. Bohnert, *EMBO J.* **1985**, *4*, 25–32.
113. W. L. Morris; L. J. Ducreux; P. D. Fraser; S. Millam; M. A. Taylor, *Metab. Eng.* **2006**, *8*, 253–263.
114. V. Mann; M. Harker; I. Pecker; J. Hirschberg, *Nat. Biotechnol.* **2000**, *18*, 888–892.
115. K. Stålberg; O. Lindgren; B. Ek; A.-S. Höglund, *Plant J.* **2003**, *36*, 771–779.
116. T. Gerjets; G. Sandmann, *J. Exp. Bot.* **2006**, *57*, 3639–3645.
117. C. Zhu; T. Gerjets; G. Sandmann, *Transgenic Res.* **2007**, *16*, 813–821.
118. T. Hasunuma; S. Miyazawa; S. Yoshimura; Y. Shinzaki; K. Tomizawa; K. Shindo; S. K. Choi; N. Misawa; C. Miyake, *Plant J.* **2008**, *55*, 857–868.
119. F. X. Cunningham, Jr.; E. Gantt, *Plant J.* **2005**, *41*, 478–492.
120. T. Lotan; J. Hirschberg, *FEBS Lett.* **1995**, *364*, 125–128.
121. S. Kajiwara; T. Kakizono; T. Saito; K. Kondo; T. Ohtani; N. Nishio; S. Nagai; N. Misawa, *Plant Mol. Biol.* **1995**, *29*, 343–352.
122. J. Jayaraj; R. Devlin; Z. Punja, *Transgenic Res.* **2008**, *17*, 489–501.
123. K. Shindo; T. Hasunuma; E. Asagi; A. Sano; E. Hotta; N. Minemura; C. Miyake; T. Maoka; N. Misawa, *Tetrahedron Lett.* **2008**, *49*, 3294–3296.

Biographical Sketch

Norihiko Misawa, Ph.D., is a senior research scientist at the Central Laboratories for Frontier Technology, Kirin Holdings Co. Ltd. and a visiting professor at Ishikawa Prefectural University.

1.21 Sterol and Steroid Biosynthesis and Metabolism in Plants and Microorganisms

Hubert Schaller, Université de Strasbourg and CNRS, Strasbourg, France

© 2010 Elsevier Ltd. All rights reserved.

1.21.1	**Introduction**	755
1.21.2	**Sterol and Steroid Biosynthetic Pathways**	756
1.21.2.1	Squalene Metabolism	756
1.21.2.1.1	Squalene or 2,3-oxidosqualene cyclizations	756
1.21.2.1.2	2,3-Oxidosqualene cyclization in cycloartenol or lanosterol in eukaryotes	757
1.21.2.1.3	Metabolization of the cyclopropane ring of 9β,19-cyclopropyl sterols	758
1.21.2.2	Structural Diversity of Sterols: The C17 Side Chain	759
1.21.2.2.1	Sterol-C24-methyltransferases	759
1.21.2.2.2	Sterol Δ^{24}-isomerase/reductase	762
1.21.2.2.3	Sterol-22-desaturase	762
1.21.2.3	Biosynthesis of Δ^5-Sterols: Metabolism of the Tetracyclic Skeleton	763
1.21.2.3.1	The C4-demethylation complex	763
1.21.2.3.2	CYP51	764
1.21.2.3.3	A common trunk of genes implicated in isomerization, desaturation, and reductions on the B and D rings	766
1.21.2.4	Polyoxidized Derivatives	768
1.21.2.4.1	Steroidal hormones in fungi	768
1.21.2.4.2	Phytoecdysteroids	769
1.21.2.4.3	Brassinosteroids	769
1.21.2.5	Steroids with Heterocycles in the Side Chain	771
1.21.2.5.1	Steroidal saponins	771
1.21.2.5.2	Steroidal glycoalkaloids	772
1.21.2.5.3	Cardiotonic steroidal glucosides	773
1.21.3	**Sterol and Steroid Conjugates**	774
1.21.3.1	Steryl Esters	774
1.21.3.2	Sterol Glucosides	775
1.21.3.3	Acylated Sterol Glucosides	775
1.21.3.4	Steroid Sulfates	775
1.21.4	**Sterol Degradation**	776
1.21.5	**Transport**	776
1.21.6	**Molecular Regulation of Sterol Biosynthesis**	777
1.21.7	**Functions of Steroids**	778
References		780

1.21.1 Introduction

The diversity of sterol and steroid chemical structures across kingdoms and the broad outline of corresponding biosynthetic schemes have been depicted.[1] The structural elucidation and biosynthesis of cholesterol was recently summarized from a historical prospect.[2] Sterols are isoprenoid lipids essential to cell membrane structure and function, and to a further metabolism into steroidal hormones in eukaryotes. Recently, phytosterol derivatives used as food additives were shown to be efficient cholesterol-lowering agents.[3] Plant sterol biosynthesis and its peculiarities were described,[4–6] as well as its general terpenoid metabolic context.[7] This chapter reviews the current state of knowledge on sterol and steroid metabolism in bacteria, fungi, and plants. Particular focus is on

functional gene discovery for a comprehensive understanding of biosynthetic pathways of steroids and the function of the latter in physiology and development. The nomenclature of sterol, sterol conjugates, steroids, steroidal saponins and glycoalkaloids, and cardenolides is given elsewhere.[8] Analysis of sterols has also been comprehensively reported elsewhere.[9,10] This chapter describes sterol and steroid metabolism across diverse organisms for which progress in the field has been achieved in recent years owing to molecular genetic approaches. Aspects of animal steroid (hormones) metabolism and action are not included here.

1.21.2 Sterol and Steroid Biosynthetic Pathways

1.21.2.1 Squalene Metabolism

1.21.2.1.1 Squalene or 2,3-oxidosqualene cyclizations

Squalene (**1**) is the C_{30} precursor of steroids (**Figure 1**). It is produced in prokaryotes and eukaryotes by squalene synthases. These enzymes are prenyl transferases that catalyze catalyzing the head-to-head condensation of two farnesyl diphosphates to yield presqualene diphosphate as an intermediate of squalene.[11] Prokaryotes generally cyclize squalene into hopane triterpenes, although it is known that a few species, such as *M. capsulatus*, can produce steroids in addition to hopane triterpenes and bacteriohopanols.[12] *M. capsulatus* contains squalene hopane cyclase and squalene oxide lanosterol cyclase activities.[13] Genomic analysis of such methanotrophic bacterium indicates the presence of functional genes encoding squalene epoxidase (SQE), the product of which is (3*S*)-2,3-oxidosqualene (**2**) and (3*S*)-2,3-oxidosqualene cyclase (OSC), the product of which is lanosterol (**3**).[14] Genome mining identified the planctomycete *Gemmata obscuriglobus* as another bacterial species containing the biosynthetic sequence squalene to lanosterol, which requires an SQE and an OSC.[15] Although *M. capsulatus* or the tubercle bacillus *Mycobacterium tuberculosis* metabolizes lanosterol into 4α-methyl-$\Delta^{8(14)}$-sterols or cholesterol, *G. obscuriglobus* synthesizes lanosterol (**3**) and parkeol (**4**) as end products with no downstream modifications. Phylogenetic and biochemical data suggest that sterol biosynthetic pathways might have been exchanged through gene transfer between bacteria and early eukaryotes. *Stigmatella aurantiaca*, another species from the myxobacteria, produce cycloartenol (**5**) as a sterol precursor. The similarity of *S. aurantiaca* and eukaryotic cycloartenol synthase gene products may also indicate an evolutionary relationship between these organisms.[16] Coexistence of squalene and squalene oxide cyclization products is not restricted to prokaryotes. *Adiantum capillus-veneris* and *Dryopteris crassirhizoma* are two fern species from the polypodiale order from which squalene cyclases (SQCs) have been cloned and shown to possess 35–40% identity with prokaryotic SQC. Functional analysis in the heterologous host *Saccharomyces cerevisiae* demonstrated that the *D. crassirhizoma* encoded a dammaradiene synthase, converting squalene (**1**) into dammara-18(28),21-diene (**6**).[17] Genes encoding SQEs from higher plants have been isolated in *Arabidopsis thaliana*,[18] *Artemisia annua*,[19] *Panax*

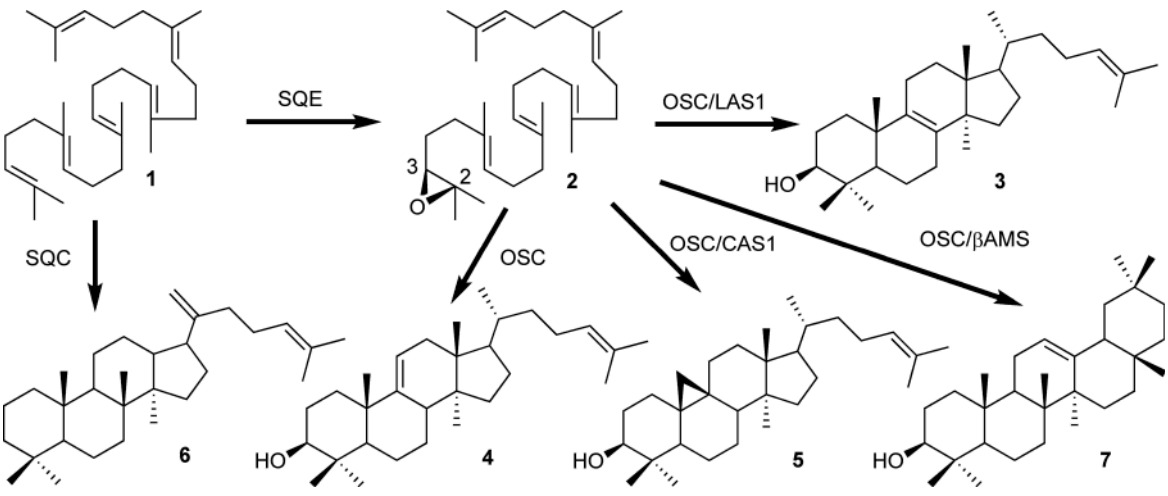


Figure 1 Squalene or 2,3-oxidosqualene cyclizations.

notoginseng,[20] and *Euphorbia tirucalli*.[21] *In situ* hybridization experiments with antisense probes of *E. tirucalli* pointed out a strong expression of *SQE* in parenchyma cells adjacent to laticifers. The Arabidopsis genome contains six genes encoding putative SQEs, among which three were shown to functionally complement a yeast mutant deficient in the endogenous SQE. Among the possibly redundant *SQE* genes of *A. thaliana*, *SQE1* was shown to be essential for triterpene and sterol biosynthesis because T-DNA (transfer DNA) insertional mutants of *SQE1* accumulated squalene. Consequently, root and seed development were impaired in these plants.[18] SQEs contain conserved flavin adenine dinucleotide (FAD)-binding domains. SQE enzymatic activity studied in yeast required molecular oxygen, nicotinamide adenine dinucleotide phosphate, reduced form (NADPH), and FAD.[22] The substrate-binding site of a mammalian SQE was identified by photoaffinity labeling.[23]

1.21.2.1.2 2,3-Oxidosqualene cyclization in cycloartenol or lanosterol in eukaryotes

Enzymatic transformation of (3*S*)-2,3-oxidosqualene (**2**) into cycloartenol (**5**), lanosterol (**3**), parkeol (**4**), or polycyclic triterpenoids such as β-amyrin (**7**) (**Figure 1**) involves carbocationic cyclizations and 1,2 rearrangements reviewed in Abe *et al*.[24] The diversity of the chemical structures produced from (3*S*)-2,3-oxidosqualene (**2**) is well illustrated when considering the 13 genes from *A. thaliana* encoding OSCs and the corresponding products formed upon expression of these genes in yeast.[25] The biodiversity in nonsteroidal triterpenes cyclization has been recently summarized.[26] Cyclization of 2,3-oxidosqualene into lanosterol (**3**) and cycloartenol (**5**) or into β-amyrin (**7**) were differentiated biochemically by the use of the inhibitor *N*-alkyl-4α,10-dimethyl-8-aza-*trans*-3β-ol, designed to mimic the high-energy carbocationic intermediates of the reaction.[27] A benzophenone photophore-containing nonterpenoid inhibitor, used as a photoaffinity label, mapped common binding sites of the mammalian lanosterol synthase and squalene hopene cyclase from *Alicyclobacillus acidocaldarius*.[28] *A. thaliana* cycloartenol synthase *CAS1* (2,3-oxidosqualene cycloartenol synthase) was isolated by metabolic interference in a yeast mutant lacking lanosterol synthase (strain *erg7*) and was transformed from being an auxotroph to being an ergosterol. A cDNA library cloned into a yeast expression plasmid vector was used to transform *erg7*. A chromatographic screen was then implemented to detect the expected formation of cycloartenol synthase in *erg7*.[29] Equivalent functional complementation strategies were developed to isolate the *ERG7* gene encoding the fungal lanosterol synthase.[30] In addition to the growing number of plant cycloartenol synthases and to the few bacterial ones, cycloartenol synthases have been characterized in the amoebas *Acantamoeba polyphaga* and *Dictyostelium discoidum*, and in the flagellate euglenoid *Astasia longa*.[31] Parasite protists from the kinetoplastideae, such as *Trypanosoma brucei*, use lanosterol synthase to produce their sterols.[32] An *A. thaliana* CAS1 mutant enzyme, bearing a valine residue at position 481 instead of an isoleucine residue, was able to produce lanosterol (**3**) and parkeol (**4**) in addition to cycloartenol (**5**).[33] Other amino acid residues important for the enzymatic formation of cycloartenol (**5**) or lanosterol (**3**) by the respective cyclases were disclosed in molecular evolution experiments.[34,35] The chemistry–biology interdisciplinary study, including site-directed mutagenesis and structural approaches of the yeast and mammalian lanosterol synthase, has been comprehensively described.[36] Position 481 of *A. thaliana CAS1* resides in the protosteryl cation-binding domain. Alignment of other OSCs with *CAS1* indicates that plant enzymes exhibit an isoleucine at position 481, whereas fungal and mammalian enzymes have a valine. *S. aurantiaca*, which uses a cycloartenol synthase, has an isoleucine at position 481, whereas *M. capsulatus* and *G. obscuriglobus*, which use lanosterol synthase, have a valine. This was crucial for the assignment of the function of an *A. thaliana* gene (At3g45130), whose deduced polypeptide sequence had 62% identity with CAS1. This protein, after being expressed in yeast, was shown to catalyze the synthesis of lanosterol (**3**),[37,38] as were the *Lotus japonicus* orthologs.[39] Although the presence of lanosterol (**3**) in certain plant species (e.g., from the genus *Euphorbia*) has been known for a long time, the fact that lanosterol synthases (LAS1) are apparently widely distributed in dicotyledonous plants raises the question of their physiological significance. In *A. thaliana*, the essential function of CAS1 has been demonstrated by a genetic approach. Complete loss-of-function of *CAS1* was lethal. Plants that were characterized by a weak allele, and therefore had a reduced transcription of the gene, accumulated 2,3-oxidosqualene (**2**). They were characterized by albino inflorescence shoots.[40] A conditional *CRE/loxP* (cyclization recombination locus of X-over P) recombination-dependent mutant allele also showed an albino phenotype at the seedling stage shortly after the *CRE/loxP*-induced onset of *CAS1* loss of function. In addition, these seedlings, which also accumulated 2,3-oxidosqualene, finally arrested their growth. It was concluded that there was no redundancy in *CAS1* for the synthesis of sterol precursors.[40] A dual biosynthetic pathway to phytosterols through cycloartenol (**5**) and lanosterol (**3**) was

investigated in *A. thaliana* seedlings overexpressing *LAS1* or deficient in *LAS1* expression using labeled mevalonate (6-$^{13}C^{2}H_3$), fed in the presence of an inhibitor of the enzyme 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR), which catalyzes the production of mevalonate. The synthesis of small amounts of lanosterol-derived sitosterol (**10**) was detected by nuclear magnetic resonance (NMR) of deuterium at carbon 19.[41] However, the fact that *CAS1*-deficient mutants are lethal indicates that *LAS1* cannot exert a compensation effect in plants on sterol biosynthesis. Lanosterol synthase might appear to be an evolutionary remnant, or it might have been recruited for a specific steroid pathway, for example, in secondary metabolism (as was proposed but not demonstrated).[41]

1.21.2.1.3 Metabolization of the cyclopropane ring of 9β,19-cyclopropyl sterols

Plants and some protists use an apparently complicated pathway to produce tetracyclic sterols via the pentacyclic triterpene cycloartenol (**5**), instead of the tetracyclic lanosterol (**3**). This implies the existence of a cyclopropane-opening enzyme or cyclopropyl isomerase (CPI), **Figure 2**). The enzymatic activity has been originally described in microsomal fractions of plant cell suspensions[42] and was later characterized thoroughly in subcellular fractions of maize. This enzyme catalyzes the isomerization of cycloeucalenol (**8**) into obtusifoliol (**9**), according to a carbocationic mechanism.[43] Obtusifoliol (**9**) (the substrate of CYP51, see Section 1.21.2.3.2)

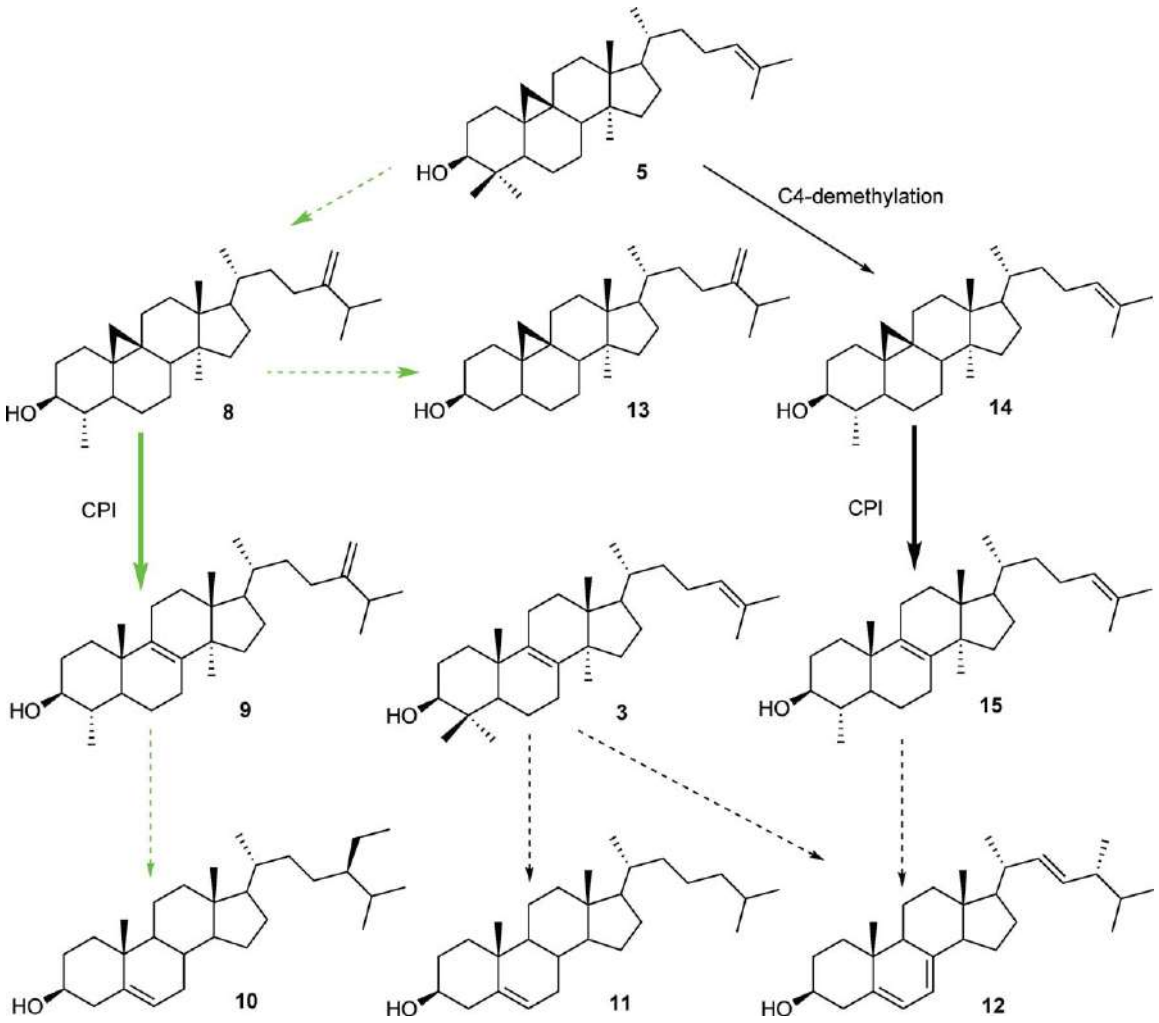


Figure 2 Metabolization of the 9β,19-cyclopropanic ring during the course of sterol biosynthesis in plants (green arrows) and yeast (black arrows).

then undergoes C14 demethylation. One can observe that downstream of obtusifoliol (**9**), the plant sterol biosynthetic pathway (dashed green arrow from obtusifoliol (**9**) to sitosterol (**10**) in **Figure 2**) is very similar to the vertebrate and fungi pathway (dashed black arrows to cholesterol (**11**) and to ergosterol (**12**), respectively, in **Figure 2**). Cycloeucalenol (**8**) or its C4-demethylated derivatives pollinastanol or 24-alkyl-pollinastanol (e.g., 24-methylene pollinastanol (**13**)) do not usually accumulate in cells except in the pollen.[44] However, cycloeucalenol (**8**) and other 9β,19-cyclopropyl sterols (**13**) may accumulate in plants[45] or protists[46] after the cyclopropyl sterol isomerase (CPI, formerly designated COI for cycloeucalenol–obtusifoliolisomerase) has been inhibited by *N*-alkyl-morpholine fungicides (tridemorph, fenpropimorph) and other C9 carbocationic transition state analogs. Chemical inhibition in growing seedlings of wheat[47] or in tobacco callus cultures isolated in a somatic genetic approach of sterol biosynthesis[48] resulted in an almost complete replacement of pathway end products by cycloeucalenol derivatives. In these experiments, cells or organisms have been viable with unusual sterols in their membranes. The significance of cycloartenol (**8**) and 9β,19-cyclopropylsterols as mandatory intermediates in sterol biosynthesis of some organisms has therefore been discussed in terms of surrogates of Δ^5-sterols. Molecular cloning of the *A. thaliana* CPI was achieved by metabolic interference in *S. cerevisiae*.[49] An ergosterol auxotroph *erg*7 strain, expressing the *A. thaliana* CAS1, and therefore accumulating 9β,19-cyclopropylsterols, was transformed with a cDNA library cloned into a yeast expression vector. A yeast transformant capable of ergosterol prototrophy was isolated. Prototrophy was most probably conferred by a protein able to open the cyclopropane of 31-nor cycloartenol (**14**) to yield 31-nor lanosterol (**15**).[8,50] This metabolite was a precursor for ergosterol biosynthesis in that yeast transformant.

Genetic inhibition of *CPI* in plants has been documented. An *A. thaliana* mutant carrying a transposable DNA element inserted into the *CPI* gene was characterized by a sterol profile fully consistent with those of plants treated with CPI inhibitors. This resulted in the accumulation of cycloeucaleneol and its derivatives, especially 24-alkyl pollinastanol.[51] Interestingly, this mutant had a severly hampered growth and development. The correct membrane sterol composition was shown to be essential for polar localization of auxin transporters.[51] This study pointed out a possible mechanism for sterol action on establishing asymmetric protein localization. From the overall *in vivo* studies, it is interesting to note apparent conflicting interpretations on the physiological significance of 9β,19-cyclopropane sterol intermediacy in plants. Genetic inhibition indeed is not compatible with the accomplishment of a complete life span,[51] whereas chemical somatic inhibition does not prevent cellular growth.[48]

1.21.2.2 Structural Diversity of Sterols: The C17 Side Chain

1.21.2.2.1 Sterol-C24-methyltransferases

Besides the lanosterol–cycloartenol bifurcation, alkylation of the sterol side chain is the other prominent peculiarity of sterol biosynthesis, which confers diversity in this pathway. Sterols from fungi and plants possess an alkyl group at C24 that sterol from vertebrate do not possess (**Figure 3**). A tremendous array of sterol structures with different side chains and the occurrence of such compounds have been described extensively.[1] Alkylation at C24 is performed by *S*-adenosyl-methionine-sterol-C-methyltransferases (SMTs), which generate 24-alkyl-sterols and *S*-adenosyl-homocysteine. These reactions have been initially studied in yeast and in the chlorophyte *Trebouxia* spp.[52] Methylation of zymosterol (**16**) into fecosterol (**17**) and methylation of cycloartenol (**5**) into 24-methylene cycloartanol (**18**) in these species were inhibited by azasterols.[53] In plant microsomal fractions, synthetic 25-azacycloartanol was described as a carbocationic transition state analog that inhibited cycloartenol C24-methyltransferase (SMT1) and 24-methylene lophenol C24^1-methyltransferase (SMT2).[54] *S. cerevisiae* contains a single SMT encoded by *ERG6*. This gene was described as nonessential because yeast strains containing *erg6* knockout alleles had a normal vegetative growth.[55] The essentiality or dispensability of ERG6 genes was challenged in a growth competition experiment. Isogenic strains differing by mutations in stereoidogenic genes were grown together with the wild type in order to test the possible competitive advantage of sterol biosynthetic mutants over ergosterol-producing strains. The conclusion was that the earlier the mutation in the biosynthetic scheme, the less able the strain was to compete with the wild type; therefore, no ergosterol biosynthetic gene/enzyme could be considered as nonessential.[56] A detailed analysis of the intracellular distribution of sterol biosynthetic enzymes was conducted with yeast. This analysis showed that SMT/ERG6 was localized almost exclusively

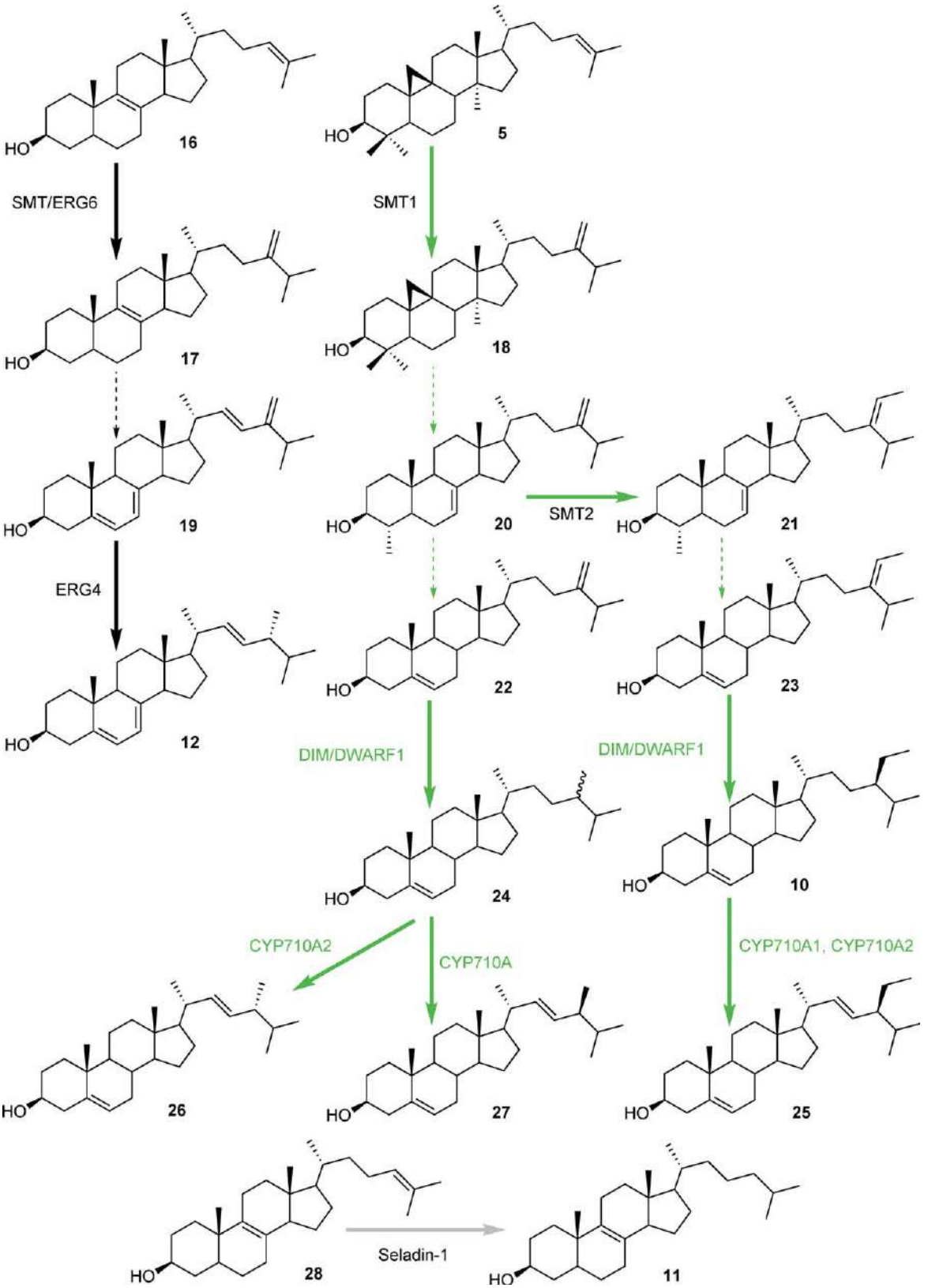


Figure 3 Sterol side-chain metabolism in yeast, plants, and vertebrates (black, green, and gray arrows).

in lipid particles and therefore was possibly implicated in sterol translocation.[57] Other sterol biosynthetic enzymes in yeast were shown to have a dual localization in the endoplasmic reticulum and in lipid particles.[58] Extensive chemical and enzymological studies have defined the mechanism and structural requirements for sterol side-chain alkylation by SMTs.[59] Sterol biosynthesis, and particularly the SMTs, have been considered as chemotherapeutic targets in trypanosomes and other kinetoplastid parasites.[60] In plants, *SMTs* have been studied mostly in *A. thaliana*, *Glycine max*, *Nicotiana tabacum*, and *Zea mays*. The capability of plants to produce alkylated sterols is due to the existence of two distinct and biosynthetically nonconsecutive sterol-C24-methyltransferases. Biosynthetic studies, including incorporation of radiolabeled precursors or treatment of cells with inhibitors, proposed that a first methylation reaction should be applied to cycloartenol to yield 24-methylene cycloartanol, then a second methylation reaction should be applied to 24-methylene lophenol to yield 24-ethylidene lophenol. A soybean gene *SMT1* was isolated by screening a cDNA library with antiplasma membrane serum. The corresponding protein expressed in *Escherichia coli* possessed lanosterol-C24-methyltransferase activity.[61] A homology-based approach led to the isolation of an *A. thaliana* SMT showing a 38% identity with the ERG6. Transformation of the wild-type and *erg6* mutants with this gene led to the synthesis of 24-ethyl sterols, indicating that this plant SMT could perform two sequential methylations in the sterol side chain.[62] Identification of two types of SMTs in plant genomes indicated that there were different biochemical functions or a (partial) redundancy of functions. The yeast mutant *erg6* provided the demonstration of the distinct reactions catalyzed by distinct SMT1 and SMT2. Transformed *erg6* strains were used to prepare delipidated microsomes to determine the substrate specificities of the SMT1 and SMT2 enzymes encoded by respective genes. Catalytic efficiencies measured in a study with *N. tabacum* enzymes indicated that SMT1 converted cycloartenol (**5**) into 24-methylene cycloartanol (**18**) but did not convert 24-methylene lophenol (**20**) into 24-ethylidene lophenol (**21**).[63] The case was slightly different with SMT2, which had a catalytic efficiency of 17 times higher with 24-methylene lophenol (**20**) (converted into 24-ethylidene lophenol (**21**)) than with cycloartenol (**5**).[63] This study concluded that the identity of SMT1 or SMT2 proteins with each other was close to 80%, whereas the identity of SMT1 and SMT2 was close to 40%. Conserved domains of the SMTs include a sterol-binding site and an *S*-adenosylmethionine-binding pocket.[64,65] In plants, SMT2 defines a branching point between the 24-methyl sterol and the 24-ethyl sterol biosynthetic segments. Transgenic approaches reported in *N. tabacum* indicated specific biochemical effects of *SMT1* or *SMT2* following up- or downregulated expression;[66] in particular, *SMT1* was shown to control the flux of carbon into sterol biosynthesis in tobacco seeds.[67] *SMT2* had a considerable impact on controlling the ratio of 24-methyl cholesterol to sitosterol.[68] A cholesterol-rich profile characterized an *A. thaliana* mutant impaired in one of its three genes (At5g13710) encoding SMTs. This *smt1* mutant was identified in a visual screen of a transposon-mutagenized population for root sensitivity to calcium.[69] A Series of allelic *smt1* mutants were characterized by slow overall growth, severely hampered embryogenesis, and altered root gravitropism.[70] In addition to an elevated cholesterol level, these *smt1* mutants had an almost-unchanged 24-methyl cholesterol but strongly reduced sitosterol levels. Such a chemical phenotype of mutants confirmed an overlap in the substrate specificity of SMT enzymes, *in vivo* particularly the fact that SMT2 can methylate cycloartenol, although with a low efficiency.[63] *A. thaliana*, overexpressing *SMT2* constitutively, accumulates more sitosterol than the wild type at the expense of campesterol and were smaller in size, most probably due to a reduced pool of campesterol at the entry of the brassinosteroid pathway (see Section 1.21.2.4.3).[71] Other transgenic lines showed cosuppression of *SMT2* and, therefore, contained four- to fivefold more 24-methyl cholesterol than the wild type at the expense of sitosterol. Pleiotropic effects on development such as reduced growth, increased branching, modified flower morphology, and low fertility were associated with modified sterol composition.[71] The *cvp1* (cotyledon vein patterning) insertional mutant of *SMT2*, displaying similar biochemical traits, was identified in a screen for novel cotyledon vascular patterns.[72] The allele *fril1* of *SMT2* had serrate petals and sepals due to ectopic endoreduplication in petal tips, suggesting a possible link between sterol composition and suppression of endoreduplication.[73] Cholesterol is naturally present in high proportions in species from the Solanaceae, which contain steroidal glycoalkaloids (see Section 1.21.2.5.2). Cholesterol has been suggested to be a precursor of solanidin in potato or tomatidin in tomato. Transgenic *Solanum tuberosum* plants expressing a *SMT1* from *G. max* had a decreased level of cholesterol associated with decreased levels of glycoalkaloids.[74]

1.21.2.2.2 Sterol Δ^{24}-isomerase/reductase

This biosynthetic step exists in all eukaryotes (**Figure 3**). Under normal conditions, *S. cerevisiae* converts ergosta-5,7,22,24(24^1)-tetraen-3β-ol (**19**) into ergosterol. The reductase encoded by the gene *ERG4* is located in the endoplasmic reticulum.[75] This enzyme in yeast was inhibited by azasteroids.[76] Besides biosynthetic aspects, ERG4 appeared to play a role in cell polarity, apical bud growth, cell wall assembly, mating, and invasive growth due to its interaction with p21-activated kinase Ste20.[77] Identification of the plant enzyme originated from a genetic approach using *A. thaliana*. This is no doubt a good example of serendipity in plant metabolic biology. The *diminuto* mutant was isolated among T-DNA transformed lines for its poor growth due to a defect in regulating cell elongation at the level of tubulin gene expression.[78] Molecular characterization of the mutated locus showed that *DIM* encoded a sterol Δ^{24}-isomerase/reductase due to the accumulation of 24-methylene cholesterol (**22**) and isofucosterol (**23**) in plants.[79] Expression of GFP fusion protein indicated an endoplasmic reticulum localized sterol Δ^{24}-isomerase/reductase. The allele *dwarf1* of a series of dwarf mutants was, in fact, the first *A. thaliana* mutant generated by T-DNA insertional mutagenesis, displaying a morphological phenotype (dwarfism) inherited as a single recessive nuclear mutation, which cosegregated with the associated marker gene (kanamycin antibiotic resistance) and the T-DNA insert.[80] In this *dwarf1* mutant, the lack of campesterol resulted in a reduced amount of bioactive brassinosteroids (see Section 1.21.2.4.3), causing dwarfism and altered development.[81] The plant sterol Δ^{24}-isomerase/reductase contains a flavin adenine dinucleotide (FAD)-binding domain indicative of a flavoenzyme. It also contains a Ca^{2+}/calmodulin-binding domain, which is essential for its function.[82] The sequence of reactions leading from $\Delta^{24(241)}$ to a C24(24^1) saturated bond through $\Delta^{24(25)}$ catalyzed by such a multifunctional enzyme was reported earlier.[83] The human desmosterol (**28**) Δ^{24}-isomerase/reductase was isolated in a mRNA differential display experiment designed to compare different brain regions in the context of neurodegeneration associated with Alzheimer's disease and was named seladin-1 for *sel*ective *A*lzheimer *d*isease *in*dicator-1 (Seladin).[84,85] This enzyme is an ortholog of the plant DIM/DWARF1, which is not the case of the yeast ERG4. Seladin-1 was also implicated in adrenocortical tumorigenesis[86] and is highly expressed in melanoma cell lines derived from cutaneous metastases.[87]

1.21.2.2.3 Sterol-22-desaturase

Cytochrome P-450 oxygenases are responsible for sterol-C22(23) desaturation in fungi[88] and plants.[8] The microsomal enzyme in yeast was purified, and its activity was reconstituted in an assay, including an animal NADPH-P-450 reductase.[89] The gene *ERG5* encoding this enzyme was cloned by functional complementation of a yeast mutant using negative selection for nystatin-sensitive transformants, which indicated the presence of ergosterol.[90] A dual biochemical function of *ERG5*/*CYP61* was considered because it showed xenobiotic metabolism, particularly with benzopyrene in yeast genotoxicity assays.[91] The *ERG5* orthologs of yeast-like symbiots, which synthesize sterols used by rice planthopper hosts (otherwise called sterol auxotrophs), had a different exon–intron organization compared to that of the *S. cerevisiae* gene.[92] The ciliated protozoan *Tetrahymena thermophila* was able to transform exogenous cholesterol into $\Delta^{7,22}$ derivatives in a cell-free assay, which required molecular oxygen, cytochrome b_5, and reduced cofactors (NADH or NADPH), pointing out an implication of cytochrome b_5 reductase in the (non-P-450 dependent) desaturation reaction.[93] *A. thaliana* has four genes encoding sterol-22-desaturases. These genes form the family *CYP710A* of cytochrome P-450 oxygenases that were isolated based on the sequence analysis and partial identity with the fungal sterol-22-desaturase encoded by *ERG5*/*CYP61*. *A. thaliana* contains multiple sterol products with a C22(23) saturation (see **Figure 3**). CYP710A1 and CYP710A2, expressed as recombinant proteins in insect cells, converted sitosterol (**10**) into stigmasterol (**25**) and converted the epimer of 24-methyl cholesterol (24) 24-*epi*-campesterol into brassicasterol (**26**).[94] Recruitment of a specific CYP710A isoform for the conversion of campesterol (**24B**) into crinosterol (**27**) was not detailed. Transgenic plants overexpressing CYP710A1 had over 30-fold increased levels of stigmasterol[94], and transgenic plants overexpressing CYP710A4 also increased their levels of stigmasterol.[95] This was associated with the esterification of Δ^5-sterols.[95] The apparent dispensability (because of highly variable levels) of stigmasterol (**25**) in plants is not well understood. The moss *Physcomitrella patens* also has *CYP710A* ortholog. Disruption of that gene by homologous recombination did not affect the viability of protonema, chloronema, or caulonema,[96] as was the case for *A. thaliana* knocked out in one of the *CYP710A*, which developed identically to the wild type.[94]

1.21.2.3 Biosynthesis of Δ^5-Sterols: Metabolism of the Tetracyclic Skeleton

1.21.2.3.1 The C4-demethylation complex

Cycloartenol (**5**) and lanosterol (**3**) are triterpene precursors that undergo a succession of enzymatic transformations, ultimately leading to functional sterols (e.g., cholesterol (**11**), ergosterol (**12**), and 24-alkyl cholesterol (**10**, **24**)). Triterpene–sterol conversion implies oxidative removal of two methyl groups at C4 and another methyl group at C14. The succession of these steps varies, depending on the organism being considered. In animal or fungal sterol biosynthetic pathways, the two methyl groups at C4 are removed sequentially by the same enzymatic complex (**Figure 4**).[97,9] This complex has been identified genetically via transcriptome analysis in *S. cerevisiae* and isolated in classical genetic approaches and in two-hybrid analysis experiments.

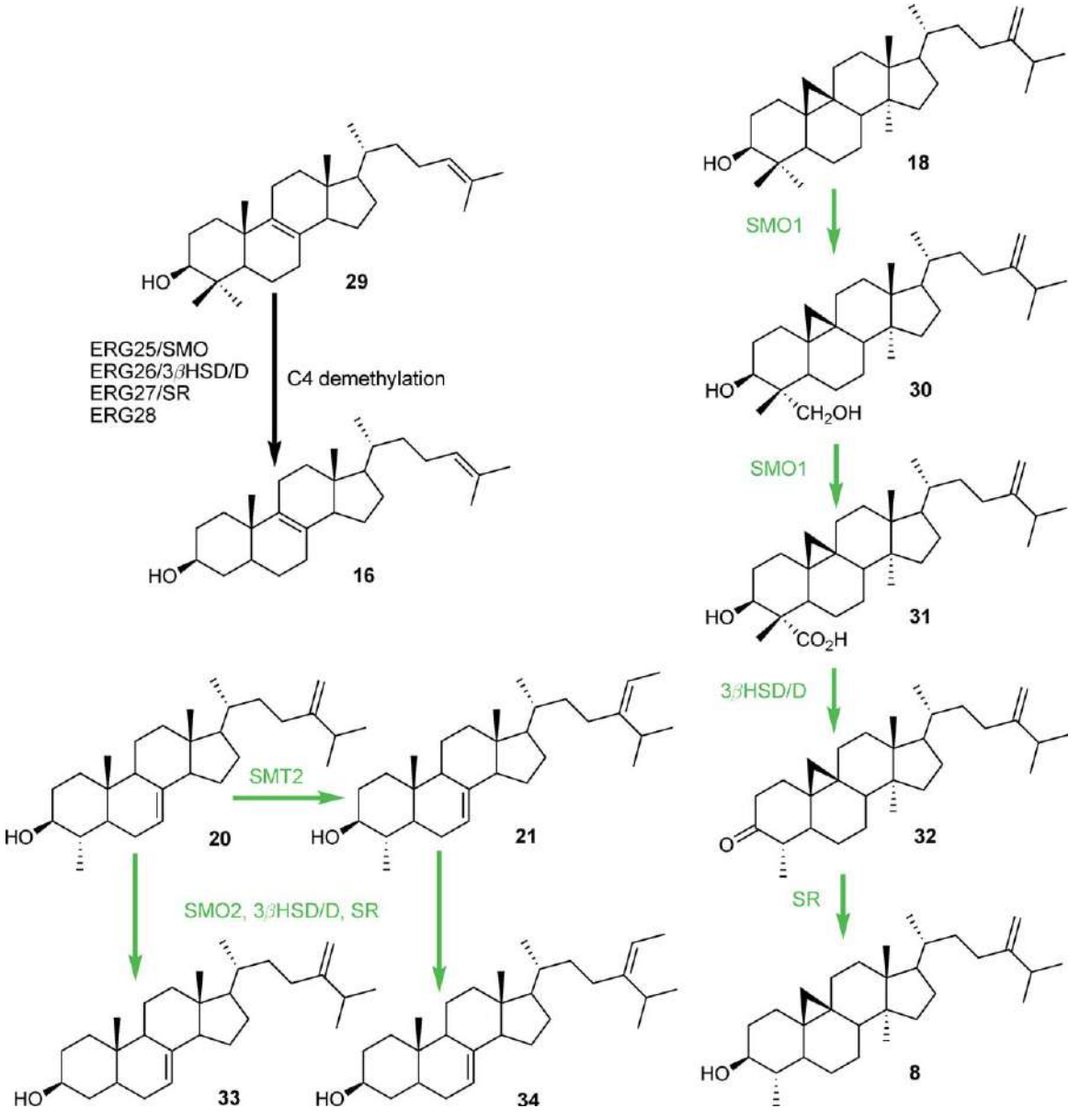


Figure 4 Sterol-C4-demethylation.

The enzymatic complex consists of ERG25, a sterol-4α-methyl-oxidase (SMO), isolated by functional complementation of the ergosterol auxotroph *erg25*, which accumulates 4,4-dimethyl-zymosterol (**29**).[98,99] The *erg25* was also complemented with the human ortholog of ERG25.[100] The ERG25-catalyzed reaction yields a 4α-hydroxymethyl sterol (**30**), then a 4α-carboxy-sterol intermediate (**31**),[101] is further used as a substrate by ERG26, a second component of the complex, which is a bifunctional 4α-carboxysterol-3β-hydroxysteroid dehydrogenase/C4-decarboxylase (3βHSD/D), belonging to the family of short-chain dehydrogenase/reductase (SDR)[102] whose product is a 3-oxosteroid (**32**). A third component of the C4-demethylation complex, ERG27, is a sterone reductase (SR) whose product is a 4-desmethyl sterol.[103] In addition to these enzymes, the C4-demethylation complex includes ERG28, a transmembrane protein[104,105] essential for the activity of the complex. A model multienzymatic membrane-bound sterol-C4-demethylation complex included cytochrome b_5 reductase and cytochrome b_5 in addition to ERG25, ERG26, ERG27, and ERG28.[9] Enzymological studies of ERG25 confirmed 4,4-dimethyl-zymosterol (**29**) as the preferred substrate of the microsomal enzyme[101] and were in full accordance with previous biochemical characterization of the reaction (in the animal and yeast system).[106,107] ERG25/SMOs are phylogenetically related to other membrane-bound nonheme iron hydroxylases widely distributed as like sterol-C5(6)-desaturases, cholesterol-25-hydroxylases, sphingolipid hydroxylases, and fatty acid desaturases.[9]

In plants, there are two distinct C4-demethylation reactions (**Figure 4**) implicated in two nonconsecutive reactions in the pathway, as is C24 alkylation of the C17 sterol side chain by SMTs (see Section 1.21.2.2.1). The first C4-demethylation applies to 24-methylene cycloartanol (**18**) which is transformed into cycloeucalenol (**8**). The second C4-demethylation applies to 24-methylene lophenol (**20**) and 24-ethylidene lophenol (**21**), which are transformed into episterol (**33**) and Δ7-avenasterol (**34**), respectively.[108] This peculiarity of the plant sterol pathway, compared to the yeast pathway, was demonstrated in *Z. mays* by the enzymology of microsomal sterol C4-methyl oxidases, 4α-carboxysterol-3β-hydroxysteroid/C4-decarboxylase (3βHSD/D), and NADPH-dependent-3-oxosteroid reductase (SR).[108–110] Virus-induced gene silencing (VIGS) in *Nicotiana bethamiana* proved that plants have *SMO1* and *SMO2* organized in small gene families, most probably implicating the redundancy of their biochemical functions.[111] Functional identification of plant genes coding for 3βHSD/D was done by via ERG26 homology-based searches for orthologs and expression in yeast.[112] Two *A. thaliana* cDNAs encoding 3βHDS/D restored ergosterol prototrophy in *erg26*. Finally, VIGS reduced the expression of *3β-HSD/D* in *N. benthamiana* and triggered the accumulation of 3β-hydroxy-4β-14-dimethyl-5α-ergosta-9β,19-cyclo-24(24^1)-en-4α-carboxylic acid (**31**).[112] The murine gene encoding NSDHL (sterol dehydrogenase) can also complement *erg26*.[113]

1.21.2.3.2 CYP51

In animals and fungi sterol biosynthesis, oxidative removal of the 14α-methyl group by the cytochrome P-450 oxidase CYP51 proceeds immediately after cyclization of 2,3-oxidosqualene into lanosterol (vertebrate and yeasts demethylate lanosterol (**3**), filamentous fungi demethylate 24-methylene lanosterol also called eburicol (**38**), **Figure 5**).[114] In plants, oxidative removal occurs at a later stage (obtusifoliol (**9**)), between the first and the second C4-demethylation reactions (**Figure 5**). CYP51 is the only known P-450 distributed in all organisms with conservation of function. CYP51s, including plant or animal pathogens, are target sites for azole inhibitors.[115] Site-directed mutagenesis, combined with genetic screens for azole-resistant mutants, indicated that azole and substrate binding had different structural requirements.[116] The pathogen *M. tuberculosis* has a soluble CYP51 ortholog that demethylates lanosterol and obtusifoliol.[117] This enzyme was expressed in *E. coli* then and crystallized in the presence of the antifungal agent fluconazole. Its structure at 2.2 Å (the first reported for a P-450 oxidase), and a mapping analysis of *Candida albicans* azole-resistant mutants, showed that drug resistance in pathogenic fungi mapped to protein regions required for catalysis rather than an azole-binding domain.[118] The proteobacterium *M. capsulatus* revealed in its genome a novel type of CYP51 bearing a ferrodoxin domain at the C-terminus and producing lanosterol-14-demethylase activity when expressed and purified from *E. coli*.[119] Trypanosomes (kinetoplastids) have CYP51s that display a strong substrate preference to obtusifoliol (**9**).[120] Plant CYP51 was purified and then cloned from *Sorghum bicolor*.[121,122] Such orthologs could complement *erg11*(*cyp51*) defective yeast mutants. Microsomal fractions from these yeasts permitted us to measure the binding constants of azole herbicides, ranging to a micromolar order of magnitude.[123] A somatic genetic approach in *N. tabacum*

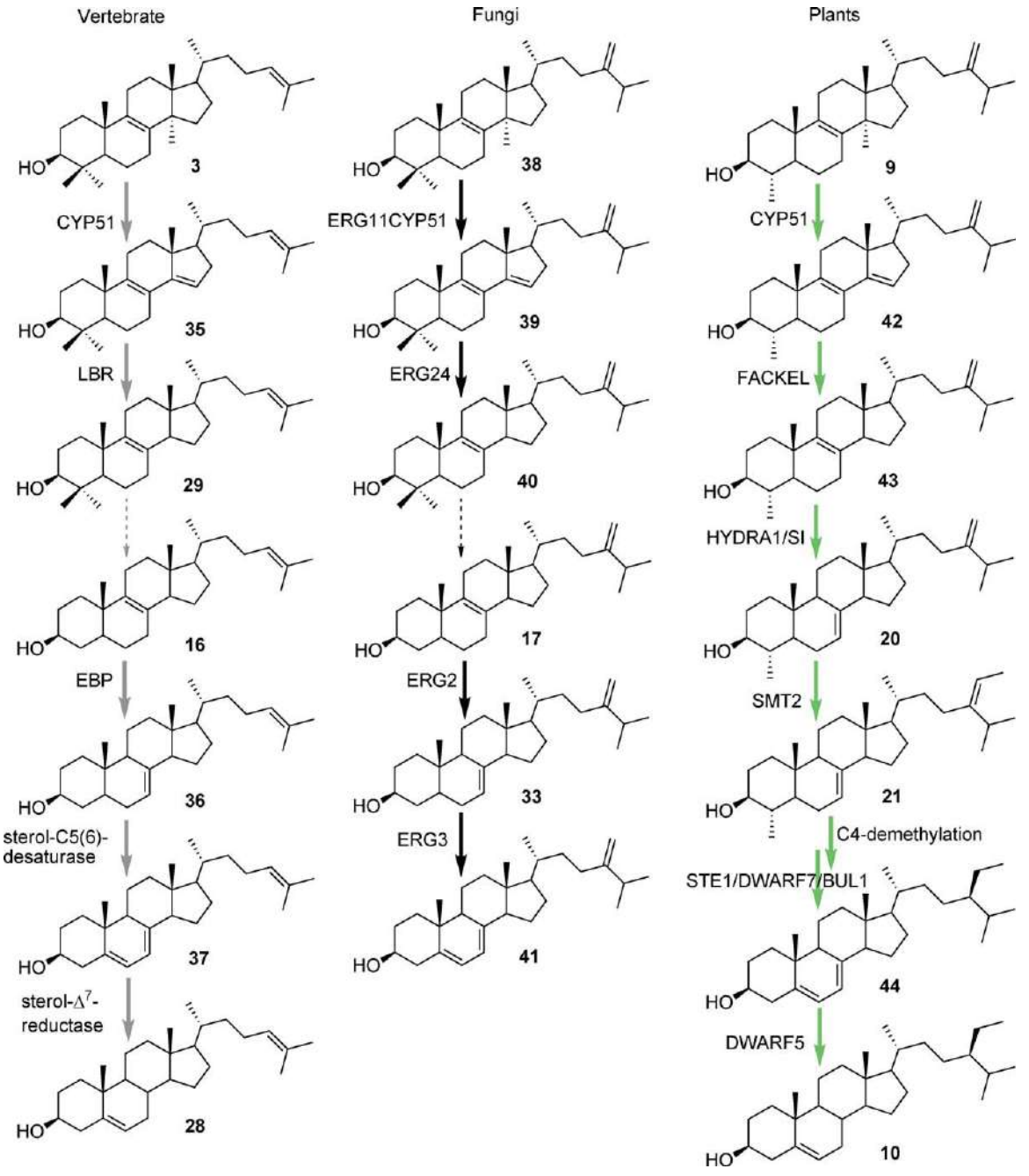


Figure 5 Comparative sterol biosynthesis in vertebrate, fungi, and plants.

led to the screening of biochemical sterol mutants presenting a marked azole-resistance phenotype.[124] Lethality characterized *A. thaliana* mutants carrying loss-of-function T-DNA alleles of CYP51, as were yeast strains deficient in ERG11/CYP51.[125] *CYP51* from *N. tabacum* mimicked the effect of inhibitors of obtusifoliol-14-demethylase in a VIGS approach, that is, accumulation of obtusifoliol (**9**), 24-dihydroobtusifoliol, and corresponding C4-demethylated and/or Δ^{24}-reduced metabolites (14α-methyl-fecosterol and 14α-methyl-24(24^1)-dihydrofecosterol).[126]

1.21.2.3.3 A common trunk of genes implicated in isomerization, desaturation, and reductions on the B and D rings

Comparative analysis of vertebrate, fungal, and plant sterol biosynthesis (**Figure 5**) led to the observation that the bioconversion of sterol intermediates possessing a cholesta-8,14-dien-3β-ol, an ergosta-8,14-dien-3β-ol, or a stigmasta-8,14-dien-3β-ol into Δ^5-sterols required enzymes that are functionally interchangeable across kingdoms in most, if not all, cases. The yeast gene *ERG24* encoding a $\Delta^{8,14}$-sterol-Δ^{14}-reductase was isolated by homologous functional complementation.[127] The *erg24* produced ignosterol (ergosta-8,14-dien-3β-ol instead of ergosterol), the C-4 demethylated substrate (**39**) of ERG24, and this did not affect viability.[128] However, in the case of *erg24* mutants of *C. albicans*, such a biochemical phenotype was shown to reduce the pathogenicity of an inoculum of fungal cells intravenously injected into mice.[129] This reinforced the potency of inhibitors of sterol-Δ^{14}-reductase as antifungals.[130,131] A 15-azasteroid (15-aza-24-methylene-D-homocholesta-8,14-dien-3β-ol), produced by the soil fungus *Geotrichum flavobrunneum*, was a strongly specific inhibitor of ERG24.[132] Enzymology and inhibition studies of the sterol-Δ^{14}-reductase, done with microsomal fractions of *Z. mays* coleoptiles, showed that this 15-azasteroid iminium analog behaved as a carbocationic transition state mimic, and this was also true for the *N*-alkyl morpholine fungicides.[133] Homology-based searches have led to the identification of a plethora of *ERG24* orthologs and of a human membrane lamin B receptor (LBR).[134,135] LBRs possess a sterol reductase domain and was therefore functional in a sterol-Δ^{14}-reductase complementation assay of *erg24*. The LBR from *Drosophila melanogaster*, a sterol auxotroph organism, did not encode a sterol-Δ^{14}-reductase function because it could not restore ergosterol prototrophy in *erg24*,[136]; therefore, it was most probably an evolutionary variant. The plant gene encoding a $\Delta^{8,14}$-sterol-Δ^{14}-reductase was isolated by positional cloning of a mutated allele called *fackel* in a genetic screening for *A. thaliana* that affected embryo development.[137] This ortholog had a 33% identity with ERG24 and also the typical sterol reductase and LBR motifs. It was able to rescue growth of *erg24* in the presence of high calcium concentrations, otherwise deleterious to these latter yeast cells.[138] *A. thaliana* embryo-defective *fackel* homozygotes cultivated on a synthetic medium contained elevated levels of $\Delta^{8,14}$-sterols, a chemotype in agreement with the mapped mutation.

Δ^8-Sterol intermediates are isomerized into Δ^7 isomers by a Δ^8-sterol-Δ^8-Δ^7-sterol isomerase (SI). In plants, Δ^8-sterols replace Δ^5-sterols in cells grown on the inhibitor AY9944.[7] The preferred substrate of the plant enzyme is 4α-methyl-5α-ergosta-8,24(24^1)-dien-3β-ol (**43**) (**Figure 5**), which accumulates upon treatment with AY9944, and is further metabolized into 4-desmethyl-Δ^8-sterol including 24-ethyl-Δ^8-sterols as alternative end products of the pathway. This is a good example of the plasticity of the sterol pathway, due to the relatively low specificity of enzymes localized downstream of the target site of an enzyme inhibitor (i.e., compound **43** may be demethylated at C4 by SMO2 and other enzymes of the C4-demethylation complex, methylated at C24 most probably by SMT2, and reduced at $\Delta^{24(241)}$ by DWARF1/DIM). This has been clearly explained in grids of alternative biosynthetic routes that parallel the main one.[7,8] The *S. cerevisiae ERG2* gene, encoding the Δ^8-sterol-Δ^8-Δ^7-sterol isomerase, was cloned by homologous functional complementation.[139] *ERG2* orthologs of plant pathogenic fungi such as *Magnaporthe grisea* (causing rice blast disease) and *Ustilago maydis* (causing corn smut disease) were also isolated.[140] ERG2 is indeed an important target for antifungal commercial compounds from the *N*-alkyl-morpholine group.[141] During the characterization of immunosuppressant compounds in the model *S. cerevisiae*, a genetic screening of UV-induced mutants that were resistant to an immunosuppressant molecule SR31747 (known to block the proliferation of lymphocytes) identified an allelic series of mutations in two *erg2* lethality suppressor genes, which conferred resistance to many structurally different sterol biosynthesis inhibitors. In the same study, overexpression in yeast of ERG2 conferred resistance to SR31747, indicating that Δ^8-sterol-Δ^8-Δ^7-sterol isomerase was the primary target of SR31747. This study discussed the relationship between immunosuppressants, sigma receptors, and Δ^8-sterol-Δ^8-Δ^7-sterol isomerase.[142] Mammalian and plant enzymes were also isolated by functional complementation of the yeast *erg2*.[143,144] In mice and humans, the ERG2 counterparts are bifunctional proteins. One function is Δ^8-sterol-Δ^8-Δ^7-sterol isomerization and the other is a pharmacological isomerization, that is, emopamil-binding proteins (EBP). Mouse and human orthologs have only 12% identity with the yeast ERG2 (at the peptidic sequence level), whereas plant and fungal isomerases share about 35% identity.[145] Sigma1 receptors in animals have a 30% identity with the yeast sterol-Δ^8-isomerase but do not possess the corresponding enzyme activity.[146,147] Yeasts expressing the plant isomerase in the *erg2* background had an ergosterol biosynthesis blocked by sigma ligands, for example, haloperidol and verapamil.[144] *Arabidopsis thaliana hydra* mutants are a

class of severe dwarfs that were identified in a genetic screen for defective embryogenesis and cell patterning in seedlings. Molecular analysis of the mutations showed that *HYDRA1* encoded the plant sterol-Δ^8-isomerase.[148] A deficiency of the sterol biosynthetic pathway at this nonredundant step led to a severe depletion of Δ^5-sterols and an accumulation of (yet nonelucidated) ergostenol and stigmastenol derivatives.[149] The physiological implications of *hydra1* (and *hydra2/fackel/sterol;-$\Delta^{8,14}$-reductase*) genetic defects in embryonic and postembryonic development were discussed in terms of sterol signaling. Most importantly, the modified sterol composition of *hydra* mutants affected auxin and ethylene signaling.[148] This showed that an appropriate sterol composition is essential for the activity of membrane-bound proteins and for membrane biology in general, as was previously indicated in the case of ATPases localized in the plasma membrane of plants.[150]

The Δ^7-sterol intermediates lathosterol (Δ^7-cholestenol) or $\Delta^{7,24}$-cholestadienol (**36**), episterol (**33**), and Δ^7-avenasterol (**34**) are the substrates of Δ^7-sterol-C5(6)-desaturases in animals, fungi, and plants (**Figure 5**). The gene *ERG3* was isolated by functional complementation of an *erg3* mutant.[151] The situation at the Δ^7 sterol intermediacy in plants was different from that occurring with Δ^8-sterol intermediates with respect to cell viability and development in autotrophic conditions. Phytochemical analyses have revealed that species from many families, including Cucurbitaceae and Chenopodiaceae, contain Δ^7-sterols as major sterols.[152,153] Nes and McKeen[1] listed plant species producing Δ^7-sterols as pathway end products. In a genetic approach of sterol biosynthesis, a chromatographic screen was applied to pooled individuals from populations of EMS (ethyl methane sulfonate)-mutagenized *A. thaliana* in order to isolate biochemical mutants. This led to the isolation of the *ste1* mutant, which accumulated about 70% of (24ξ)-24-methyl-5α-cholest-7-en-3-ol (Δ^7-campesterol) and (24*R*)-24-ethyl-5α-cholest-7-en-3-ol (Δ^7-sitosterol), at the expense of the Δ^5-sterols campesterol and sitosterol.[154] The mutation had almost no effect on morphogenesis and growth. An allelic series of *STE1*, *dwarf7* and *bul1*, was characterized by an extremely dwarf phenotype of plants. Such dwarfism and biosynthetic defects could be complemented biochemically by feeding exogenous brassinosteroids to growing seedlings[155,156] and genetically by the expression of *ERG3*.[154] In the tiny dwarf *bul1*, a cellular analysis showed that microtubule polymerization/depolymerization was compromised,[156] just like in the *diminuto/dwarf1 (sterol-Δ^{24}-isomerase/reductase)* mutant. These plants expressing strong alleles of *STE1* contained Δ^7-sterols and as little as 2% of residual Δ^5-sterols. This indicated that the dwarfism was due to the lack of a sufficient pool of campesterol that served as a precursor for the synthesis of brassinosteroids, and that in Arabidopsis at least, Δ^7-sterols cannot prime the synthesis of brassinosteroids (see Section 1.21.2.4.3). The *A. thaliana* sterol-C5(6)-desaturase *STE1* was cloned by functional complementation of the yeast *erg3* mutant. A plant cDNA expression library was transformed in *erg3* and transformants were screened for cycloheximide resistance, nystatin sensitivity, and sterol content.[157] One cDNA expressed in the *ste1* mutant restored a wild-type sterol composition and was therefore a Δ^7-sterol-C5(6)-desaturase. A second gene encoding a putative Δ^7-sterol-C5(6)-desaturase located beside *STE1* was found in the genome of *A. thaliana*: the encoded protein shared 80% identity with STE1. The functional analysis of this gene was not reported. Molecular analysis of *ste1* showed that the encoded polypeptide contained a single amino acid substitution T114I.[158] Δ^7-sterol-C5(6)-desaturase enzymatic activities were characterized in rat liver microsomes[159] and in *Z. mays* coleoptile microsomes.[160] The human gene *SC5DL* complemented the mutant erg3.[161] The plant Δ^7-sterol-C5(6)-desaturase STE1, studied by functional expression in *erg3*, is a nonheme iron oxygenase, requiring cytochrome b_5 as an electron carrier from the reductant NADPH to the Δ^7-sterol-C5(6)-desaturase, via cytochrome b_5 reductase. Site-directed mutagenesis identified the histidine-rich motifs of the protein as ligands for a catalytic Fe center, as was the case for membrane-bound fatty acid desaturases.[162] The plant enzyme carrying the T114I mutation expressed in *erg3* had a higher K_m and a lower catalytic efficiency, in agreement with the biochemical phenotype of *ste1*. Interestingly, the conservative T114S mutation had a 28-fold higher V_{max} value and an increased catalytic efficiency compared to the wild type, indicating that this amino acid residue played an essential role in the catalytic process.[162] The molecular mechanism of sterol C5(6) desaturation was performed either with C5α- or C6α-deuterated Δ^7-cholestenol analogs as mechanistic probes. These substrates showed deuterium kinetic isotope effects in accordance with the chemical activation of the C6α–H bond prior to its cleavage by the enzyme as a rate-limiting step in the desaturation reaction.[163]

The last enzymes of the sterol pathway that modify the tetracyclic skeleton are $\Delta^{5,7}$-sterol Δ^7-reductases. These reductases in plants and animals but not in yeasts (**Figure 5**). In plants, the reduction of $\Delta^{5,7}$-cholestadienol into cholesterol was enzymatically characterized in microsomal preparations from

Z. mays coleoptiles.[164] This reduction reaction is NADPH dependent and is strongly inhibited by carbocationic transition state substrate analogs. In agreement with such a carbocationic mechanism of the reduction of the Δ^7 double bond, synthetic azasteroids, including 6-aza-B-homo-5α-cholest-7-en-3β-ol, were particularly efficient in inhibiting the *Z. mays* microsomal $\Delta^{5,7}$-sterol Δ^7-reductase *in vitro*. *In vivo*, 6-aza-B-homo-5α-cholest-7-en-3β-ol–treated *Rubus fruticosus* cells contained (24*R*)-24-ethyl-5α-cholest-5,7-dien-3β-ol (**44**).[165] In mammals, inhibition by a piperazine derivative of the $\Delta^{5,7}$-sterol Δ^7-reductase caused an accumulation of $\Delta^{5,7}$-cholestadienol.[166] A plant cDNA encoding the $\Delta^{5,7}$-sterol Δ^7-reductase was isolated from an *A. thaliana* library expressed in wild-type yeast. With this strategy of metabolic interference of the reductase with the ergosterol biosynthetic pathway, a transformant displaying nystatin resistance was isolated due to the absence of ergosterol that was further metabolized into Δ^5-sterols. The cloned protein presented sequence similarities with other sterol reductases.[167] The efficient expression of the *A. thaliana* $\Delta^{5,7}$-sterol Δ^7-reductase (DWARF5) in yeast supported a further biotechnological strategy designed to produce mammalian steroids.[168] Functional analysis of the plant $\Delta^{5,7}$-sterol Δ^7-reductase was done in *A. thaliana*. An allelic series of *dwarf5* mutants was characterized by knockout mutations or other null mutations (deletion, splice-site, missense, and nonsense mutations). This resulted in an accumulation of $\Delta^{5,7}$-sterols at the expense of campesterol and sitosterol.[169,170] Consequently, strong mutations in *DWARF5* (as in the Δ^7-*sterol-C5(6)-desaturase/STE1/DWARF7* or in the Δ^{24}-*sterol isomerase/reductase DIMINUTO/DWARF1* of *A. thaliana*) result in a brassinosteroid deficiency and, therefore, are a typical dwarfism. Biochemical and genetic complementation of the *dwarf5* mutants with exogenous brassinosteroids and with an expressed cDNA encoding DWARF5, respectively, were in full accordance with the crucial role of Δ^5-sterols as precursors of plant steroid hormones and as structural components of membranes.[8,169,170]

1.21.2.4 Polyoxidized Derivatives

1.21.2.4.1 Steroidal hormones in fungi

Antheridiol (**46**) and oogoniol (**47**) are derivatives of fucosterol (**45**) (**Figure 6**) acting as pheromones in the oomycetes of *Achlya ambisexualis* (water mold). These compounds control sexual morphogenesis.[171] Their action on cellular metabolism includes ribosomal RNA or protein synthesis.[172,173] Biochemical analysis of a high-affinity steroid-binding protein suggested that it could function as a steroid receptor.[174] An antheridiol-induced chaperone HSP90 heat-shock protein was shown to be associated with the steroid receptor complex.[175,176] Molecular and genetic aspects of the biosynthesis of antheridiol and oogoniol have not been investigated in detail.

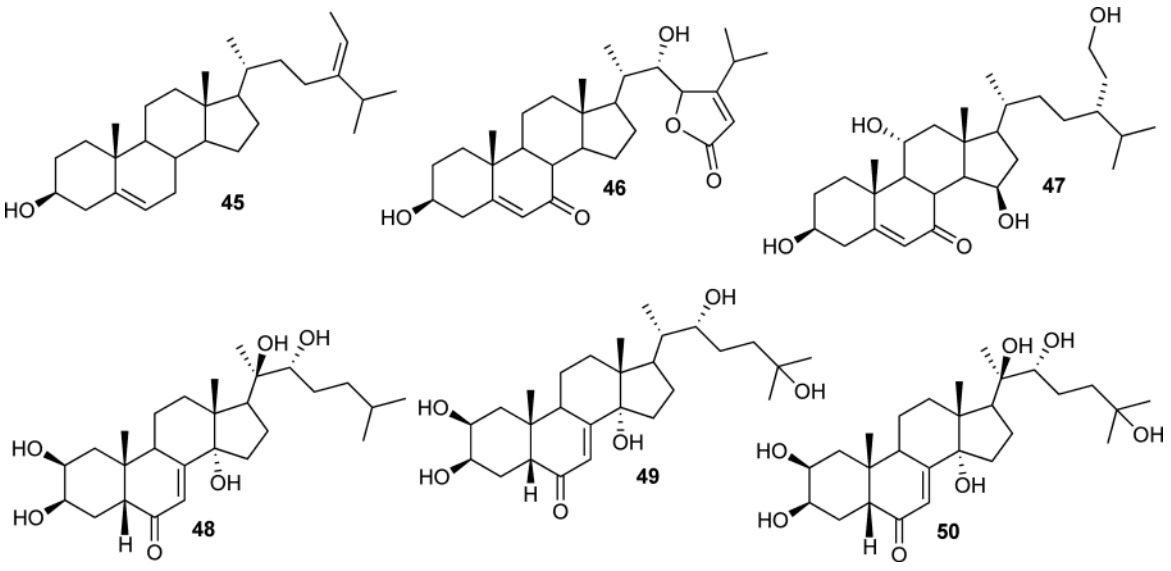


Figure 6 Structures of steroid hormones in Oomycetes and of some ecdysteroids.

1.21.2.4.2 Phytoecdysteroids

Phytoecdysteroids are distributed in a large number of land plants[177] including polypodine B (**48**) in ferns[178] and α-ecdysone (**49**) or 20-hydroxyecdysone (**50**) (**Figure 6**) in Chenopodiaceae.[179,180] Mycoecdysteroids have been described in fungi.[181] These compounds have the same structural features as ecdysteroids found in insects or other arthropods, for instance, α-ecdysone (**49**) or 20-hydroxyecdysone (**50**), which are hormones acting in important developmental cellular processes such as molting.[182] The biology and molecular regulation of this process in insect development has been reviewed.[183] It is assumed that plant ecdysteroids could exert deterrent or antifeedant effects on predators[184] or develop interfering effects in molting.[185] Although the phytochemical diversity of ecdysteroids has largely been covered, the biosynthesis of this class of compound is still largely unknown. Hundreds of C27, C28, or C29 5β-steroids collectively named 'phytoecdysteroids' have been reported.[186] Conjugated forms of ecdysteroids are known in plants and animals.[187] Ecdysteroid biosynthesis and transport have been studied in spinach leaves. Radiolabeled [^{14}C]-mevalonate was efficiently incorporated into 20-hydroxecdysone[188], therefore cholesterol might be a precursor. A possible biosynthetic intermediate between cholesterol (or lathosterol in Caryophyllaceae that do not produce Δ^5-sterols) is a 14α-hydroxy-7-en-6-one derivative.[186] Structural characteristics of ecdysteroids, in addition to the chromophore in the B ring and hydroxyl groups at 3β and 14α, typically show a high degree of oxidation (additional double bonds, hydroxyl or oxo groups). Biotransformations of putative phytoecdysteroid precursors were studied in tissue cultures of *Polypodium vulgare*. Incubation of calli or prothalli with various labeled ecdysteroids led to the detection of C2-hydroxylase enzyme activity when compounds with a hydroxyl group at C3 were tested as substrates. The same material was used to study the stereospecificity of the enzymatic conversion of 22-hydroxycholesterol or 25-hydroxycholesterol into ecdysone and 20-hydroxyecdysone.[189] In ecological approaches, plant families have been surveyed for the presence of ecdysteroid agonist or antagonist activities using *D. melanogaster* cellular bioassays.[190] Biological activities of ecdysteroids and brassinosteroids have been compared using this insect cell bioassay as well as the rice lamina inclination assay, classically used in brassinosteroid biology. There was no interference of brassinosteroid with ecdysteroid signaling in insects observed, nor was any ecdysteroid with brassinosteroid signaling in plants observed.[191]

1.21.2.4.3 Brassinosteroids

Brassinolide was discovered in rape pollen in 1979[192] and castasterone in chestnut insect gall in 1982.[193] An array of plant steroids named brassinosteroids were subsequently described, and their natural occurrence and biosynthesis was documented extensively.[194] Chemical analysis of brassinosteroids, present in plants at the nanomolar range, first included bioassays and radioimmunoassays, as well as methods in GC–MS (gas chromatography–mass spectrometry). Many aspects of the physiological effects of brassinosteroids were studied, and their possible applications in agriculture were described.[194] The elucidation of the biosynthetic pathway and the characterization of molecular regulation of brassinosteroid signaling was achieved within a decade, mostly through the use of genetic approaches with the model plant *A. thaliana*. Oxidative conversion of sterols into brassinosteroids was shown to consist of a grid of multiple pathways. Arabidopsis mutants showing impaired skotomorphogenesis were characterized as small dwarfs whose growth can be restored to the wild-type level by exogenous brassinosteroids.[195,196] The gene *DET2s*, isolated through this mutational approach, was shown to encode a protein of 40% identity with the mammalian steroid-5α-reductase. Interestingly, it was shown that plant *DET2* orthologs of the DET2 can substitute for each other, indicating a structural and functional conservation of steroid hormone signaling to a certain extent.[197] DET2 was shown to catalyze the reduction of (24*R*)-24-methylcholest-4-en-3-one (**51**) to (24*R*)-24-methyl-5α-cholestan-3-one (**52**) en route to campestanol (**53**) (**Figure 7**) in *A. thaliana* and in *Pisum sativum*.[198,199] The DET2 ortholog of *Gossypium hirsutum* was shown to play an important role in the basipetal growth of cotton fiber.[200] In *Solanum malacoxylon*, two isozymes of DET2 were identified by comparing the metabolization of 5α-campestanone to that of progesterone in different tissues.[201] The occurrence of DET2s was extended to fungi. A DET2 fungal ortholog was isolated from *U. maydis* and was shown to be induced in plant–host parasitic interaction (corn smut disease).[202] The fungal DET2 was expressed in the Arabidopsis *det2-1* mutant, and this was sufficient to restore a wild-type phenotype in the plant mutant. Consequently, fungal, plant, and mammalian DET2s are most probably evolutionarily related. A series of dwarf Arabidopsis mutants isolated in a T-DNA insertional mutagenesis approach was also considered in the elucidation of the brassinosteroid biosynthetic pathway. The conversion of

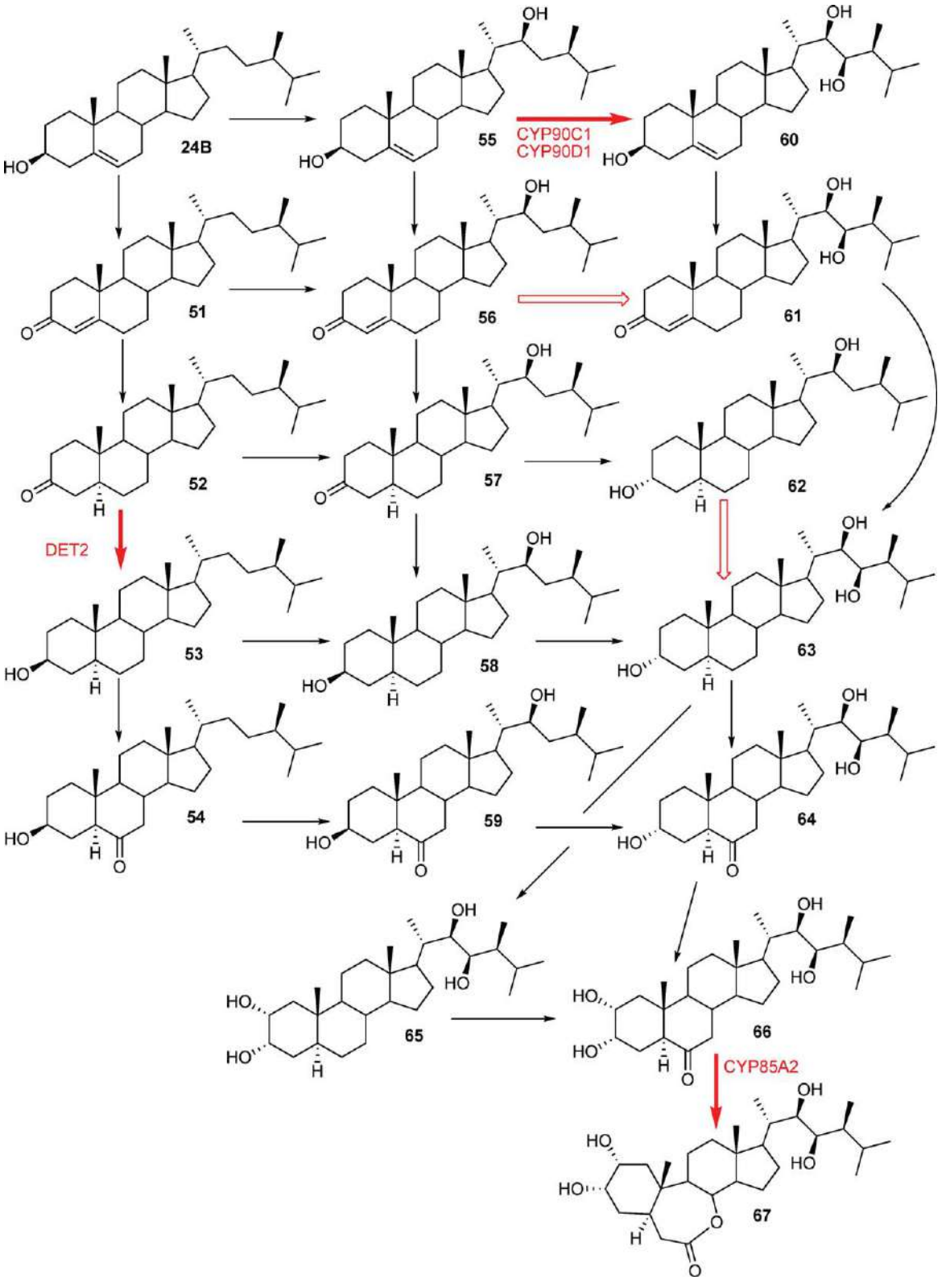


Figure 7 Brassinosteroid biosynthesis emphasizing short cuts of the C23-hydroxylation (large red arrows). Adapted from T. Ohnishi; A. M. Szatmari; B. Watanabe; S. Fujita; S. Bancos; C. Koncz; M. Lafos; K. Shibata; T. Yokota; K. Sakata; M. Szekeres; M. Mizutani, *Plant Cell* **2006**, *18*, 3275–3288. **56**, 22-hydroxy-ergost-4-en-3-one; **57**, 22-hydroxy-5α-ergost-3-one; **58**, 6-deoxo-cathasterone; **59**, cathasterone; **60**, 22,23-dihydroxy-campesterol; **61**, 22,23-dihydroxy-ergost-4-en-3-one; **62**, 3-*epi*-6-deoxo-cathasterone; **63**, 6-deoxo-typhasterone; **64**, typhasterone; **65**, 6-deoxo-castasterone.

campesterol into C28 brassinosteroid intermediates, and then into catasterone (**66**) and brassinolide (**67**) after Baeyer–Villiger C6 oxidation to form a 7-oxolactonic B ring typical of brassinolide, was depicted in detail in *A. thaliana* and *Catharanthus roseus*.[194,203] Biosynthetic segments that form an intricate network of alternate routes between campesterol (**24**), (22*S*)-22-hydroxycampesterol (**55**), campestanol (**53**), 6-oxo-campestanol (**54**), 6-oxo-brassinosteroid intermediates, or 6-deoxo-brassinosteroid intermediates have been described as an early C22 oxidation branch,[204] a late C6 oxidation pathway, or an early C6 oxidation pathway. C23-hydroxylation shortcuts interspersed into those segments, which allow conversion of C22-hydroxylated intermediates into 6-deoxo intermediates,[205] are summarized in **Figure 7**. Cytochrome P-450 oxidases are involved not only in biosynthesis but also in the catabolism of brassinosteroids. Hydroxylation at C26 described in *A. thaliana* and *Lycopersicon esculentum* represents a possibility for the plant cell to inactivate brassinosteroids.[206] The essential functions of brassinosteroids in the regulation of biological processes are mediated by a signaling pathway whose components and molecular mechanisms include, BRI1, a plasma membrane receptor kinase, the transcription factors BES1 and BRZ1, and kinases and phosphatases at play in this signaling pathway.[207–209]

1.21.2.5 Steroids with Heterocycles in the Side Chain

1.21.2.5.1 Steroidal saponins

Two types of steroidal saponins are distributed within monocotyledones from the Asparagaceae, Costaceae, Poaceae, Dioscoreaceae, and Liliaceae and dicotyledones from the Solanaceae or Fabaceae.[210] Structurally, these compounds are spirostan (diosgenin (**68**) or furostan (nuatigenin (**69**)) sapogenins of which the hydroxyl at C3 is linked to an oligosaccharide (**Figure 8**). These compounds have been studied from the phytochemical and pharmacological aspects due to their medicinal properties and widespread uses. The occurrence and structural elucidation of steroidal saponins from *Dioscorea* species has driven several lines of research because of the industrial interest of diosgenin for steroid production. Steroidal saponins are described as anticancer agents along with other pharmacological properties.[211] Dioscin (**70**, **Figure 8**) from the yam[212] has antifungal activity against the human pathogenic yeast *C. albicans*. Other derivatives have antiallergic activity as was monitored using biochemical markers from specific test cell lines.[213] Molecular and genetic aspects of steroidal saponin biosynthesis are scarce. In *Avena sativa*, avenacosides (**71**) were converted into the fungicidal 26-desglucoavenacosides (**72**) upon infection by pathogens, and the hydrolytic fungal enzymes implicated in this process were isolated and described.[214] This type of hydrolytic enzyme, the 26-*O*-β-glucosidase, was purified from *Costus speciosus*. Its activity was high in stored rhizomes, where saponins underwent the transformation of spirostanol glycosides to furostanol glycosides.[215] Glycosylation of steroidal sapogenins and saponins is of

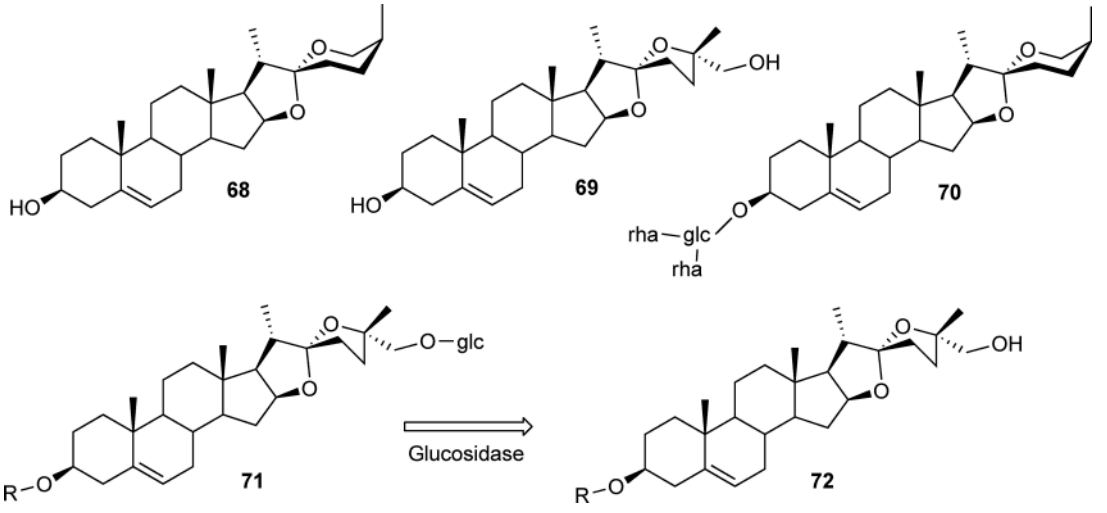


Figure 8 Sapogenins and steroidal saponins.

considerable importance for pharmacological activity. Glucosyltransferases involved in diosgenin (**68**), nuatigenin (**69**), and tigogenin metabolism have been cloned from *Solanum aculeatissimum*.[216] Their physiological role in plant biology has been related to plant defense. A functional study of these glycosyltransferases expressed in *E. coli* led to the identification of essential residues from the donor–sugar recognition domain of the protein.[217] One of these glycosyltransferases from *S. aculeatissimum* was shown to glucosylate steroidal alkaloids, in addition to saponins.

1.21.2.5.2 Steroidal glycoalkaloids

Steroidal glycoalkaloids are secondary metabolites found mainly in species belonging to the Solanaceae family. The structure is based on aglycones of the solanidane (e.g., solanidine (**73**)) or of the spirosolane (e.g., tomatidine (**74**)) types (**Figure 9**). Structural elucidation of these metabolites was documented over 50 years ago,[218] and structures continue to be reported.[219] Steroidal glycoalkaloids have been shown to play a role in plant–pathogen interactions. Resistance to bacterial and fungal diseases[220,221] and to insects[222,223] were considered in relation to the high contents of glycoalkaloids. Similarly, the tomato steroidal glycoalkaloid α-tomatine was shown to exert an antifungal activity associated with membrane permeabilization.[224] The well-known toxicity of α-chaconine and α-solanine (**75**) present in potato tubers produced for human consumption has also been linked to cell membrane permeabilization.[225] Upregulation of cholesterol biosynthetic genes in Caco-2 intestinal epithelial cells after treatment with α-chaconine was also reported.[226] The biosynthesis of steroidal glycoalkaloids has been studied in *Solanum melongena* and *S. tuberosum*. Biochemical studies support the view that cholesterol is the most probable precursor for the biosynthesis of solanidine.[227–229] A UDP-glucose:solasodine glucosyltransferase was partially purified from *S. melongena* leaves. The glucosyltransferase was able to glucosylate aglycones from the spirosolane type but not from the solanidane type.[230] In *S. tuberosum*, the biosynthesis of α-solanine and α-chaconine downstream to solanidine includes an UDP-galactose:solanidine galactosyltransferase (SGT1) catalyzing the conversion of solanidine into γ-solanine.[231] This gene was originally cloned as a UDP-glucose glucosyltransferase expressed from a potato cDNA library, transformed in yeast.[232] Other glucosyltransferases were isolated: a UDP-glucose:solanidine glucosyltransferase (SGT2) was implicated in the production of α-chaconine. In tubers of potato plants expressing an antisense SGT2 construct, the accumulation of α-solanine was increased and α-chaconine was reduced.[232] Another transferase, SGT3, was identified as a glycosterol rhamnosyltransferase that was able to catalyze the terminal step in the formation of the triose side chain of α-solanine and α-chaconine from β-solanine and β-chaconine, respectively.[233] Regulatory aspects of potato steroidal glycoalkaloid biosynthesis have been addressed in the context of the isoprenoid metabolism. A high steroidal glycoalkaloid content in tissues was associated with high expression levels of HMGR and squalene synthase.[234] The biosynthetic link between cholesterol and glycoalkaloid biosynthesis was addressed in the transgenic potato overexpressing a cycloartenol-C24-methyltransferase from *G. max* (*Gm*SMT1). The expression of *GmSMT1* led to increased levels of isofucosterol and sitosterol at the expense of cholesterol. Shortage of cholesterol resulted in an associated reduction of steroidal glycoalkaloids in leaves and tubers,[235] reinforcing the role of cholesterol as biosynthetic precursor of α-solanine and α-chaconine.

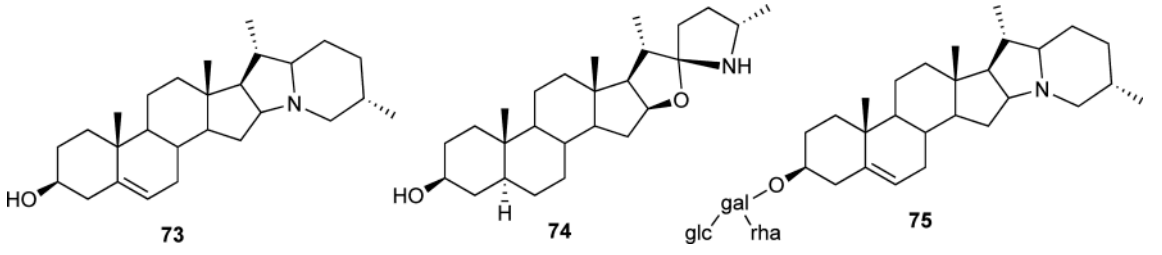


Figure 9 Steroidal glycoalkaloids.

1.21.2.5.3 Cardiotonic steroidal glucosides

These compounds are produced by plant species from diverse families such as Apocynaceae or Plantaginaceae but are also produced by amphibians. Historically, they have attracted considerable interest due to their medicinal properties. An epoxybufenolide series displaying a growth inhibition effect on the cancer cell line KB was described.[236] Recently, a series of plant cardiotonic steroids were shown to behave like potent splicing modulators in a test system (a reporter gene construct) designed for screening chemical libraries.[237] The steroidal aglycones of these compounds are cardenolides, bearing a lactone ring of five atoms at C17, or bufenolides, bearing a lactone ring of six atoms at C17.[2] *Digitalis purpurea* accumulates high levels of the cardenolides digitoxigenin (**76**) and digoxigenin (**77**) and high levels of the cardenolide glycosides (e.g., digoxin (**78**)) in leaves, with a diversity of sugar residues.[238,239] Cardenolide formation has been studied in *Digitalis lanata*, *D. purpurea*, and *Asclepias incarnata*, with an important focus on the activity of pregnane-modifying enzymes.[240–243] These steroids are thought to originate from cholesterol or possibly from 24-alkyl-sterols.[244] Cardenolide and pregnane biosynthesis requires a 3α-hydroxysteroid-5β-reductase.[245] A cDNA encoding a progesterone 5β-reductase (5β-POR) was cloned from *Digitalis lanata* leaves and functionally expressed in *E. coli* (**Figure 10**).[246] When progesterone was used as a substrate in this assay, the 5β isomers were formed exclusively. The crystal structures of 5β-POR in complexes containing progesterone indicated an architecture of the active site similar to that of SDR.[247] Other plant orthologs of 5β-POR were identified in *Isoplexis canariensis*,[248] another cardenolide-rich plant species, and in the model *A. thaliana*, which has not been reported as producing cardenolide.[249] The Arabidopsis protein functionally expressed in *E. coli* stereospecifically reduced progesterone to 5β-pregnane-3, 20-dione.[249] The corresponding gene was strongly transcribed in leaves. This Arabidopsis 5β-POR gene was originally described as a mutant allele negatively affecting cotyledon and leaf *v*ein *p*atterning, therefore called VEP1, required for normal vascular strand development.[250] A set of enzymes implicated in the bioconversion of pregnenolone into cardenolides in the model *D. lanata* included a Δ^5-3β-hydroxysteroid dehydrogenase (βHSD) and a Δ^5-3-oxosteroid isomerase. The *D. lanata* 3βHSD expressed in *E. coli* uses pregnenolone but not cholesterol as a substrate (**Figure 10**).[246] A malonyl-coenzymeA:21-hydroxypregnane 21-*O*-malonyltransferase involved in the formation of the butenolide ring of digitoxigenin was reported in leaves of *D. lanata*.[251] Cardenolides, produced in laticifers of Apocynaceae,[252] have been looked at in ecological approaches of plant–insect interactions to illustrate the defense-escalation theory and the evolutionary trends of secondary metabolism moving toward a decline of plant chemical defence strategies.[253]

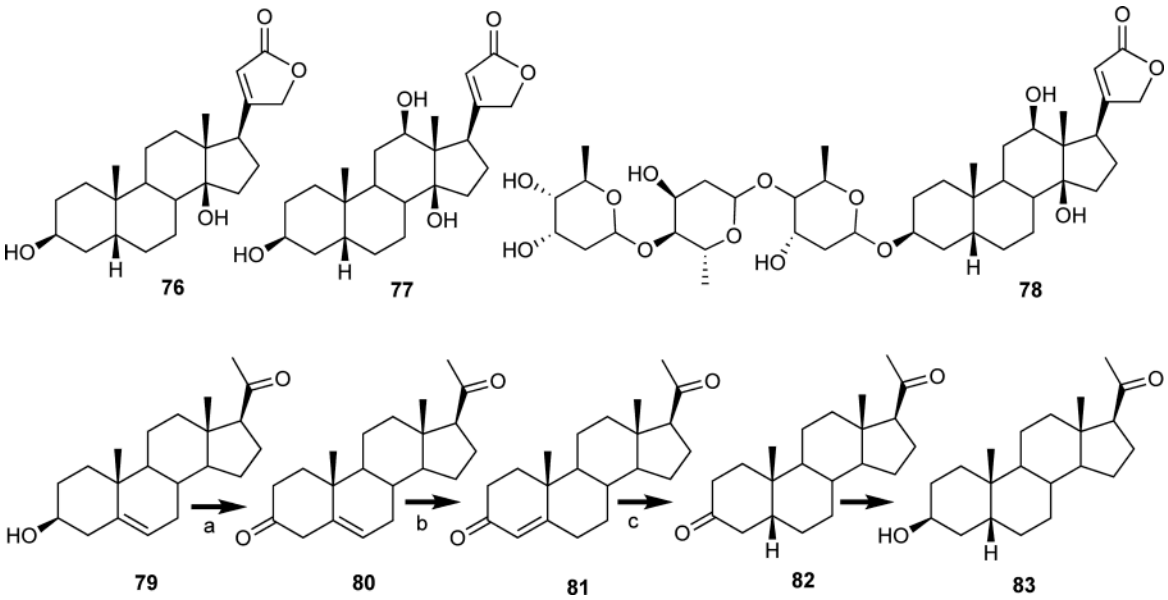


Figure 10 Cardenolide structure (**76–78**) and biosynthesis (**79–83**). a = 3β HSD, b = Δ^5-3-oxosteroid isomerase, and c = 5βPOR. 79, pregnenolone; 80, isoprogesterone; 81, progesterone; 82, 5β-pregnane-3, 20-dione; 83, 5β-pregnane-3β-ol-20-one.

1.21.3 Sterol and Steroid Conjugates

1.21.3.1 Steryl Esters

Steryl esters are ubiquitous sterol conjugates.[254] The fatty acyl moieties of these conjugates are usually representative of abundant compounds in a given organism such as C16 or C18 fatty acids. A diversity of sterol esters has been described in mammals,[255] in fungi,[256] and in plants.[257] Cholesterol is esterified in mammals by two distinct types of enzymes. Membrane-bound acyl-coA:cholesterol acyltransferases (ACAT) catalyze acyl-coA-dependent acylations in cells.[258] Soluble lecithin:cholesterol acyl transferase, a circulating enzyme present in the bloodstream, is evolutionarily unrelated to the ACAT type of enzymes.[259] Ergosterol is esterified in yeast by two ACAT-related enzymes (AREs), ARE1 and ARE2. These two proteins share a 49% identity, and their functional redundancy is indicated by the fact that sterol ester biosynthesis is not affected in an *are1* mutant, is reduced to 75% of its physiological level in an *are2* mutant, but is totally abolished in a double mutant *are1 are2*.[260] The absence of steryl ester-forming enzymes has no effect on cell viability in laboratory conditions. Enzyme activity measurements and fluorescence microscopy of proteins fused to the green fluorescent protein (GFP) indicated that both ARE1 and ARE2 proteins were localized in the endoplasmic reticulum. Are1p was shown to esterify the precursor and the end product, namely lanosterol and ergosterol, whereas are2p has a strong substrate preference for ergosterol.[261] Steryl esters are stored in lipid droplets or particles in yeast. Sterols in this storage form may be mobilized by the action of steryl ester hydrolases.[262] Three enzymes in yeast contribute to this process.[263] Experiments performed with *S. cerevisiae* support storage and mobilization of sterols as a dispensable process. However, enzymes implicated in there are the key elements of sterol homeostasis. Acetylation of sterols and steroids in yeast has been described as a detoxification pathway, including also a deacetylase. The acetyltransferase ATF2 and the steryldeacetylase SAY1 have been functionally characterized in a steroid export process.[264] In plants, the situation resembles that of mammals. Although biochemical studies performed with subcellular fractions have described sterol and acyl donors implicated in the reactions, molecular characterization of sterol ester-forming enzymes is recent. In the model *A. thaliana*, two genes encoding steryl ester-forming enzymes have been reported. One is related to the mammalian (lecithin cholesterol acetyltransferase) LCAT. Leaf microsomal membranes enriched with phospholipid:sterol acyltransferase PSAT1 (phospholipid sterol acyltransferase) catalyze the transacylation of fatty acyl moieties from the *sn-2* position of phosphatidylethanolamine.[265] The implication of this enzyme in regulatory aspects of sterol metabolism was indicated by the fact that sterol intermediates were preferentially esterified by the PSAT-rich fraction in the presence of pathway end products. Arabidopsis mutant lines deficient in PSAT1 are strongly depleted in steryl esters, but this has no effect on viability. Another sterol acyltransferase belongs to the family of plant membrane-bound *O*-acyltransferases related to the ACATs of yeast and animal. This plant sterol-*O*-acyltransferase, ASAT1, when expressed in yeast, catalyzes the production of lanosterol esters. Similarly, enzyme assays performed with subcellular fractions of this yeast indicated a substrate preference for cycloartenol as acyl acceptor and saturated fatty acyl coenzyme A as acyl donor. A seed-specific overexpression of *ASAT1* in *A. thaliana* resulted in elevated amounts of cycloartenol esters.[266] The sites of steryl esters accumulation in plants are of two types. Steryl esters accumulate in cellular lipid droplets when the amount of free sterol synthesized is higher than that normally required to build membrane structures. This has been shown with the tobacco mutant *sterov* isolated in a somatic genetic approach[267–268] and in plant cell cultures fed with the upstream sterol precursor mevalonate.[269] Steryl esters are also deposited in the lipid bodies of elaioplasts in the tapetum. Tapetal cells in developing anthers of *Brassica napus* contain tapetosomes and elaioplasts, two types of organelles that are required for the pollen coat elaboration. Comparative analysis of sterols from the sterol ester-rich lipid bodies of elaioplasts and from the pollen coat indicated that a lipid coating of pollen is made with neutral lipid produced in the tapetum.[270,271] Species from the Poaceae (corn, wheat, rice, triticale, barley, and oat) contain steryl ferulates and other esters of phenylpropanoids localized in aleurone cells.[272] Cycloartenyl ferulates present in rice bran oil revealed their anti-inflammatory properties during pharmacological studies.[273]

1.21.3.2 Sterol Glucosides

Uridine diphospho (UDP)-glucose sterol-β-D-glucosyltransferase (USGT) is a plasma membrane–bound enzyme.[274] Sterols recovered from the acid hydrolysis of sterol glucoside fractions are usually Δ^5-sterols (cholesterol, ergosterol, campesterol, and sitosterol),[254] but others have been reported, such as Δ^7-sterols in Leguminosae.[275] The USGT from *A. sativa* etiolated shoots was solubilized and purified up to 12,500-fold using a sepharose-based chromatographic process[276], and a cDNA was functionally expressed in *E. coli*.[277] Genes encoding USGT were isolated from *S. cerevisiae*, *C. albicans*, *Pichia pastoris*, and *D. discoidum* and functionally expressed.[278] The characterization of a yeast mutant deficient in sterol glucosylation indicated that this biochemical process was apparently dispensable because the lack of USGT had no effect on cell viability. An alternative pathway for the synthesis of sterol glucosides has been proposed in Arabidopsis, based on the use of a mutant of *P. pastoris* deficient in the production of glucosylceramides and of sterol glucosides. A plant glucosylceramide synthase was expressed in this double mutant. This resulted in the synthesis of glucosylceramides and of sterol glucosides, therefore indicating the presence of multiple pathways for sterol glucoside synthesis in plants.[279] Sterol glucosides in *P. pastoris* have been described as enhancers for the autophagic process of peroxisome degradation.[280] Sterol glucoside biosynthetic capability has been reported for prokaryotes. *Helicobacter pylori*, the pathogen responsible for gastric ulcers and carcinoma, is a sterol auxotroph but contains a gene Hp0421 that encodes cholesterol-α-glucosyltransferase.[281] A bacterial strain lacking a functional Hp0421 was generated and used to point out the important role of sterol glucosides in the interaction of bacteria with the host. It was concluded that cholesterol glycosylation promotes immune evasion by *H. pylori*.[282] Plasma membrane lipid alterations, and particularly the sterol glucosides alterations, were shown to be associated with cold acclimatization of plants.[283] Biophysical studies have addressed the functional role of sterol glucosides in membrane models.[284,285] The possible role of sterol glucosides in plants as primers for cellulose biosynthesis has been addressed experimentally. Microsomes prepared from cotton fibers synthesize sitosterol-cellodextrins from sitosterol glucoside and UDP-glucose under conditions that favor cellulose synthesis.[286] Studies on plant–microbe interaction revealed that USGT is required for pathogenicity. Indeed, conidia of *Colletotrichum gloeosporioides* use it as a virulence factor whose synthesis is induced by surface contact.[287]

1.21.3.3 Acylated Sterol Glucosides

The synthesis of acylated steryl glucosides has been reported in plants.[254] Fatty acyl moieties have been identified as well as acyl donors.[288] A membrane-bound phospholipid:steryl glucoside acyltransferase from *Solanum melalonga* was partially purified and shown to acylate phytosteryl glucosides. The preferred acyl donors were phosphoglycerolipids compared to 1,2-diacylglycerols, particularly 1,2-dimirystoylphosphatidylic acid, providing the acyl moiety from the *sn-1* position.[289] An acylated sterol glucoside, β-sitosteryl-3-*O*-β-D-glucopyranosyl-6′-*O*-palmitate, exhibited high anticomplementary activity following an activity-guided isolation from *Orostachys japonicus* extracts. Interestingly, the corresponding nonacylated sterol glucoside had no pharmacological activity.[290]

1.21.3.4 Steroid Sulfates

Sulfate conjugation by sulfotransferases is a process that participates in regulating the biological activity of steroid hormones in mammals,[291] a process that has been described in plants.[292] In *B. napus*, a steroid sulfotransferase catalyzes an O-sulfonation of brassinosteroids and also of animal steroids. This plant enzyme is specific for the hydroxyl group at C22 and preferentially uses biosynthetic intermediates. It has been shown that sulfonation of 24-epibrassinolide abolishes its activity in a functional assay. Hormone inactivation by sulfonation is, therefore, a common mechanism distributed in eukaryotes.[293] As in the case of *B. napus*, *A. thaliana* contain two distinct brassinosteroid sulfotransferases that share a 44% identity and defined distinct plant sulfotransferase families. These enzymes exhibited partial overlapping functions with human dehydroepiandrosterone sulfotransferase. Molecular characterization of the plant steroid sulfotransferases revealed transcription specificities, particularly induction by cytokinins. The reaction catalyzed by the sulfotransferase *AtST1* was shown to be stereospecific for 24-epibrassinosteroids.[294] The expression of the sulfotransferase in *B.*

napus was induced by salicylic acid, indicating that steroid-dependent biological response may be at play in plant defense processes.

1.21.4 Sterol Degradation

Sterol degradation is known to occur in actinobacteria such as *Nocardia rhodochrous*. The potential for bacteria to provide genetic resources for steroid biotechnology has recently driven more research. Cholesterol oxidases produce cholest-4-en-3-one, as they were shown to have both 3β-hydroxysteroid dehydrogenase and 3-oxo-steroid Δ^4–Δ^5-isomerase activities in acellular fractions.[295] *Sterolibacterium denitrificans* is a proteobacteria from the Rhodocyclaceae that can also degrade cholesterol to carbon dioxide under anoxic conditions. A cholest-4-en-3-one-Δ^1-dehydrogenase partially purified, sequenced as tryptic peptides, then cloned and expressed in *E. coli*, accepted cholest-4-en-3-one and other mammalian steroids as substrates.[296] This FAD-dependent enzyme is similar to 3-oxo-steroid-Δ^1-dehydrogenase (and to other enzymes from the SDR group), which is involved in aerobic degradation of steroids by another proteobacteria *Pseudomonas testosteroni*.[297] A bioinformatic-based study of an actinobacteria from the *Rhodococcus* genus (also known for plasiticity in bioconversion/biodegradation processes) provided a compendium of genes responsible for catabolism of cholesterol to propionyl coenzyme A and pyruvate.[298] The related pathogenic genus *Mycobacterium* displayed a conserved cholesterol catabolic pathway that enabled *M. tuberculosis*, for example, to grow on cholesterol *in vitro* and, most importantly, to survive in macrophages.

1.21.5 Transport

Aspects of steroid transport at the cellular or organism levels are known across prokaryotes and eukaryotes. Actinomycetales (*Mycobacterium*, *Rhodococcus* for instance) possess *mce* loci (mammalian cell entry), on which *mce* genes are arranged in operons.[299] These genes are upregulated during growth on cholesterol. A genetic approach using *Rhodoccocus jostii* deletion mutants of genes belonging to the *mce4* locus, combined with cholesterol and other steroid uptake assays, demonstrated that *mce4* encodes a sterol uptake system that is a steroid transporter of the ABC type.[300] Cholesterol metabolism and circulation in mammals is closely linked to its transport across cell membranes toward lipoproteins of the bloodstream, and this is mediated by ABC transporters.[301] In human enterocytes, transporters ABC G5 and G8 are responsible for 24-alkyl-sterols (plant sterols from the diet) efflux to the intestinal lumen. Mutations in these transporters cause the rare disease sitosterolemia.[302] Sterol carrier proteins (SCP) in animals are described as intracellular transporters of lipids with important roles in membrane biology and have been extensively studied over the past 30 years.[303] Structure–activity of the human SCP2, which facilitated such lipid transport, was determined using an assay of cholesterol and phosphatidylcholine transferred from small donor unilamellar vesicles to acceptor membrane of bacterial protoplasts.[304] Binding experiments suggested an interaction of cholesterol or fatty acids with SCP2, consistent with the role of an intracellular aqueous carrier or of an enhancer of sterol desorption from membranes.[305,306] The crystal structure of an ortholog of SCP2, determined at 1.8 Å resolution, revealed a hydrophobic tunnel suitable for lipid binding. Closely similar cavities were found in plant lipid transfer proteins (LTPs).[307] An ortholog of SCP2 was found in insects.[308] Another ortholog of SCP2 identified in *A. thaliana* showed lipid transfer activity.[309] The *A. thaliana* SCP2 is a ubiquitous peroxisomal protein essential for seed morphology and germination.[310] The evolution of SCP2 indicated a large distribution among living organisms and essential functions.[311] This is also true for oxysterol-binding proteins, which are implicated in sterol metabolism, transport, trafficking, and signaling, and which are found in animals, yeast, and plants (reviewed in Fairn and McMaster[312] and Javitt[313]). An *o*xysterol-binding protein-*r*elated *p*rotein (ORP) from *Petunia inflata* was shown to interact with a receptor kinase in pollen. This ORP allowed the detection of 12 genes encoding ORPs in *A. thaliana*, which were not yet assigned to physiological functions.[314] In oomycetes from the *Phytophthora* genus, a secreted protein called cryptogein was able to bind a sterol and to transfer it between phospholipidic bilayers.[315] The function of such proteins was related to plant–pathogen interactions.[316]

1.21.6 Molecular Regulation of Sterol Biosynthesis

Although cholesterol and structurally related sterols are distributed in bacteria, fungi, plants, and animals, the molecular regulation of the biosynthesis and accumulation of these compounds in different organisms may be diverse.

Cholesterol homeostasis in mammalian cells is controlled by feedback mechanisms acting at the transcriptional and the posttranscriptional levels of genes implicated in the synthesis of steroidogenic or lipidogenic enzymes or in the production of regulatory elements. Transcription of genes encoding HMGR, HMGcoA synthase, and the LDL (low-density lipoprotein) membrane-bound receptor increases when there is a cellular need for sterol.[317] Common *cis*-elements found in promoter sequences of these genes named SRE[318] were found to be the binding sites of a class of transcription factors, the SREBP (sterol regulatory element-binding proteins).[319] SREBP precursors of 125 kDa are composed of an N-terminal domain, which is the transcription factor itself, a member of the basic helix–loop–helix leucine zippers (bHLHZ), a central hydrophobic region composed of two transmembrane domains, and a C-terminal domain implicated in SREBP processing.[320] SREBPs are anchored in the endoplasmic reticulum and in the nuclear envelope. In the absence of cholesterol, SREBPs undergo intramembrane proteolytic cleavage, and the transcription activator bHLHZ translocates to the nucleus where it increases expression of target genes.[321] The proteolysis of SREBP implicates a SREBP cleavage activation protein (SCAP) in physical interaction with SREBP. SCAP–SREBP interaction enables the action of two distinct proteases S1P and S2P (sites 1 or 2 proteases).[322,323] HMGR has a central regulatory role in cholesterol homeostasis at the transcriptional but also at the translational levels. The mammalian HMGR contains eight helical membrane-spanning domains essential to sterol-mediated proteolysis. This was shown using reporter proteins whose half-life was monitored in the presence of cholesterol.[324] Sterol-sensing domains are present in HMGR and in SCAP and consist of approximately 180 amino acids forming five membrane-spanning domains. These domains have also been observed in other proteins implicated in lipid trafficking such as the Niemann–Pick disease type C1 protein (NPC1).[325] INSIG-1 (INSulin-Induced Gene) is another membrane protein that binds the sterol-sensing domain of SCAP in order to facilitate the retention of SCAP–SREBP complex in the ER.[326] In the absence of cholesterol, SCAP escorts SREBP toward the Golgi bodies proteolytic cleavage, thereby stimulating cholesterol synthesis.

Different types of ergosterol regulatory biosynthetic schemes are known in yeasts. A similar SREBP pathway was described in *Schizosaccharomyces pombe*.[327] It contains functional homologs of the SREBP pathway such as SREBP and SCAP. A microarray analysis has demonstrated that these regulatory elements enhance the sterol biosynthetic enzymes (as is the case in the animal system) but also functions as an oxygen sensor because these regulatory gene products activate hypoxia marker genes and, therefore, mediate oxygen-dependent sterol synthesis.[327] The pathogenic yeast *Cryptococcus neoformans* (causing meningoencephalitis) also possesses an SREBP, which is activated under low oxygen and triggers ergosterol biosynthesis.[328] Although insects are sterol auxotrophs, the genome of *D. melanogaster* contain all components of the SREBP pathway and were shown to be functional in Drosophila cells in culture.[329] Processing of the transcription factor is regulated by fatty acids, and not by sterols, and therefore favors transcription of genes implicated in fatty acid synthesis. From an evolutionary point of view, these observations underline the essential role of SREBP pathways in membrane integrity in the first place, and not only in cholesterol biosynthesis. Insects also share other components with mammals that are implicated in the regulatory effect of cholesterol synthesis and in neurodegeneration.[330]

The yeast model *S. cerevisiae* does not contain the SREBP pathway. However, it controls the biosynthesis of ergosterol at the HMGR level and other mevalonate pathway enzymes.[331,332] Additionally, another rate-limiting biosynthetic step was identified beyond the committed precursor squalene: the overexpression of a cytosolic HMGR led indeed to high squalene accumulation,[333] therefore identifying SQE as a bottleneck. This is not the case for other organisms, for example, plants in which HMGR represents a major bottleneck: its overexpression results in the accumulation of high amounts of pathway end product as steryl esters stored in cytosolic lipid droplets.[334] The regulated degradation of HMGR is an element of ergosterol homeostasis.[335] Levels of ergosterol have been implicated in the transcriptional regulation of sterol biosynthetic genes such as ERG10[336] or ERG3.[337] In the latter experiments, the promoter sequence of ERG3 that was fused to the bacterial reporter lacZ drove increasing levels of β-galactosidase activity, which was inversely proportional to

the quantity of ergosterol. A regulatory protein SUT1 of *S. cerevisiae* was involved in sterol uptake under hypoxia.[338] A transcriptional regulatory mechanism alternative to the SREBP pathway at play in *S. pombe* was described in *S. cerevisiae*. It is made up of two transcription factors upc2p and ecm22p, which are members of the fungal Zn2-Cys(6) binuclear cluster family, and bind to promoter sequences of *ERG* genes[339,340] and other regions of the same ERG genes.[341] *S. cerevisiae* also has a transcriptional repressor mot3, which interacts with ergosterol biosynthetic genes and plays a role in vacuolar and membrane transport.[342]

A series of oxysterols have been implicated in cellular cholesterol homeostasis. The oxidases that catalyze the synthesis of 7-hydroxycholesterol, 24-hydroxycholesterol, and 27-hydroxycholesterol are cytochrome P-450 oxygenases, whereas 25-hydroxycholesterol is made by a nonheme iron oxygenase. Among this series of molecules, 25(*R*)-hydroxycholesterol was the most potent regulator of HMGR in hepatic cells.[343] Oxysterol-binding proteins have been recently described as possible key elements of nonvesicular sterol transport.[344] Liver X receptors are nuclear proteins that bind oxysterols and act as transcription factors of genes implicated in sterol and lipid biosynthesis.[345] Pharmacological studies have shown that 7-hydroxycholesterol shows antitumor activity.[346] Interestingly, oxysterols were shown to be also nonenzymatic products in food due to autooxidation. The consequence of this process in functional foods was discussed.[10]

Plants must regulate their sterol biosynthesis and cellular lipid homeostasis by alternate systems because they do not have a SREBP pathway (as indicated by genome data mining). These alternate pathways are neither understood nor identified. However, the key regulatory role of the enzyme HMGR in the production and accumulation of steady-state levels of phytosterols is a trait that plants share with other eukaryotes. Arabidopsis or tobacco plants overexpressing HMGR, or deficient in the expression of HMGR, consequently have a higher or lower amount of sterol.[334,347] The cellular machinery linking a putative sensing system in membranes with action on gene activity is unknown. Light has a probable role in this process. It was indeed shown that phytochrome A, B, C, and D, which are major light receptors in plants, act as negative regulators of HMGR in Arabidopsis.[348]

1.21.7 Functions of Steroids

Major functions of sterols and steroids are related to cell membrane structure and hormonal functions and, are, therefore, typical of essential metabolites. Steroidal saponins, steroidal glycoalkaloids, and cardenolides belong to the category of natural products (secondary metabolites), which possess various biological and pharmacological activities linked to environmental interactions.

The crucial importance of sterols in regulating physical properties of membranes has long been known.[349] The formation of membrane microdomains, also known as lipid rafts[350] was shown to be favored by cholesterol, ergosterol, or plant sterols, together with sphingolipids.[351] These membrane domains have been described over the last 15 years as platforms supporting signal transduction[352] and host–pathogen interactions[353] in mammals. Membrane microdomains have been operationally characterized by their insolubility in nonionic detergents.[354] Such lipid rafts have been isolated and characterized in plant cells[355] and in yeast.[356] In the latter, ergosterol and membrane microdomains have an essential role in generation and maintainance of cell polarity during mating.[357] In the case of higher plants that are sessile organisms, the multiplicity of pathway end products (24-alkyl-Δ^5-sterols), compared to the uniqueness of cholesterol in vertebrates or ergosterol in fungi, has been discussed as a way to regulate membrane thermal shocks.[358] Indeed, ordering/disordering properties of membrane models made of sterols, sphingolipids, and deuterated dipalmitoylphosphatidylcholine (therefore mimicking natural rafts) were analyzed by solid-state ^{2}H-NMR in this study. Raft mimics with plant sterols showed less sensitivity to temperature variations than mimics with cholesterol or ergosterol.

The essentialness of Δ^5-sterols (therefore considered as primary metabolites) in multicellular organisms has been illustrated over the last decade by mostly biological and genetic approaches. The nematode *Caenorhabditis elegans*, which is an auxotroph to cholesterol, was grown on cholesterol or on synthetic *ent*-cholesterol for comparison. The enantiomer of cholesterol was unable to support development of *C. elegans*, a result which demonstrated that the absolute configuration of cholesterol, in addition to its biophysical properties, was essential.[359] In the model plant *A. thaliana*, the characterization of sterol biochemical mutants, which were affected in biosynthetic steps between squalene and the end products campesterol and sitosterol, clearly showed

that sterol intermediates are unable to support normal growth and development in replacing Δ^5-sterols structurally or as precursors of brassinosteroids (see Section 1.21.2.4.3). The same consideration is true in humans in the case of loss-of-function mutations that affect cholesterol biosynthetic genes: mutations in NSDHL (3β-hydroxysteroid dehydrogenase) cause the CHILD syndrome,[360] mutations in the Δ^8–Δ^7-sterol isomerase cause the Conradi–Hünermann–Happle syndrome,[361] mutations in the human CYP51 cause the Antley–Bixler syndrome,[362] and mutations in the Δ^7-reductase cause the Smith–Lemli–Opitz syndrome (SLOS).[363] All these syndromes consist of severe developmental defects.

Cholesterol also exerts a crucial role in animal (insect and vertebrates) development as part of its signaling processes. In fact, cholesterol covalently modifies HEDGEHOG-secreted signaling proteins that are essential in embryogenesis.[364–366] Covalent binding of cholesterol to the N-terminal end of the signaling protein is achieved by the cholesterol transferase activity of its C-terminal end. In the signaling process, HEDGEHOG binds its PATCHED receptor, which contains a conserved sterol-sensing domain.[367] Interestingly, mammalian meiosis was activated by intermediates of the cholesterol pathway MAS (meiosis-activating sterols).[368]

Abbreviations

ABC	ATP-binding cassette
ACAT	acyl-coA-cholesterol acyltransferase
ARE	ACAT-related enzymes
bHLHZ	basic helix–loop–helix leucine zipper
BRI	brassinosteroid insensitive
BZR	brassinazole resistant
CPI	cyclopropylsterol isomerase
CRE/*lox*P	cyclization recombination locus of X-over P
CVP	cotyledon vein patterning
CYP51	sterol-14-demethylase
CYP710	sterol-22-desaturase
CYP85A2	7-oxolactone synthase (brassinolide synthase)
CYP90C1	brassinosteroid-C23-hydroxylase
CYP90D1	brassinosteroid-C23-hydroxylase
DET2	steroid-5α-reductase
DIMINUTO/DWARF1	sterol-Δ^{24}-isomerase/reductase
DWARF5	$\Delta^{5,7}$-sterol Δ^7-reductase
EBP	emopamil-binding protein
EMS	ethyl methane sulfonate
ERG11/CYP51	lanosterol-14-demethylase
ERG11/CYP51	sterol-14-demethylase
ERG2	Δ^8-sterol-Δ^8-Δ^7-sterol isomerase
ERG24	$\Delta^{8,14}$-sterol-Δ^{14}-reductase
ERG25/SMO	sterol C4-methyl oxidase
ERG26/ 3βHSD/D	4α-carboxysterol-3β-hydroxysteroid/C4-decarboxylase
ERG27/SR	3-oxosteroid reductase
ERG3	Δ^7-sterol-C5(6)-desaturase
ERG4	sterol-Δ^{24}-isomerase/reductase
ERG5/CYP61	sterol-22-desaturase
ERG7	lanosterol synthase
FACKEL	$\Delta^{8,14}$-sterol-Δ^{14}-reductase
FAD	flavin adenine dinucleotide
GC–MS	gas chromatography–mass spectrometry
GFP	green fluorescent protein
HMGR	3-hydroxy-3-methylglutaryl coenzyme A reductase

HSP	heat-shock protein
HYDRA1/SI	Δ^8-sterol-Δ^8-Δ^7-sterol isomerase
LBR	lamin B receptor
LCAT	lecithin cholesterol acetyltransferase
LDL	low-density lipoprotein
LTP	lipid transfer protein
NADH	nicotinamide adenine dinucleotide, reduced form
NADPH	nicotinamide adenine dinucleotide phosphate, reduced form
NMR	nuclear magnetic resonance
NSDHL	3β-hydroxysteroid dehydrogenase
ORP	oxysterol-binding protein-related protein
OSC/CAS1	oxidosqualene cyclase/cycloartenol synthase
OSC/LAS1	oxidosqualene cyclase/lanosterol synthase
OSC/βAMS	oxidosqualene cyclase/β-amyrin synthase
PSAT	phospholipid sterol acyltransferase
SCAP	SREBP cleavage activation protein
SCP	sterol carrier protein
SDR	short-chain dehydrogenase/reductase
Seladin	selective Alzheimer's disease indicator
SGT	solanidine glucosyl transferase
SMO	sterol-4α-methyl-oxidase
SMT/ERG6	zymosterol-C24-methyltransferase
SMT1	cycloartenol-C24-methyltransferase
SMT2	24-methylene-C24^1-methyltransferase
SQC	squalene cyclase
SQE	squalene epoxidase
SR	sterone reductase
SRE	sterol regulatory element
SREBP	sterol regulatory element-binding proteins
STE1/DWARF7/BUL1	Δ^7-sterol-C5(6)-desaturase
T-DNA	transfer DNA
UDP	uridine diphosphate
USGT	(UDP)-glucose sterol-β-D-glucosyltransferase
VEP	vein patterning
VIGS	virus-induced gene silencing
3βHSD/D	4α-carboxysterol-3β-hydroxysteroid dehydrogenase/C4-decarboxylase
5βPOR	progesterone 5β-reductase

References

1. W. R. Nes; M. McKeen, *Biochemistry of Steroids and Other Isopentenoids*; University Park Press: Baltimore, 1977.
2. J. W. Cornforth, *Biochem. Biophys. Res. Commun.* **2002**, *292*, 1129–1138.
3. R. A. Moreau; B. Whitaker; K. B. Hicks, *Prog. Lipid. Res.* **2002**, *41*, 457–500.
4. P. Benveniste, *Annu. Rev. Plant Physiol.* **1986**, *37*, 275–307.
5. G. Ourisson, *J. Plant Physiol.* **1994**, *143*, 434–439.
6. P. Benveniste, *Annu. Rev. Plant Biol.* **2004**, *55*, 429–457.
7. F. Bouvier; A. Rahier; B. Camara, *Prog. Lipid Res.* **2005**, *44*, 357–429.
8. G. P. Moss, *Pure Appl. Chem.* **1989**, *61*, 1783–1822.
9. A. Rahier; P. Benveniste, Mass Spectral Identification of Phytosterols. In *Analysis of Sterols and Other Biologically Significant Steroids*; W. D. Nes, E. J. Parish, Eds.; Academic Press: New York, 1989; pp 223–250.
10. J. L. Goad; T. Akihisha, *Analysis of Sterols*; Blackie Academic and Professional: London, 1997.

11. B. S. J. Blagg; M. B. Jarstfer; D. H. Rogers; C. D. Poulter, *J. Am. Chem. Soc.* **2002**, *124*, 8846–8853.
12. A. Tippelt; L. Jahnke; K. Poralla, *Biochim. Biophys. Acta* **1998**, *1391*, 223–232.
13. M. Rohmer; P. Bouvier; G. Ourisson, *Eur. J. Biochem.* **1980**, *112*, 557–560.
14. C. Nakano; A. Motegi; T. Sato; M. Onodera; T. Hoshino, *Biosci. Biotechnol. Biochem.* **2007**, *71*, 2543–2550.
15. A. Pearson; M. Budin; J. J. Brocks, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15352–15357.
16. H. B. Bode; B. Zeggel; B. Silakowski; S. C. Wenzel; H. Reichenbach; R. Müller, *Mol. Microbiol.* **2003**, *47*, 471–481.
17. J. Shinozaki; M. Shibuya; K. Masuda; Y. Ebizuka, *Phytochemistry* **2008**, *69*, 2559–2564.
18. J. M. Rasbery; H. Shan; R. J. LeClair; M. Norman; S. P. Matsuda; B. Bartel, *J. Biol. Chem.* **2007**, *282*, 17002–17013.
19. R. Y. Yang; L. L. Feng; X. Q. Yang; L. L. Yin; X. L. Xu; Q. P. Zeng, *Planta Med.* **2008**, *74*, 1510–1516.
20. F. He; Y. Zhu; M. He; Y. Zhang, *DNA Seq.* **2008**, *19*, 270–273.
21. H. Uchida; R. Sugiyama; O. Nakayachi; M. Takemura; K. Ohyama, *Planta* **2007**, *226*, 1109–1115.
22. C. Ruckenstuhl; A. Eidenberger; S. Lang; F. Turnowsky, *Biochem. Soc. Trans.* **2005**, *33*, 1197–1201.
23. H. K. Lee; Y. F. Zheng; X. Y. Xiao; M. Bai; J. Sakakibara; T. Ono; G. D. Prestwich, *Biochem. Biophys. Res. Commun.* **2004**, *27*, 1–9.
24. I. Abe; M. Rohmer; G. D. Prestwitch, *Chem. Rev.* **1993**, *93*, 2189–2206.
25. S. Lodeiro; Q. Xiong; W. K. Wilson; M. D. Kolesnikova; C. S. Onak; S. P. T. Matsuda, *J. Am. Chem. Soc.* **2007**, *129*, 11213–11222.
26. D. R. Phillips; J. M. Rasbery; B. Bartel; S. P. T. Matsuda, *Curr. Opin. Plant Biol.* **2006**, *9*, 305–314.
27. M. Taton; P. Benveniste; A. Rahier; *Biochem. Biophys. Res. Commun.* **1986**, *138*, 764–770.
28. I. Abe; Y. F. Zheng; G. D. Prestwich, *Biochemistry* **1998**, *37*, 5779–5784.
29. E. J. Corey; S. P. Matsuda; B. Bartel, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11628–11632.
30. E. J. Corey; S. P. Matsuda; B. Bartel, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 2211–2215.
31. S. M. Godzina; M. A. Lovato; M. M. Meyer; K. A. Foster; W. K. Wilson; W. Gu; E. L. de Hostos; S. P. Matsuda, *Lipids* **2000**, *35*, 249–255.
32. F. S. Buckner; L. N. Nguyen; B. M. Joubert; S. P. Matsuda, *Mol. Biochem. Parasitol.* **2000**, *110*, 399–403.
33. S. P. Matsuda; L. B. Darr; E. A. Hart; J. B. Herrera; K. E. McCann; M. M. Meyer; J. Pang; H. G. Schepmann, *Org. Lett.* **2000**, *2*, 2261–2263.
34. T. K. Wu; J. H. Griffin, *Biochemistry* **2002**, *41*, 8238–8244.
35. M. J. Segura; S. Lodeiro; M. M. Meyer; A. J. Patel; S. P. Matsuda, *Org. Lett.* **2002**, *4*, 4459–4462.
36. T. K. Wu; C. H. Chang; Y. T. Liu; T. T. Wang, *Chem. Rec.* **2008**, *8*, 302–325.
37. M. D. Kolesnikova; Q. Xiong; S. Lodeiro; L. Hua; S. P. Matsuda, *Arch. Biochem. Biophys.* **2006**, *447*, 87–95.
38. M. Suzuki; T. Xiang; K. Ohyama; H. Seki; K. Saito; T. Muranaka; H. Hayashi; Y. Katsube; T. Kushiro; M. Shibuya; Y. Ebizuka, *Plant Cell Physiol.* **2006**, *47*, 565–571.
39. S. Sawai; T. Akashi; N. Sakurai; H. Suzuki; D. Shibata; S. Ayabe; T. Aoki, *Plant Cell Physiol.* **2006**, *47*, 673–677.
40. E. Babiychuk; P. Bouvier-Navé; V. Compagnon; M. Suzuki; T. Muranaka; M. Van Montagu; S. Kushnir; H. Schaller, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3163–3168.
41. K. Ohyama; M. Suzuki; J. Kikuchi; K. Saito; T. Muranaka, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 725–730.
42. R. Heintz; P. Benveniste, *J. Biol. Chem.* **1974**, *249*, 4267–4274.
43. A. Rahier; M. Taton; P. Benveniste, *Eur. J. Biochem.* **1989**, *181*, 615–626.
44. S. S. Wu; R. A. Moreau; B. D. Whitaker; A. H. Huang, *Lipids* **1999**, *34*, 517–523.
45. A. Grandmougin; P. Bouvier-Navé; P. Ullmann; P. Benveniste; M. A. Hartmann, *Plant Physiol.* **1989**, *90*, 591–597.
46. D. Raederstorff; M. Rohmer, *Eur. J. Biochem.* **1987**, *164*, 421–426.
47. M. F. Costet-Corio; P. Benveniste, *Pesticide Sci.* **1987**, *22*, 343–357.
48. H. Schaller; P. Maillot-Vernier; P. Benveniste; G. Belliard, *Phytochemistry* **1991**, *30*, 2547–2554.
49. M. A. Lovato; E. A. Hart; M. J. Segura; J. L. Giner; S. P. Matsuda, *J. Biol. Chem.* **2000**, *275*, 13394–13397.
50. S. Darnet; A. Rahier; *Biochem. Biophys. Acta* **2003**, *1633*, 106–117.
51. S. Men; Y. Boutté; Y. Ikeda; X. Li; K. Palme; Y. D. Stierhof; M. A. Hartmann; T. Moritz; M. Grebe, *Nat. Cell Biol.* **2008**, *10*, 237–244.
52. Z. A. Wojciechowski; L. J. Goad; T. W. Goodwin, *Biochem. J.* **1973**, *136*, 405–412.
53. A. C. Oehlschlager; R. H. Angus; A. M. Pierce; H. D. Pierce, Jr; R. Srinivasan, *Biochemistry,* **1984**, *23*, 3582–3589.
54. A. Rahier; J. C. Génot; F. Schuber; P. Benveniste; A. S. Narula, *J. Biol. Chem.* **1984**, *259*, 15215–15223.
55. R. F. Gaber; D. M. Copple; B. K. Kennedy; M. Vidal; M. Bard, *Mol. Cell Biol.* **1989**, *9*, 3447–3456.
56. L. M. Palermo; F. W. Leak; S. Tove; L. W. Parks, *Curr. Genet.* **1997**, *32*, 93–99.
57. E. Zinser; F. Paltauf; G. Daum, *J. Bacteriol.* **1993**, *175*, 2853–2858.
58. R. Leber; K. Landl; E. Zinser; H. Ahorn; A. Spök; S. D. Kohlwein; F. Turnowsky; G. Daum, *Mol. Biol. Cell* **1998**, *9*, 375–386.
59. G. G. Janssen; W. D. Nes, *J. Biol. Chem.* **1992**, *267*, 25856–25863.
60. P. A. Haughan; M. L. Chance; L. J. Goad, *Biochem. J.* **1995**, *308*, 31–38.
61. J. Shi; R. A. Gonzales; M. K. Bhattacharyya, *J. Biol. Chem.* **1996**, *271*, 9384–9389.
62. T. Husselstein; D. Gachotte; T. Desprez; M. Bard; P. Benveniste, *FEBS Lett.* **1996**, *381*, 87–92.
63. P. Bouvier-Navé; T. Husselstein; P. Benveniste, *Eur. J. Biochem.* **1998**, *256*, 88–96.
64. J. A. Marshall; W. D. Nes, *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1533–1536.
65. W. D. Nes; J. A. Marshall; Z. Jia; T. T. Jaradat; Z. Song; P. Jayasimha, *J. Biol. Chem.* **2002**, *277*, 42549–42556.
66. A. Schaeffer; P. Bouvier-Navé; P. Benveniste; H. Schaller, *Lipids* **2000**, *35*, 263–269.
67. N. Homberg; M. Harker; C. L. Gibbard; A. D. Wallace; J. C. Clayton; R. Safford, *Plant Physiol.* **2002**, *130*, 303–311.
68. H. Schaller; P. Bouvier-Navé; P. Benveniste, *Plant Physiol.* **1998**, *118*, 461–469.
69. A. C. Diener; H. Li; W. Zhou; W. J. Whoriskey; W. D. Nes; G. R. Fink, *Plant Cell* **2000**, *12*, 853–870.
70. V. Willemsen; J. Friml; M. Grebe; A. van de Toorn; K. Palme; B. Scheres, *Plant Cell* **2003**, *15*, 612–625.
71. A. Schaeffer; R. Bronner; P. Benveniste; H. Schaller, *Plant J.* **2001**, *25*, 605–615.
72. F. M. Carland; S. Fujioka; S. Takatsuto; S. Yoshida; T. Nelson, *Plant Cell* **2002**, *14*, 2045–2058.

73. Y. Hase; S. Fujioka; S. Yoshida; G. Sun; M. Umeda; A. Tanaka, *J. Exp. Bot.* **2005**, *56*, 1263–1268.
74. L. Arnqvist; P. C. Dutta; L. Jonsson; F. Sitbon, *Plant Physiol.* **2003**, *131*, 1792–1799.
75. D. Zweytick; C. Hrastnik; S. D. Kohlwein; G. Daum, *FEBS Lett.* **2000**, *470*, 83–87.
76. A. M. Pierce; A. M. Unrau; A. C. Oehlschlager; R. A. Woods, *Can. J. Biochem.* **1979**, *57*, 201–208.
77. C. Tiedje; D. G. Holland; U. Just; T. Höfken, *J. Cell Sci.* **2007**, *120*, 3613–3624.
78. T. Takahashi; A. Gasch; N. Nishizawa; N. H. Chua, *Genes Dev.* **1995**, *9*, 97–107.
79. U. Klahre; T. Noguchi; S. Fujioka; S. Takatsuto; T. Yokota; T. Nomura; S. Yoshida; N. H. Chua, *Plant Cell* **1998**, *10*, 1677–1690.
80. K. A. Feldmann; M. D. Marks; M. L. Christianson; R. S. Quatrano, *Science* **1989**, *243*, 1351–1354.
81. S. Choe; B. P. Dilkes; B. D. Gregory; A. S. Ross; H. Yuan; T. Noguchi; S. Fujioka; S. Takatsuto; A. Tanaka; S. Yoshida; F. E. Tax; K. A. Feldmann, *Plant Physiol.* **1999**, *119*, 897–907.
82. L. Du; B. W. Poovaiah, *Nature* **2005**, *437*, 741–745.
83. C. Largeau; J. L. Goad; T. W. Goodwin, *Phytochemistry* **1977**, *16*, 1925–1930.
84. I. Greeve; I. Hermans-Borgmeyer; C. Brellinger; D. Kasper; T. Gomez-Isla; C. Behl; B. Levkau; R. M. Nitsch, *J. Neurosci.* **2000**, *20*, 7345–7352.
85. A. Peri; G. Danza; M. Serio, *J. Endocrinol. Invest.* **2005**, *28*, 285–293.
86. D. Sarkar; T. Imai; F. Kambe; A. Shibata; S. Ohmori; A. Siddiq; S. Hayasaka; H. Funahashi; H. Seo, *J. Clin. Endocrinol. Metab.* **2001**, *86*, 5130–5137.
87. D. Di Stasi; V. Vallacchi; V. Campi; T. Ranzani; M. Daniotti; E. Chiodini; S. Fiorentini; I. Greeve; A. Prinetti; L. Rivoltini; M. A. Pierotti; M. Rodolfo, *Int. J. Cancer* **2005**, *115*, 224–230.
88. S. Hata; T. Nishino; H. Katsuki; Y. Aoyama; Y. Yoshida, *Biochem. Biophys. Res. Commun.* **1983**, *116*, 162–166.
89. S. L. Kelly; D. C. Lamb; A. J. Corran; B. C. Baldwin; L. W. Parks; D. E. Kelly, *FEBS Lett.* **1995**, *377*, 217–220.
90. B. A. Skaggs; J. F. Alexander; C. A. Pierson; K. S. Schweitzer; K. T. Chun; C. Koegel; R. Barbuch; M. Bard, *Gene* **1996**, *169*, 105–109.
91. S. L. Kelly; D. C. Lamb; D. E. Kelly, *FEBS Lett.* **1997**, *412*, 233–235.
92. H. Noda; Y. Koizumi, *Insect Biochem. Mol. Biol.* **2003**, *33*, 649–658.
93. A. D. Nusblat; L. Muñoz; G. A. Valcarce; C. B. Nudel, *J. Eukaryot. Microbiol.* **2005**, *52*, 61–67.
94. T. Morikawa; M. Mizutani; N. Aoki; B. Watanabe; H. Saga; S. Saito; A. Oikawa; H. Suzuki; N. Sakurai; D. Shibata; A. Wadano; K. Sakata; D. Ohta, *Plant Cell* **2006**, *18*, 1008–1022.
95. L. Arnqvist; M. Persson; L. Jonsson; P. C. Dutta; F. Sitbon, *Planta* **2008**, *227*, 309–317.
96. T. Morikawa; H. Saga; H. Hashizume; D. Ohta, *Planta* **2009**, *229*, 1311–1322.
97. G. F. Gibbons, *Biochem. J.* **1974**, *144*, 59–68.
98. M. Bard; D. A. Bruner; C. A. Pierson; N. D. Lees; B. Biermann; L. Frye; C. Koegel; R. Barbuch, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 186–190.
99. S. Darnet; M. Bard; A. Rahier, *FEBS Lett.* **2001**, *508*, 39–43.
100. L. Li; J. Kaplan, *J. Biol. Chem.* **1996**, *271*, 16927–16933.
101. S. Darnet; A. Rahier, *Biochim. Biophys. Acta* **2003**, *1633*, 106–117.
102. D. Gachotte; R. Barbuch; J. Gaylor; E. Nickel; M. Bard, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13794–13799. Erratum in: *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1810.
103. D. Gachotte; S. E. Sen; J. Eckstein; R. Barbuch; M. Krieger; B. D. Ray; M. Bard, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 12655–12660.
104. C. Mo; M. Valachovic; S. K. Randall; J. T. Nickels; M. Bard, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9739–9744.
105. C. Mo; M. Valachovic; M. Bard, *Biochim. Biophys. Acta* **2004**, *1686*, 30–36.
106. J. L. Gaylor; Y. Miyake; T. Yamano, *J. Biol. Chem.* **1975**, *250*, 7159–7167.
107. Y. Aoyama; Y. Yoshida; R. Sato; M. Susani; H. Ruis, *Biochim. Biophys. Acta* **1981**, *663*, 194–202.
108. S. Pascal; M. Taton; A. Rahier, *J. Biol. Chem.* **1993**, *268*, 11639–11654.
109. S. Rondet; M. Taton; A. Rahier, *Arch. Biochem. Biophys.* **1999**, *366*, 249–260.
110. S. Pascal; M. Taton; A. Rahier, *Arch. Biochem. Biophys.* **1994**, *312*, 260–271.
111. S. Darnet; A. Rahier, *Biochem. J.* **2004**, *378*, 889–898.
112. A. Rahier; S. Darnet; F. Bouvier; B. Camara; M. Bard, *J. Biol. Chem.* **2006**, *281*, 27264–27277.
113. M. E. Lucas; Q. Ma; D. Cunningham; J. Peters; B. Cattanach; M. Bard; B. K. Elmore; G. E. Herman, *Mol. Genet. Metab.* **2003**, *80*, 227–233.
114. Y. Aoyama; M. Noshiro; O. Gotoh; S. Imaoka; Y. Funae; N. Kurosawa; T. Horiuchi; Y. Yoshida, *J. Biochem* **1996**, *119*, 926–933.
115. P. Benveniste; A. Rahier, Target Sites of Sterol Biosynthesis Inhibition in Plants. In *Target Sites of Fungicide Action*; W. Koellers, Ed.; CRC Press: Boca Raton, 1992; pp 207–225.
116. S. L. Kelly; D. C. Lamb; D. E. Kelly, *FEMS Microbiol. Lett.* **1999**, *180*, 171–175.
117. A. Bellamine; A. T. Mangla; W. D. Nes; M. R. Waterman, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 8937–8942.
118. L. M. Podust; T. L. Poulos; M. R. Waterman, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3068–3073.
119. C. J. Jackson; D. C. Lamb; T. H. Marczylo; A. G. Warrilow; N. J. Manning; D. J. Lowe; D. E. Kelly; S. L. Kelly, *J. Biol. Chem.* **2002**, *277*, 46959–46965.
120. G. I. Lepesheva; W. D. Nes; W. Zhou; G. C. Hill; M. R. Waterman, *Biochemistry* **2004**, *43*, 10789–10799.
121. R. A. Kahn; S. Bak; C. E. Olsen; I. Svendsen; B. L. Moller, *J. Biol. Chem.* **1996**, *271*, 32944–32950.
122. S. Bak; R. A. Kahn; C. E. Olsen; B. A. Halkier, *Plant J.* **1997**, *11*, 191–201.
123. F. Cabello-Hurtado; M. Taton; N. Forthoffer; R. Kahn; S. Bak; A. Rahier; D. Werck-Reichhart, *Eur. J. Biochem.* **1999**, *262*, 435–446.
124. H. Schaller; P. Maillot-Vernier; L. Gondet; G. Belliard; P. Benveniste, *Biochem. Soc. Trans.* **1993**, *21*, 1052–1057.
125. H. B. Kim; H. Schaller; C. H. Goh; M. Kwon; S. Choe; C. S. An; F. Durst; K. A. Feldmann; R. Feyereisen, *Plant Physiol.* **2005**, *138*, 2033–2047.
126. C. Burger; S. Rondet; P. Benveniste; H. Schaller, *J. Exp. Bot.* **2003**, *54*, 1675–1683.
127. R. T. Lorenz; L. W. Parks, *DNA Cell Biol.* **1992**, *11*, 685–692.

128. J. H. Crowley; S. J. Smith; F. W. Leak; L. W. Parks, *J. Bacteriol.* **1996**, *178*, 2991–2993.
129. N. Jia; B. Arthington-Skaggs; W. Lee; C. A. Pierson; N. D. Lees; J. Eckstein; R. Barbuch; M. Bard, *Antimicrob. Agents Chemother.* **2002**, *46*, 947–957.
130. A. Akins, *Med. Mycol.* **2005**, *43*, 285–318.
131. G. Giaever; P. Flaherty; J. Kumm; M. Proctor; C. Nislow; D. F. Jaramillo; A. M. Chu; M. I. Jordan; A. P. Arkin; R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 793–798.
132. C. P. Woloshuk; H. D. Sisler; S. R. Dutky, *Antimicrob. Agents Chemother.* **1979**, *16*, 81–86.
133. M. Taton; P. Benveniste; A. Rahier, *Eur. J. Biochem.* **1989**, *185*, 605–614.
134. S. Smith, *Gene* **1995**, *155*, 139–140.
135. S. Silve; P. H. Dupuy; P. Ferrara; G. Loison, *Biochim. Biophys. Acta* **1998**, *1392*, 233–244.
136. N. Wagner; D. Weber; S. Seitz; G. Krohne, *J. Cell Sci.* **2004**, *117*, 2015–2028.
137. K. Schrick; U. Mayer; A. Horrichs; C. Kuhnt; C. Bellini; J. Dangl; J. Schmidt; G. Jürgens, *Genes Dev.* **2000**, *14*, 1471–1484.
138. J. H. Crowley; S. Tove; L. W. Parks, *Curr. Genet.* **1998**, *34*, 93–99.
139. W. H. Ashman; R. J. Barbuch; C. E. Ulbright; H. W. Jarrett; M. Bard, *Lipids* **1991**, *26*, 628–632.
140. J. P. Keon; C. S. James; S. Court; C. Baden-Daintree; A. M. Bailey; R. S. Burden; M. Bard; J. A. Hargreaves, *Curr. Genet.* **1994**, *25*, 531–537.
141. D. E. Kelly; M. E. Rose; S. L. Kelly, *FEMS Microbiol. Lett.* **1994**, *122*, 223–226.
142. S. Silve; P. Leplatois; A. Josse; P. H. Dupuy; C. Lanau; M. Kaghad; C. Dhers; C. Picard; A. Rahier; M. Taton; G. Le Fur; D. Caput; P. Ferrara; G. Loison, *Mol. Cell. Biol.* **1996**, *16*, 2719–2727.
143. S. Silve; P. H. Dupuy; C. Labit-Lebouteiller; M. Kaghad; P. Chalon; A. Rahier; M. Taton; J. Lupker; D. Shire; G. Loison, *J. Biol. Chem.* **1996**, *271*, 22434–22440.
144. R. J. Grebenok; T. E. Ohnmeiss, A. Yamamoto; E. D. Huntley ED; D. W. Galbraith; D. Della Penna, *Plant Mol. Biol.* **1998**, *38*, 807–815.
145. F. F. Moebius; K. Bermoser; R. J. Reiter; M. Hanner; H. Glossmann, *Biochemistry* **1996**, *35*, 16871–16888.
146. F. F. Moebius; R. J. Reiter; M. Hanner; H. Glossmann, *Br. J. Pharmacol.* **1997**, *121*, 1–6.
147. F. F. Moebius; B. U. Fitzky; G. Wietzorrek; A. Haidekker; A. Eder; H. Glossmann, *Biochem. J.* **2003**, *374*, 229–237.
148. M. Souter; J. Topping; M. Pullen; J. Friml; K. Palme; R. Hackett; D. Grierson; K. Lindsey K, *Plant Cell* **2002**, *14*, 1017–1031.
149. K. Schrick; S. Fujioka; S. Takatsuto; Y. D. Stierhof; H. Stransky; S. Yoshida; G. Jürgens, *Plant J.* **2004**, *38*, 227–243. Erratum in *Plant J.* **2004**, *38*, 562.
150. M. A. Hartmann, In *Lipid Metabolism and Membrane Biogenesis*; G. Daum, Ed.; Top Current Genetics (Rev Ser); Springer-Verlag: Berlin, 2003; pp 183–211.
151. B. A. Arthington; L. G. Bennett; P. L. Skatrud; C. J. Guynn; R. J. Barbuch; C. E. Ulbright; M. Bard, *Gene* **1991**, *102*, 39–44.
152. T. Akihisa (ne Itoh); S. Thakur; F. U. Rosenstein; T. Matsumoto, *Lipids* **1986**, *21*, 39–47.
153. S. Xu; C. W. Patterson; W. R. Lusby; K. M. Schmid; T. A. Salt, *Lipids* **1990**, *25*, 61–64.
154. D. Gachotte; R. Meens; P. Benveniste, *Plant J.* **1995**, *8*, 403–416.
155. S. Choe; T. Noguchi; S. Fujioka; S. Takatsuto; C. P. Tissier; B. D. Gregory; A. S. Ross; A. Tanaka; S. Yoshida; F. E. Tax; K. A. Feldmann, *Plant Cell* **1999**, *11*, 207–221.
156. M. Catterou; F. Dubois; H. Schaller; L. Aubanelle; B. Vilcot; B. S. Sangwan-Norreel; R. S. Sangwan, *Planta* **2001**, *212*, 659–672.
157. D. Gachotte; T. Husselstein; M. Bard; F. Lacroute; P. Benveniste, *Plant J.* **1996**, *9*, 391–398.
158. T. Husselstein; H. Schaller; D. Gachotte; P. Benveniste, *Plant Mol. Biol.* **1999**, *39*, 891–906.
159. S. Kawata; J. M. Trzaskos; J. L. Gaylor, *J. Biol. Chem.* **1985**, *260*, 6609–6617.
160. M. Taton; A. Rahier, *Arch. Biochem. Biophys.* **1996**, *325*, 279–288.
161. M. Matsushima; J. Inazawa; E. Takahashi; K. Suzumori; Y. Nakamura, *Cytogenet. Cell Genet.* **1996**, *74*, 252–254.
162. A. Rahier; P. Benveniste; T. Husselstein; M. Taton, *Biochem. Soc. Trans.* **2000**, *28*, 799–803.
163. A. Rahier, *Biochemistry* **2001**, *40*, 256–267.
164. M. Taton; A. Rahier, *Biochem. Biophys. Res. Commun.* **1991**, *181*, 465–473.
165. A. Rahier; M. Taton, *Biochemistry* **1996**, *35*, 7069–7076.
166. J. Aufenanger; J. Pill; F. H. Schmidt; K. Stegmeier, *Biochem. Pharmacol.* **1986**, *35*, 911–916.
167. E. Lecain; X. Chenivesse; R. Spagnoli; D. Pompon, *J. Biol. Chem.* **1996**, *271*, 10866–10873.
168. C. Duport; R. Spagnoli; E. Degryse; D. Pompon, *Nat. Biotechnol.* **1998**, *16*, 186–189.
169. S. Choe; A. Tanaka; T. Noguchi; S. Fujioka; S. Takatsuto; A. S. Ross; F. E. Tax; S. Yoshida; K. A. Feldmann, *Plant J.* **2000**, *21*, 431–443.
170. H. Schaller, *Prog. Lipid Res.* **2003**, *42*, 163–175.
171. J. T. Mullins; E. A. Ellis, *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 1347–1350.
172. P. A. Horgen; R. Smith; J. C. Silver; G. Craig, *Can. J. Biochem.* **1975**, *53*, 1341–1345.
173. B. Groner; N. Hynes; A. Sippel; G. Schutz, *Nature* **1976**, *261*, 599–601.
174. R. M. Riehl; D. O. Toft, *J. Biol. Chem.* **1984**, *259*, 15324–15330.
175. S. A. Brunt; R. Riehl; J. C. Silver, *Mol. Cell. Biol.* **1990**, *10*, 273–281.
176. S. A. Brunt; J. C. Silver, *Fungal Genet. Biol.* **2004**, *41*, 239–252.
177. J. H. Adler; R. J. Grebenok, *Lipids* **1995**, *30*, 257–262.
178. K. Nakanishi, *Steroids* **1992**, *57*, 649–657.
179. L. Dinan; S. D. Sarker; P. Bourne; P. Whiting; V. Sik; H. H. Rees, *Arch. Insect Biochem. Physiol.* **1999**, *41*, 18–23.
180. L. Zibareva; V. Volodin; Z. Saatov; T. Savchenko; P. Whiting; R. Lafont; L. Dinan, *Phytochemistry* **2003**, *64*, 499–517.
181. K. Vokac; M. Budesnsky; J. Harmatha; J. Kohoutova, *Phytochemistry* **1998**, *49*, 2109–2114.
182. M. Buszczak; W. A. Segraves, *Curr. Biol.* **2000**, *10*, 830–833.
183. R. S. Hewes, *Trends Endocrinol. Metab.* **2008**, *19*, 317–323.
184. D. Calas; D. Thiéry; F. Marion-Poll, *J. Chem. Ecol.* **2006**, *32*, 2443–2454.
185. J. Harmatha; L. Dinan, *Arch. Insect. Biochem. Physiol.* **1997**, *35*, 219–225.
186. L. Dinan, *Phytochemistry* **2001**, *57*, 325–339.

187. A. Maria; J. P. Girault; Z. Saatov; J. Harmatha; L. Dinan; R. Lafont, *J. Chromatogr. Sci.* **2005**, *43*, 149–157.
188. A. Bakrim; A. Maria; F. Sayah; R. Lafont; N. Takvorian, *Plant Physiol. Biochem.* **2008**, *46*, 844–854.
189. N. Reixach; R. Lafont; F. Camps; J. Casas, *Eur. J. Biochem.* **1999**, *266*, 608–615.
190. T. Savchenko; P. Whiting; A. Germade; L. Dinan, *Biochem. Syst. Ecol.* **2000**, *28*, 403–419.
191. B. Voigt; P. Whiting; L. Dinan, *Cell Mol. Life Sci.* **2001**, *58*, 1133–1140.
192. M. D. Grove; G. F. Spencer; W. K. Rohwedder, *Nature* **1979**, *281*, 216–217.
193. T. Yokota; M. Arima; N. Takahashi, *Tetrahedron Lett.* **1982**, *23*, 1275–1278.
194. A. Sakurai; T. Yokota; S. D. Clouse, Eds., *Brassinosteroids. Steroidal Plant Hormones*; Springer: Tokyo, 1999.
195. J. Li; P. Nagpal; V. Vitart; T. C. McMorris; J. Chory, *Science* **1996**, *272*, 398–401.
196. M. Szekeres; K. Németh; Z. Koncz-Kálmán; J. Mathur; A. Hauschmann; T. Altmann; G. Rédei; F. Nagy; J. Schell; C. Koncz, *Cell* **1996**, *85*, 171–182.
197. J. Li; M. G. Biswas; A. Chao; D. W. Russell; J. Chory, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3554–3559.
198. T. Noguchi; S. Fujioka; S. Takatsuto; A. Sakurai; S. Yoshida; J. Li; J. Chory, *Plant Physiol.* **1999**, *120*, 833–840.
199. T. Nomura; C. E. Jager; Y. Kitasaka; K. Takeuchi; M. Fukami; K. Yoneyama; Y. Matsushita; H. Nyunoya; S. Takatsuto; S. Fujioka; J. J. Smith; L. H. Kerckhoffs; J. B. Reid; T. Yokota, *Plant Physiol.* **2004**, *135*, 2220–2229.
200. M. Luo; Y. Xiao; X. Li; X. Lu; W. Deng; D. Li; L. Hou; M. Hu; Y. Li; Y. Pei, *Plant J.* **2007**, *51*, 419–430.
201. F. Rosati; G. Danza; A. Guarna; N. Cini; M. L. Racchi; M. Serio, *Endocrinology* **2003**, *144*, 220–229.
202. C. W. Basse; C. Kerschbamer; M. Brustmann; T. Altmann; R. Kahmann, *Plant Physiol.* **2002**, *129*, 717–732.
203. T. Katsumata; A. Hasegawa; T. Fujiwara; M. Notomi; H. Abe; M. Natsume; H. Kawaide, *Biosci. Biotechnol. Biochem.* **2008**, *72*, 2110–2117.
204. S. Fujita; T. Ohnishi; B. Watanabe; T. Yokota; S. Takatsuto; S. Fujioka; S. Yoshida; K. Sakata; M. Mizutani, *Plant J.* **2006**, *45*, 765–774.
205. T. Ohnishi; A. M. Szatmari; B. Watanabe; S. Fujita; S. Bancos; C. Koncz; M. Lafos; K. Shibata; T. Yokota; K. Sakata; M. Szekeres; M. Mizutani, *Plant Cell* **2006**, *18*, 3275–3288.
206. T. Ohnishi; T. Nomura; B. Watanabe; D. Ohta; T. Yokota; H. Miyagawa; K. Sakata; M. Mizutani, *Phytochemistry* **2006**, *67*, 1895–1906.
207. G. Vert; J. L. Nemhauser; N. Geldner; F. Hong; J. Chory, *Annu. Rev. Cell Dev. Biol.* **2005**, *21*, 177–201.
208. G. Vert; J. Chory, *Nature* **2006**, *441*, 96–100.
209. Y. Yin; D. Vafeados; Y. Tao; S. Yoshida; T. Asami; J. Chory, *Cell* **2005**, *120*, 249–259.
210. W. Eichenberger, *Plant Cell Rep.* **1982**, *1*, 253–256.
211. M. Sautour; A. C. Mitaine-Offer; M. A. Lacaille-Dubois, *J. Nat. Med.* **2007**, *61*, 91–101.
212. M. Sautour; A. C. Mitaine-Offer; T. Miyamoto; A. Dongmo; M. A. Lacaille-Dubois, *Chem. Pharm. Bull. (Tokyo)* **2004**, *52*, 1353–1355.
213. S. Tewtrakul; A. Itharat, *Bioorg. Med. Chem.* **2006**, *14*, 8707–8711.
214. J. P. Morissey; J. P. Wubben; A. E. Osbourn, *Mol. Plant Microbe Interact.* **2000**, *13*, 1041–1052.
215. K. Inoue; Y. Ebizuka, *FEBS Lett.* **1996**, *378*, 157–160.
216. A. Kohara; C. Nakajima; K. Hashimoto; T. Ikenaga; H. Tanaka; Y. Shoyama; S. Yoshida; T. Muranaka, *Plant Mol. Biol.* **2005**, *57*, 225–239.
217. A. Kohara; C. Nakajima; S. Yoshida; T. Muranaka; *Phytochemistry* **2007**, *68*, 478–486.
218. R. Kuhn; I. Löw, *Angew. Chem. Int. Ed.* **1954**, *66*, 639–640.
219. S. Abouzid; N. Fawzy; N. Darweesh; Y. Orihara, *Nat. Prod. Res.* **2008**, *22*, 147–153.
220. S. Austin; E. Lojkowska; M. K. Ehlenfeldt; A. Kelman; J. P. Helgeson, *Phytopathology* **1988**, *78*, 1216–1220.
221. G. C. Percival; M. S. Karim; G. R. Dixon, *Plant Pathol.* **1998**, *47*, 665–670.
222. K. L. Deahl; W. W. Cantelo; S. L. Sinden; L. L. Sanford, *Am. Potato J.* **1991**, *68*, 659–666.
223. L. L. Sanford; K. L. Deahl; S. L. Sinden; T. L. Ladd, Jr, *Am. Potato J.* **1997**, *69*, 693–703.
224. V. Simons; J. P. Morrissey; M. Latijnhouwers; M. Csukai; A. Cleaver; C. Yarrow; A. Osbourn, *Antimicrob. Agents Chemother.* **2006**, *50*, 2732–2740.
225. E. A. J. Keukens; T. de Vrije; L. A. M. Jansen; H. de Boer; M. Janssen; A. I. P. M. de Kroon; W. M. F. Jongen; B. de Kruijff, *Biochim. Biophys. Acta* **1996**, *1279*, 243–250.
226. T. Mandimika; H. Baykus; J. Poortman; C. Garza; H. Kuiper; A. Peijnenburg, *Food Chem. Toxicol.* **2007**, *45*, 1918–1927.
227. R. Tschesche; H. Hulpke, *Z. Naturforsch.* **1967**, *22b*, 791.
228. L. Canonica; L. Ronchietti; G. Russo; G. Sportoletti, *Chem. Commun.* **1977**, *286*, 213.
229. E. Heftmann, *Phytochemistry* **1983**, *22*, 1843–1860.
230. C. Paczkowski; M. Kalinowska; Z. A. Wojciechowski, *Acta Biochim. Pol.* **1997**, *44*, 43–53.
231. K. F. McCue; P. V. Allen; L. V. Shepherd; A. Blake; J. Whitworth; M. M. Maccree; D. R. Rockhold; D. Stewart; H. V. Davies; W. R. Belknap, *Phytochemistry* **2006**, *15*, 1590–1597.
232. C. P. Moehs; P. V. Allen; M. Friedman; W. R. Belknap, *Plant J.* **1997**, *11*, 227–236.
233. K. F. McCue; P. V. Allen; L. V. Shepherd; A. Blake; M. M. Maccree; D. R. Rockhold; R. G. Novy; D. Stewart; H. V. Davies; W. R. Belknap, *Phytochemistry* **2007**, *68*, 327–334.
234. P. Krits; E. Fogelman; I. Ginzberg, *Planta* **2007**, *227*, 143–150.
235. L. Arnqvist; P. C. Dutta; L. Jonsson; F. Sitbon, *Plant Physiol.* **2003**, *131*, 1792–1799.
236. Y. Kamano; T. Nogawa; A. Yamashita; M. Hayashi; M. Inoue; P. Drasar; G. R. Pettit, *J. Nat. Prod.* **2002**, *65*, 1001–1005.
237. P. Stoilov; C. H. Lin; R. Damoiseaux; J. Nikolic; D. L. Black, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 11218–11223.
238. C. Theurer; W. Kreis; E. Reinhard, *Planta Med.* **1998**, *64*, 705–710.
239. F. Schaller; W. Kreis, *Planta Med.* **2006**, *72*, 1149–1156.
240. T. Warashina; T. Noro, *Chem. Pharm. Bull. (Tokyo)* **2000**, *48*, 516–524.
241. P. Lindemann; M. Luckner, *Phytochemistry* **1997**, *46*, 507–513.
242. H. U. Seitz; D. E. Gaertner, *Plant Cell* **1994**, *38*, 337–344.
243. U. Stuhlemmer; W. Kreis, *Plant Physiol.* **1996**, *34*, 85–91.

244. F. Milek; E. Reinhard; W. Kreis, *Plant Physiol. Biochem.* **1997**, *35*, 111–121.
245. U. Stuhlemmer; W. Haussmann; F. Milek; W. Kreis; E. Reinhard, *Z. Naturforsch.* **1993**, *48*, 713–721.
246. V. Herl; J. Frankenstein; N. Meitinger; F. Müller-Uri; W. Kreis, *Planta Med.* **2007**, *73*, 704–710.
247. A. Thorn; C. Egerer-Sieber; C. M. Jäger; V. Herl; F. Müller-Uri; W. Kreis; Y. A. Muller, *J. Biol. Chem.* **2008**, *283*, 17260–17269.
248. V. Herl; G. Fischer; R. Bötsch; F. Müller-Uri; W. Kreis, *Planta Med.* **2006**, *72*, 1163–1165.
249. V. Herl; G. Fischer; V. A. Reva; M. Stiebritz; Y. A. Muller; F. Müller-Uri; W. Kreis, *Biochimie* **2009**, *91*, 517–525.
250. J. H. Jun; C. M. Ha; H. G. Nam, *Plant Cell Physiol.* **2002**, *43*, 323–330.
251. U. Stuhlemmer; W. Kreis, *Tetrahedron Lett.* **1996**, *371*, 2221–2224.
252. J. M. Hagel; E. C. Yeung; P. J. Facchini, *Trends Plant Sci.* **2008**, *13*, 631–639.
253. A. A. Agrawal; M. Fishbein, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 10057–10060.
254. B. Mudd, Sterol Interconversions. In *The Biochemistry of Plants*; P. K. Stumpf, E. E. Conn, Eds.; Academic Press: New York, **1980**; *4*, 509–534.
255. B. Cheng; J. Kowal, *J. Lipid Res.* **1994**, *35*, 1115–1121.
256. J. P. Yuan; H. C. Kuang; J. H. Wang; X. Liu, *Appl. Microbiol. Biotechnol.* **2008**, *80*, 459–465.
257. L. Dyas; L. J. Goad, *Phytochemistry* **1993**, *34*, 17–29.
258. P. Oelkers; A. Behari; D. Cromley; J. T. Billheimer; S. L. Sturley, *J. Biol. Chem.* **1998**, *273*, 26765–26771.
259. A. Jonas, *Biochim. Biophys. Acta* **2000**, *1529*, 245–256.
260. H. Yang; M. Bard; D. A. Bruner; A. Gleeson; R. J. Deckelbaum; G. Aljinovic; T. M. Pohl; R. Rothstein; S. L. Sturley, *Science* **1996**, *272*, 1353–1356.
261. D. Zweytick; E. Leitner; S. D. Kohlwein; C. Yu; J. Rothblatt; G. Daum, *Eur. J. Biochem.* **2000**, *267*, 1075–1082.
262. R. Schneiter; G. Daum, *Methods Mol. Biol.* **2006**, *313*, 75–84.
263. R. Köffel; R. Tiwari; L. Falquet; R. Schneiter, *Mol. Cell Biol.* **2005**, *25*, 1655–1668.
264. R. Tiwari; R. Köffel; R. Schneiter, *EMBO J.* **2007**, *26*, 5109–5119.
265. A. Banas; A. S. Carlsson; B. Huang; M. Lenman; W. Banas; M. Lee; A. Noiriel; P. Benveniste; H. Schaller; P. Bouvier-Navé; S. Stymne, *J. Biol. Chem.* **2005**, *280*, 34626–34634.
266. Q. Chen; L. Steinhauer; J. Hammerlindl; W. Keller; J. Zou, *Plant Physiol.* **2007**, *145*, 974–984.
267. P. Maillot-Vernier; L. Gondet; H. Schaller; P. Benveniste; G. Bellierd, *Mol. Gen. Genet.* **1991**, *231*, 33–40.
268. L. Gondet; R. Bronner; P. Benveniste, *Plant Physiol.* **1994**, *105*, 509–518.
269. S. C. Wilkinson; R. Powls; L. J. Goad, *Phytochemistry* **1994**, *37*, 1031–1035.
270. I. Hernández-Pinzón; J. H. Ross; K. A. Barnes; A. P. Damant; D. Murphy, *Planta* **1999**, *208*, 588–598.
271. D. J. Murphy, *Protoplasma* **2006**, *228*, 31–39.
272. R. A. Moreau; V. Singh; A. Nuñez; K. B. Hicks, *Biochem. Soc. Trans.* **2000**, *28*, 803–806.
273. M. S. Islam; T. Murata; M. Fujisawa; R. Nagasaka; H. Ushio; A. M. Bari; M. Hori; H. Ozaki, *Br. J. Pharmacol.* **2008**, *154*, 812–824.
274. M. A. Hartmann-Bouillon; P. Benveniste, *Phytochemistry* **1978**, *17*, 1037–1042.
275. C. S. Freire; D. S. Coelho; N. M. Santos; A. J. Silvestre; C. Pascoal Neto, *Lipids* **2005**, *40*, 317–322.
276. D. C. Warnecke; E. Heinz, *Plant Physiol.* **1994**, *105*, 1067–1073.
277. D. C. Warnecke; M. Baltrusch; F. Buck; F. P. Wolter; E. Heinz, *Plant Mol. Biol.* **1997**, *35*, 597–603.
278. D. Warnecke; R. Erdmann; A. Fahl; B. Hube; F. Müller; T. Zank; U. Zähringer; E. Heinz, *J. Biol. Chem.* **1999**, *274*, 13048–13059.
279. I. Hillig; M. Leipelt; C. Ott; U. Zähringer; D. Warnecke; E. Heinz, *FEBS Lett.* **2003**, *553*, 365–369.
280. T. Y. Nazarko; J. C. Farré; A. S. Polupanov; A. A. Sibirny; S. Subramani, *Autophagy* **2007**, *3*, 263–265.
281. A. H. Lebrun; C. Wunder; J. Hildebrand; Y. Churin; U. Zähringer; B. Lindner; T. F. Meyer; E. Heinz; D. Warnecke, *J. Biol. Chem.* **2006**, *281*, 27765–27772.
282. C. Wunder; Y. Churin; F. Winau; D. Warnecke; M. Vieth; B. Lindner; U. Zähringer; H. J. Mollenkopf; E. Heinz; T. F. Meyer, *Nat. Med.* **2006**, *12*, 1030–1038.
283. D. V. Lynch; P. L. Steponkus, *Plant Physiol.* **1987**, *83*, 761–767.
284. M. S. Webb; T. C. Irving; P. L. Steponkus, *Biochim. Biophys. Acta* **1995**, *1239*, 226–238.
285. K. K. Halling; B. Ramstedt; J. P. Slotte, *Biochim. Biophys. Acta* **2008**, *1778*, 1100–1111.
286. L. Peng; Y. Kawagoe; P. Hogan; D. Delmer, *Science* **2002**, *295*, 59–60.
287. Y. K. Kim; Y. Wang; Z. M. Liu; P. E. Kolattukudy, *Plant J.* **2002**, *30*, 177–187.
288. A. Zdzislaw; A. Wojciechowski; J. Zimowski, *Biochim. Biophys. Acta* **1975**, *398*, 111–117.
289. A. Potocka; J. Zimowski, *Acta Biochim. Pol.* **2008**, *55*, 135–140.
290. N. Y. Yoon; B. S. Min; H. K. Lee; J. C. Park; J. S. Choi, *Arch. Pharm. Res.* **2005**, *28*, 892–896.
291. C. A. Strott, *Endocr. Rev.* **1997**, *17*, 670–697.
292. L. Varin; F. Marsolais; M. Richard; M. Rouleau, *FASEB J.* **1997**, *11*, 517–525.
293. M. Rouleau; F. Marsolais; M. Richard; L. Nicolle; B. Voigt; G. Adam; L. Varin, *J. Biol. Chem.* **1999**, *274*, 20925–20930.
294. F. Marsolais; J. Boyd; Y. Paredes; A. M. Schinas; M. Garcia; S. Elzein; L. Varin, *Planta* **2007**, *225*, 1233–1244.
295. P. S. Cheetham; P. Dunnill; M. D. Lilly, *Biochem. J.* **1982**, *201*, 515–521.
296. Y. R. Chiang; W. Ismail; S. Gallien; D. Heintz; A. Van Dorsselaer; G. Fuchs, *Appl. Environ. Microbiol.* **2008**, *74*, 107–113.
297. P. Plesiat; M. Grandguillot; S. Harayama; S. Vragar; Y. Michel-Briand, *J. Bacteriol.* **1991**, *173*, 7219–7227.
298. R. Van der Geize; K. Yam; T. Heuser; M. H. Wilbrink; H. Hara; M. C. Anderton; E. Sim; L. Dijkhuizen; J. E. Davies; W. W. Mohn; L. D. Eltis, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1947–1952.
299. N. Casali; L. W. Riley, *BMC Genom.* **2007**, *26*, 60.
300. W. W. Mohn; R. van der Geize; G. R. Stewart; S. Okamoto; J. Liu; L. Dijkhuizen; L. D. Eltis; *J. Biol. Chem.* **2008**, *283*, 35368–35374.
301. A. Baldán; D. D. Bojanic; P. A. Edwards, *J. Lipid Res.* **2008**, *50*, S80–S85.
302. K. E. Berge; H. Tian; G. A. Graf; L. Yu; N. V. Grishin; J. Schultz; P. Kwiterovich; B. Shan; R. Barnes; H. H. Hobbs, *Science* **2000**, *290*, 1771–1775.

303. A. M. Gallegos; B. P. Atshaves; S. M. Storey; O. Starodub; A. D. Petrescu; H. Huang; A. L. McIntosh; G. G. Martin; H. Chao; A. B. Kier; F. Schroeder, *Prog. Lipid Res.* **2001**, *40*, 498–563.
304. U. Seedorf; S. Scheek; T. Engel; C. Steif; H. J. Hinz; G. Assmann, *J. Biol. Chem.* **1994**, *269*, 2613–2618.
305. S. M. Colles; J. K. Woodford; D. Moncecchi; S. C. Myers-Payne; L. R. McLean; J. T. Billheimer; F. Schroeder, *Lipids* **1995**, *30*, 795–803.
306. N. J. Stolowich; A. Frolov; B. Atshaves; E. J. Murphy, C. A. Jolly; J. T. Billheimer; A. I. Scott; F. Schroeder, *Biochemistry* **1997**, *36*, 1719–1729.
307. T. Choinowski; H. Hauser; K. Piontek, *Biochemistry* **2000**, *39*, 1897–1902.
308. D. H. Dyer; S. Lovell; J. B. Thoden; H. M. Holden; I. Rayment; Q. Lan, *J. Biol. Chem.* **2003**, *278*, 39085–39091.
309. J. Edqvist; E. Rönnberg; S. Rosenquist; K. Blomqvist; L. Viitanen; T. A. Salminen; M. Nylund; J. Tuuf; P. Mattjus, *J. Biol. Chem.* **2004**, *279*, 44–53 *535*.
310. B. S. Zheng; E. Rönnberg; L. Viitanen; T. A. Salminen; K. Lundgren; T. Moritz; J. Edqvist, *J. Exp. Bot.* **2008**, *59*, 3485–3499.
311. J. Edqvist; K. Blomqvist, *J. Mol. Evol.* **2006**, *62*, 292–306.
312. G. D. Fairn; C. R. McMaster, *Cell Mol. Life Sci.* **2008**, *65*, 228–236.
313. N. B. Javitt, *Steroids* **2008**, *73*, 149–157.
314. A. L. Skirpan; P. E. Dowd; P. Sijacic; C. J. Jaworski; S. Gilroy; T. H. Kao, *Plant Mol. Biol.* **2006**, *61*, 553–565.
315. V. Mikes; M. L. Milat; M. Ponchet; P. Ricci; J. P. Blein, *FEBS Lett.* **1997**, *416*, 190–192.
316. J. Lochman; T. Kasparovsky; J. Damborsky; H. Osman; A. Marais; R. Chaloupkova; M. Ponchet; J. P. Blein; V. Mikes, *Biochemistry* **2005**, *44*, 6565–6572.
317. J. L. Goldstein; M. S. Brown, *Nature* **1990**, *343*, 425–430.
318. J. R. Smith; T. F. Osborne; M. S. Brown; J. L. Goldstein; G. Gil, *J. Biol. Chem.* **1988**, *263*, 18480–18487.
319. M. S. Brown; J. L. Goldstein, *Cell* **1997**, *89*, 331–340.
320. C. Yokoyama; X. Wang; M. R. Briggs; A. Admon; J. Wu; X. Hua; J. L. Goldstein; M. S. Brown, *Cell* **1993**, *75*, 187–197.
321. X. Wang; R. Sato; M. S. Brown; X. Hua; J. L. Goldstein, *Cell* **1994**, *77*, 53–62.
322. M. S. Brown; J. L. Goldstein, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11041–11048.
323. M. S. Brown; J. Ye; R. B. Rawson; J. L. Goldstein, *Cell* **2000**, *100*, 391–398.
324. L. Xu; R. D. Simoni, *Arch. Biochem. Biophys.* **2003**, *414*, 232–243.
325. T. Y. Chang; C. C. Chang; N. Ohgami; Y. Yamauchi, *Annu. Rev. Cell Dev. Biol.* **2006**, *22*, 129–157.
326. T. Yang; P. J. Espenshade; M. E. Wright; D. Yabe; Y. Gong; R. Aebersold; J. L. Goldstein; M. S. Brown, *Cell* **2002**, *110*, 489–500.
327. A. L. Hughes; B. L. Todd; P. J. Espenshade, *Cell* **2005**, *120*, 831–842.
328. Y. C. Chang; C. M. Bien; H. Lee; P. J. Espenshade; K. J. Kwon-Chung, *Mol. Microbiol.* **2007**, *64*, 614–629.
329. A. C. Seegmiller; I. Dobrosotskaya; J. L. Goldstein; Y. K. Ho; M. S. Brown; R. B. Rawson, *Dev. Cell* **2002**, *2*, 229–238.
330. J. A. Tschäpe; C. Hammerschmied; M. Mühlig-Versen; K. Athenstaedt; G. Daum; D. Kretzschmar, *EMBO J.* **2002**, *21*, 6367–6376.
331. M. Bard; J. F. Downing, *J. Gen. Microbiol.* **1981**, *125*, 415–420.
332. M. Servouse; F. Karst, *Biochem. J.* **1986**, *240*, 541–547.
333. T. Polakowski; U. Stahl; C. Lang, *Appl. Microbiol. Biotechnol.* **1998**, *49*, 66–71.
334. H. Schaller; B. Grausem; P. Benveniste; M. L. Chye; C. T. Tan; Y. H. Song; N. Chua, *Plant Physiol.* **1995**, *109*, 761–770.
335. R. Y. Hampton; J. Rine, *J. Cell Biol.* **1994**, *125*, 299–312.
336. D. Dimster-Denk; J. Rine, *Mol. Cell Biol.* **1996**, *16*, 3981–3989.
337. S. J. Smith; J. H. Crowley; L. W. Parks, *Mol. Cell Biol.* **1996**, *16*, 5427–5432.
338. F. Ness; S. Bourot; M. Régnacq; R. Spagnoli; T. Bergès; F. Karst, *Eur. J. Biochem.* **2001**, *268*, 1585–1595.
339. A. Vik; J. Rine, *Mol. Cell Biol.* **2001**, *21*, 6395–6405.
340. B. S. Davies; H. S. Wang; J. Rine, *Mol. Cell Biol.* **2005**, *25*, 7375–7385.
341. M. Germann; C. Gallo; T. Donahue; R. Shirzadi; J. Stukey; S. Lang; C. Ruckenstuhl; S. Oliaro-Bosso; V. McDonough; F. Turnowsky; G. Balliano; J. T. Nickels, Jr., *J. Biol. Chem.* **2005**, *280*, 35904–35913.
342. C. Hongay; M. Bard; F. Winston, *EMBO J.* **2002**, *21*, 4114–4124.
343. S. E. Saucier; A. A. Kandutsch; F. R. Taylor; T. A. Spencer; S. Phirwa; A. K. Gayen, *J. Biol. Chem.* **1985**, *260*, 14571–14579.
344. H. Yang, *Trends Cell Biol.* **2006**, *16*, 427–432.
345. Y. Wang; P. M. Rogers; K. R. Stayroock; C. Su; G. Varga; Q. Shen; S. Nagpal; T. P. Burris, *Mol. Pharmacol.* **2008**, *74*, 1716–1721.
346. G. Maier; G. Bing; U. Falken; E. Wagner; C. Unger, *Anticancer Res.* **1999**, *19*, 4251–4256.
347. M. Suzuki; Y. Kamide; N. Nagata; H. Seki; K. Ohyama; H. Kato; K. Masuda; S. Sato; T. Kato; S. Tabata; S. Yoshida; T. Muranaka, *Plant J.* **2004**, *37*, 750–761.
348. M. Rodríguez-Concepción; O. Forés; J. F. Martinez-García; V. González; M. A. Phillips; A. Ferrer; A. Boronat, *Plant Cell* **2004**, *16*, 144–156.
349. R. A. Demel; B. De Kruyff, *Biochim. Biophys. Acta* **1976**, *457*, 109–132.
350. D. A. Brown; J. K. Rose, *Cell* **1992**, *68*, 533–544.
351. X. Xu; R. Bittman; G. Duportail; D. Heissler; C. Vilcheze; E. London, *J. Biol. Chem.* **2001**, *276*, 33540–33546.
352. D. C. Hoessli; S. Ilangumaran; A. Soltermann; P. J. Robinson; B. Borisch; Nasir-Ud-Din, *Glycoconj. J.* **2000**, *17*, 191–197.
353. D. W. Zaas; M. Duncan; J. Rae Wright; S. N. Abraham, *Biochim. Biophys. Acta* **2005**, *1746*, 305–313.
354. S. Mongrand; J. Morel; J. Laroche; S. Claverol; J. P. Carde; M. A. Hartmann; M. Bonneu; F. Simon-Plas; R. Lessire; J. J. Bessoule, *J. Biol. Chem.* **2004**, *279*, 36277–36286.
355. Y. Roche; P. Gerbeau-Pissot; B. Buhot; D. Thomas; L. Bonneau; J. Gresti; S. Mongrand; J. M. Perrier-Cornet; F. Simon-Plas, *FASEB J.* **2008**, *22*, 3980–3991.
356. M. Bagnat; S. Keränen; A. Shevchenko; A. Shevchenko; K. Simons, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3254–3259.
357. M. Bagnat; K. Simons, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14183–14188.
358. J. G. Beck; D. Mathieu; C. Loudet; S. Buchoux; E. J. Dufourc, *FASEB J.* **2007**, *21*, 1714–1723.
359. C. M. Crowder; E. J. Westover; A. S. Kumar; R. E. Ostlund, Jr; D. F. Covey, *J. Biol. Chem.* **2001**, *276*, 44369–44372.
360. A. König; R. Happle; D. Bornholdt; H. Engel; K. H. Grzeschik, *Am. J. Med. Genet.* **2000**, *90*, 339–346.

361. H. Traupe; C. Has, *Eur. J. Dermatol.* **2000**, *10*, 425–428.
362. R. I. Kelley; L. E. Kratz; R. L. Glaser; M. L. Netzloff; L. M. Wolf; E. W. Jabs, *Am. J. Med. Genet.* **2002**, *110*, 95–102.
363. B. U. Fitzky; M. Witsch-Baumgartner; M. Erdel; J. N. Lee; Y. K. Paik; H. Glossmann; G. Utermann; F. F. Moebius, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8181–8186.
364. J. A. Porter; K. E. Young; P. A. Beachy, *Science* **1996**, *274*, 255–259. Erratum in: *Science* **1996**, *274*, 1597.
365. J. Alcedo; M. Noll, *Biol. Chem.* **1997**, *378*, 583–590.
366. T. R. Bürglin, *Genome Biol.* **2008**, *9*, 241.
367. P. E. Kuwabara; M. Labouesse, *Trends Genet.* **2002**, *18*, 193–201.
368. D. Rozman; M. Seliskar; M. Cotman; M. Fink; *Mol. Cell Endocrinol.* **2005**, *234*, 47–56.

Biographical Sketch

Hubert Schaller obtained the Ph.D. degree under the supervision of P. Benveniste in 1992 from the Université Pasteur in Strasbourg, France, on the '*In Vitro* Selection and Characterization of Tobacco Mutants Resistants to Pesticide Inhibitors of Sterol Biosynthesis'. He was a visiting scientist in the cooperative framework of the Phytochemical Survey of Malaysia at the Chemistry Department, University of Malaya, Kuala Lumpur, in collaboration with the Institut de Chimie des Substances Naturelles du CNRS in Gif-sur-Yvette, from 1989 to 1990. He was a postdoctoral scientist at the Max Planck Institut in Cologne from 1993 to 1994 in the Genetic Principles of Plant Development and Breeding Department. Hubert Schaller works in the field of metabolic plant biology with a focus on steroids and triterpenoids in the Plant Metabolic Networks Department of the Institut de Biologie Moléculaire des Plantes du CNRS, Strasbourg.

1.22 Isoprenoid in Actinomycetes

Tohru Dairi, Toyama Prefectural University, Toyama, Japan

© 2010 Elsevier Ltd. All rights reserved.

1.22.1	**Introduction**	789
1.22.2	**Cyclic Isoprenoids**	790
1.22.2.1	Cyclic Monoterpene	790
1.22.2.2	Cyclic Sesquiterpene	791
1.22.2.3	Cyclic Diterpene	796
1.22.2.4	Other Cyclic Isoprenoids	799
1.22.3	**Hybrid Isoprenoids**	801
1.22.3.1	Novobiocin/Clorobiocin	802
1.22.3.2	Polyketide-Isoprenoid Hybrid Compound	802
1.22.3.3	Moenomycin A/Pholipomycin/AC326-Alpha	805
1.22.3.4	Phenazine-Isoprenoid Hybrid Compounds	806
1.22.3.5	Carbazole-Isoprenoid Hybrid Compounds	807
1.22.3.6	Other Isoprenoid Hybrid Compounds	807
References		811

1.22.1 Introduction

Isoprenoids form the largest single family of compounds found in nature with over 24 000 known examples and contain industrially useful compounds such as rubber, flavors, antibiotics, and plant hormones. Most of these compounds have been isolated from eukaryotes such as plants and fungi. In contrast to eukaryotes, prokaryotes are known to produce a limited number of isoprenoids. Among prokaryotes, actinomycete strains, which are known to produce many kinds of nonribosomal peptides, polyketides, aminoglycosides, and so on, have recently been reported to produce many types of isoprenoids. The structures of these isoprenoids are unique and different from those of eukaryotic origin. Therefore, in this chapter, the structural diversity of isoprenoids produced by actinomycetes is summarized.

Besides the structural diversity, biosynthetic studies on genes and enzymes responsible for biosynthesis of isoprenoids produced by actinomycetes are also described in this chapter. Since recombinant DNA technology has been used to understand the biosynthetic machinery of natural compounds in the past two decades, gene clusters encoding many natural products have been cloned from many organisms and their functions are extensively characterized. Moreover, whole-genome sequencing has uncovered hundreds of candidates for secondary metabolic pathways. Among them, the biosynthetic machinery of nonribosomal peptide and polyketide natural products has been extensively investigated. On the other hand, only limited studies on biosynthetic genes and enzymes of isoprenoids have been reported, probably because most of the isoprenoids were isolated from eukaryotes and biosynthetic enzymes of isoprenoids are usually membrane bound, resulting in technical difficulties while purifying such enzymes from natural resources and while expressing recombinant enzymes in *Escherichia coli*. Moreover, scattering of biosynthetic genes on genomic DNA in eukaryotes makes it difficult to clone all the biosynthetic genes. Recently, whole-genome sequencing of plants and fungi has been reported, but the number of organisms analyzed is limited, because of their large genome size. In contrast, a pioneering study on pentalenene synthase, a sesquiterpene synthase from an actinomycete strain, demonstrated that the enzyme is not membrane bound and is expressed with relative ease as a recombinant enzyme in *E. coli*.[1] This enabled us to investigate the properties of prokaryotic isoprenoid biosynthetic enzymes in detail. To date, many studies in this line such as cloning of isoprenoid biosynthetic genes, identification of isoprenoid-related genes in genome databases, and characterization of these genes with recombinant enzymes have been reported. Therefore, these biosynthetic studies are also summarized in this chapter.

1.22.2 Cyclic Isoprenoids

The first step in isoprenoid biosynthesis is catalyzed by polyprenyl diphosphate synthase (PDS), which sequentially condenses isopentenyl diphosphates (IPP) into allylic diphosphates, such as dimethylallyl diphosphate (DMAPP), to produce linear polyprenyl diphosphates with specific chain lengths. Geranyl diphosphate (GDP) synthase, farnesyl diphosphate (FDP) synthase, and geranylgeranyl diphosphate (GGDP) synthase catalyze the additions of one, two, and three molecules of IPP to DMAPP, yielding GDP (C_{10}), FDP (C_{15}), and GGDP (C_{20}), respectively. In plants and fungi, these polyprenyl diphosphates usually undergo a range of cyclizations to produce the cyclic parent skeletons of monoterpenes (C_{10}), sesquiterpenes (C_{15}), and diterpenes (C_{20}), followed by a variety of modifications, such as hydroxylation, methylation, and glycosylation, to give rise to thousands of different isoprenoid metabolites.

Isoprenoid cyclases are classified into two major types with respect to their modes of cyclization. One type of the reaction is initiated by ionization of polyprenyl diphosphate to an allylic carbocation followed by cyclization and deprotonation to olefin (type A). Almost all monoterpene and sesquiterpene cyclases are classified in this group. The other type of the reaction is initiated by protonation at the terminal double bond of polyprenyl diphosphate (type B). Squalene (hopene) cyclases, which are triterpene cyclases, belong to this group. As for diterpene cyclases, both types are known. The former class and the latter class of enzymes usually possess a DDXXD motif and a DXDD motif, respectively, which mediate substrate binding by chelation of a divalent metal ion.

1.22.2.1 Cyclic Monoterpene

In contrast to plants and fungi, limited examples of cyclic isoprenoids that are formed by cyclization of GDP, FDP, or GGDP have been reported from actinomycete origin. To date, 2-methylisoborneol[2] (**1**) and 2-methylenebornane (**2**),[3] both of which are known as odorous and volatile metabolites, are the sole cyclic monoterpenes found in actinomycetes. A monoterpene cyclase responsible for the formation of **1** from GDP has recently been identified in *Streptomyces lasaliensis*.[3] Very interestingly, in this reaction, *S*-adenosylmethionine (SAM)-dependent methyltransferase first catalyzes methylation at C2 position of GDP and the methylated GDP is successively cyclized by the monoterpene cyclase to form **1** or **2** (**Figure 1**).

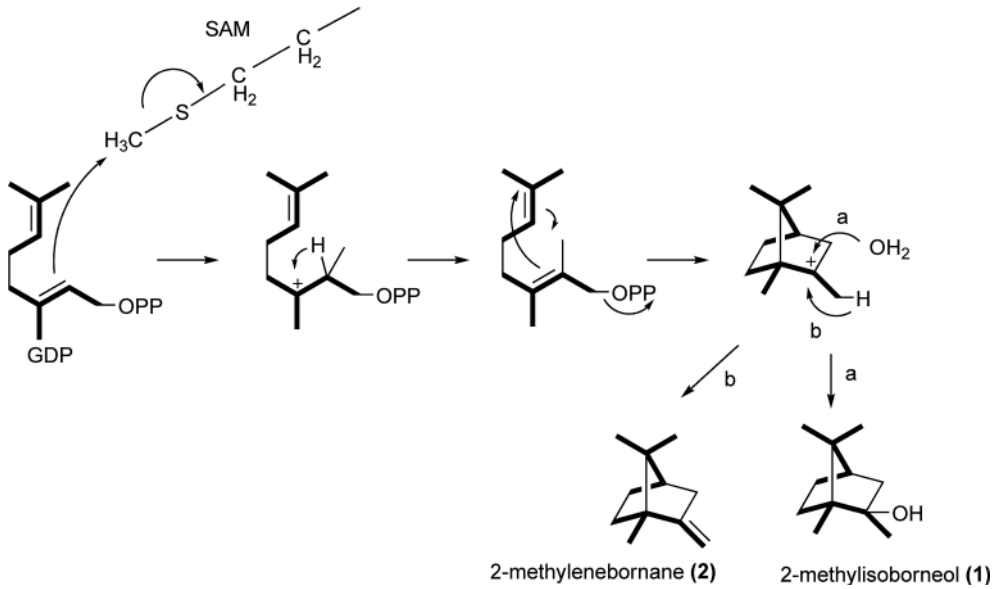


Figure 1 Cyclization of GDP to 2-methylisoborneol **(1)** and 2-methylenebornane **(2)**.

1.22.2.2 Cyclic Sesquiterpene

As for cyclic sesquiterpenes, pentalenene (**3**),[4] pentalenolactone (**4**),[5,6] (4*S*,7*R*)-germacra-1(10)*E*,5*E*-diene-11-ol (**5**),[7] geosmin (**6**),[8] epi-isozizaene (**7**),[9] albaflavenone (**8**),[10] amorphane sesquiterpenes (**9**),[11] selina-4(14),7(11)-diene-9-ol (**10**),[12] selina-4(14),7(11)-diene-8,9-diol (**10**),[13] africantriol (**11**),[14] and BE-40644 (**12**)[15] were isolated from actinomycetes. All these compounds except for **12** are biosynthesized from FDP by specific type A isoprenoid cyclases. Compound **12** is composed of a cyclic C_{15} unit derived from FDP and an aromatic moiety formed by successive dehydrations of 2-*epi*-5-*epi*-valiolone (see below). As described above, the pioneering biosynthetic studies on prokaryotic isoprenoid cyclases were performed by D. Cane *et al.*[1] with pentalenene synthase found in *Streptomyces* strain. After the first isolation of **4** in the 1960s, **4**-related compounds were extensively isolated and characterized in the 1980s to elucidate the biosynthetic mechanism. Later, **3**, a hydrocarbon and a probable precursor of **4**,[4] was isolated. In the mid-1980s, D. Cane *et al.* succeeded in detecting an enzyme activity that catalyzes the conversion of FDP into **3** in the cell-free extract of *Streptomyces* UC5319 and the detailed cyclization mechanism was studied with labeled precursors. In 1994, the enzyme (pentalenene synthase) was purified and a DNA fragment encoding it was cloned by polymerase chain reaction (PCR) using primers based on N-terminal and internal peptide sequences of a prokaryotic isoprenoid cyclase for the first time.[1] A recombinant enzyme was also successfully overproduced in *E. coli* and kinetic parameters of the enzyme were calculated. Active site residues of the enzyme were analyzed by site-directed mutagenesis.[16,17] However, the relationship between **3** and **4** remained unclear after that. In 2003, whole-genome sequencing of *Streptomyces avermitilis* was completed and a putative gene cluster (*ptl* genes) for biosynthesis of **4** was identified.[18] To date, of these genes, the function of *ptl*A (pentalenene synthase), *ptl*I (cytochrome P-450),[19] *ptl*H (nonheme iron dioxygenase),[20] and *ptl*F (short-chain dehydrogenase)[21] was studied with recombinant enzymes (**Figure 2**).

Geosmin (**6**), a sesquiterpenoid, is a well-known compound that is responsible for the undesirable 'earthy' flavor of red table beets. It was recently shown that the recombinant enzymes SCO6073[22–24] and SAV2163,[25] which existed in *Streptomyces coelicolor* A3(2) and *S. avermitilis*, respectively, and had significant sequence

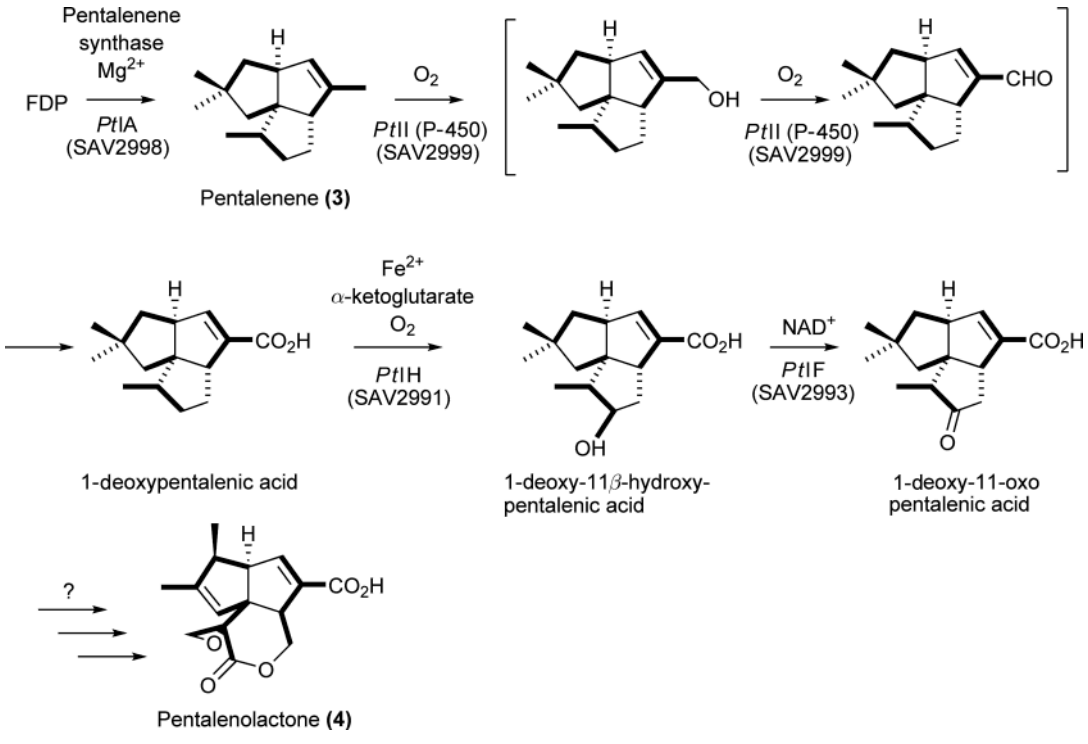


Figure 2 Biosynthesis of pentalenolactone **(4)** via pentalenene **(3)**.

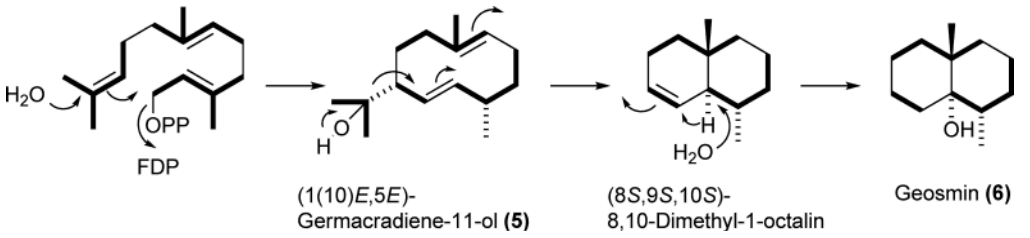


Figure 3 Cyclization of FDP to geosmin **(6)** via (1(10)*E*,5*E*)-germacradiene-11-ol **(5)**.

similarity to the abovementioned sesquiterpene synthase pentalenene synthase, converted FDP into geosmin in the presence of Mg^{2+} though the yield was very low and the major product was (4*S*,7*R*)-germacra-1(10)*E*,5*E*-diene-11-ol (**5**), which had been isolated as a natural compound in the culture broth of *Streptomyces* strain[7] (**Figure 3**). Considering the structure of **6**, C_3 unit was perhaps removed from a biosynthetic intermediate. This unique reaction mechanism has recently been clarified as follows: the N-terminal half of the recombinant SCO6073 converts FDP into **5** and germacrene D in the presence of Mg^{2+} and the C-terminal domain catalyzes the Mg^{2+}-dependent conversion of **5** to **6**.[26–28]

Albaflavenone (**8**), a sesquiterpene ketone, was isolated from highly odorous *Streptomyces albidoflavus*.[10] Previously, **8**-related compound **7** was shown to be formed from FDP by SCO5222 by successive reactions after ionization and isomerization of FDP.[9] Very recently, **8** was confirmed to be synthesized from (−)-*epi*-isozizaene (**7**) via (4*S*)- or (4*R*)-albaflavenol by SCO5223 (*CYP170A1*), a P-450 enzyme, of *S. coelicolor* A3(2) with a recombinant enzyme[29] (**Figure 4**). The detailed reaction mechanism of cyclization was also studied with deuterated FDPs.[9]

As for **9** and **10**, no biosynthetic studies were reported. However, these compounds were perhaps biosynthesized by sesquiterpene cyclases from FDP and some monooxygenases such as P-450 (**Figure 5**).

In contrast to the compounds described above, **12** is a 'cyclic isoprenoid quinone hybrid compound'. Interestingly, *Actinoplanes* sp. strain A40644, which is a producer of **12**, is equipped with the mevalonate (MV) pathway for the formation of IPP in addition to the 2-*C*-methyl-D-erythritol-4-phosphate (MEP) pathway, which usually operates in actinomycetes, and the strain was proved to produce **12** through the MV

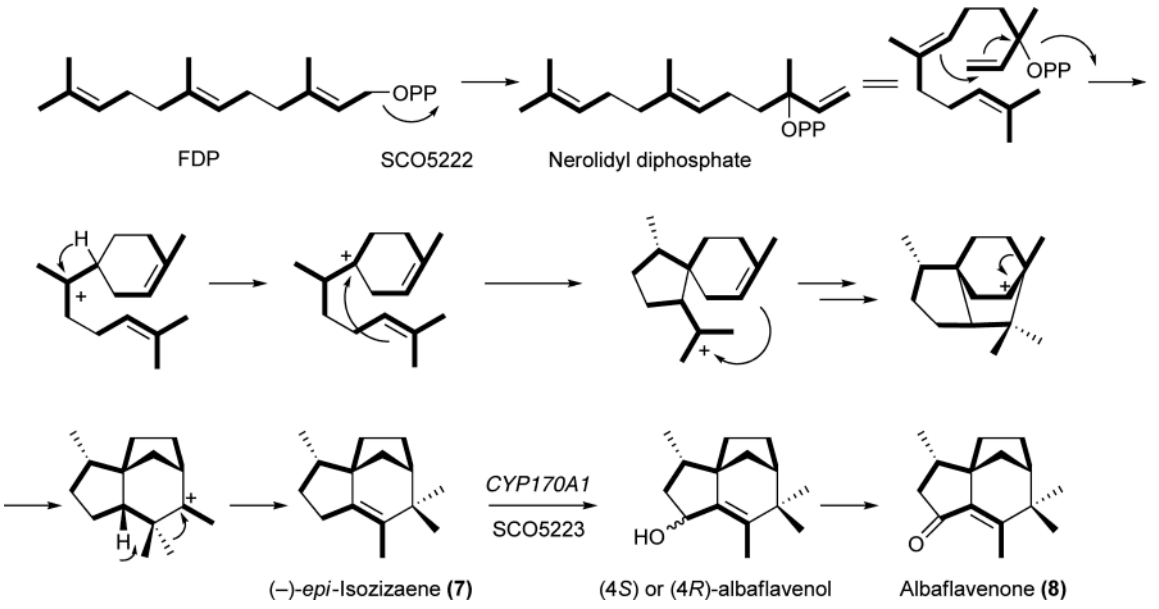


Figure 4 Biosynthesis of albaflavenone **(8)** via (-)-*epi*-isozizaene (**7**).

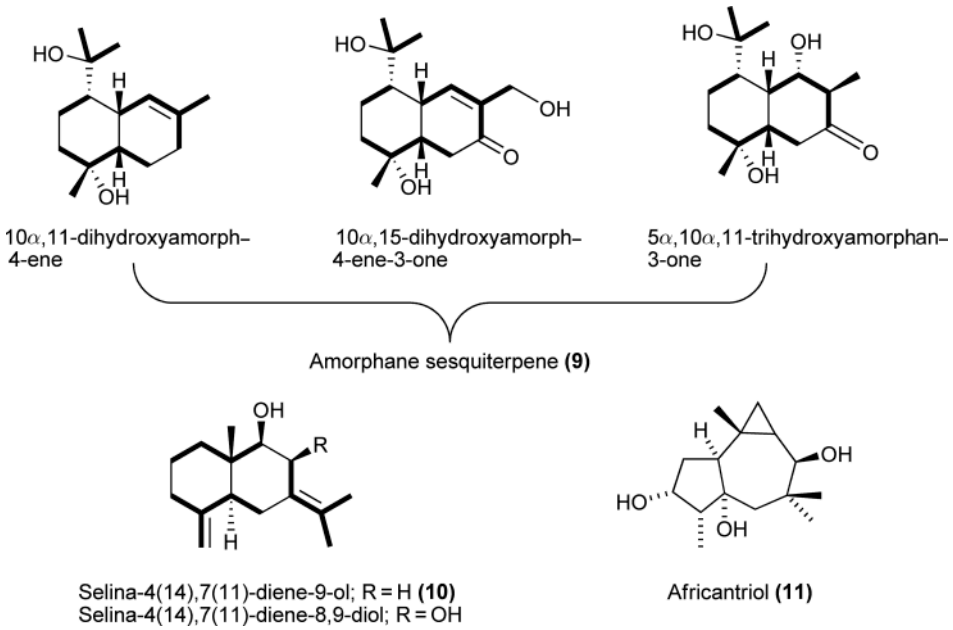
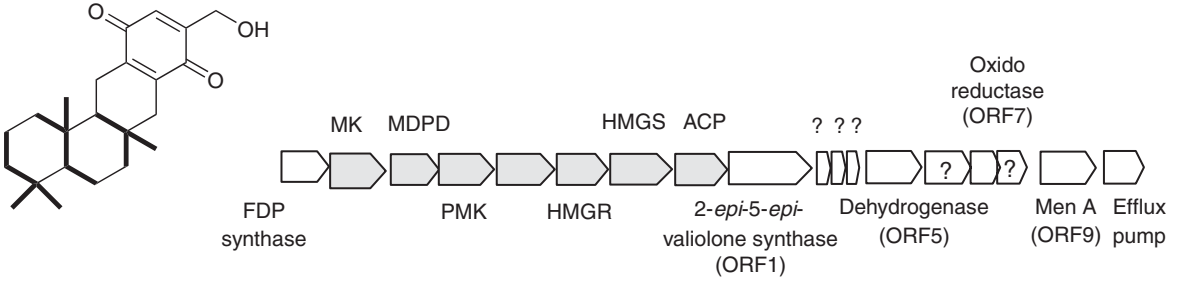


Figure 5 Structures of other cyclized sesquiterpenes.

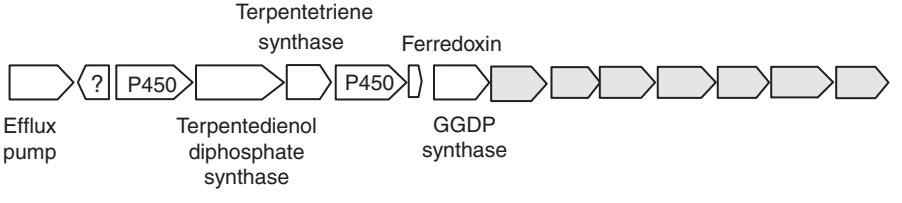
pathway.[30] In order to better understand the biological significance of the presence of the MV pathway in actinomycetes, the MV pathway gene clusters and their flanking regions have been cloned and analyzed from actinomycetes that produce isoprenoid compounds.[31] All the MV pathway gene clusters contained genes coding for mevalonate kinase (MK), mevalonate diphosphate decarboxylase (MDPD), phosphomevalonate kinase (PMK), type 2 IPP isomerase, HMG-CoA reductase (HMGR), HMG-CoA synthase (HMGS), and 3-oxoacyl-(acyl-carrier-protein) (ACP). The order of each of the open reading frames (ORFs) is also the same, and the respective homologous ORFs show more than 70% amino acid identity with each other. In contrast to these conservative gene organizations, very interestingly, the biosynthetic genes of **12**,[31] terpentecin (cyclic diterpene (**13**)),[32] furaquinocin A (polyketide-isoprenoid hybrid compound (**14**)),[33] viguiepinol (3-hydroxy-pimara-9(11),15-diene) (**16**),[34] and napyradiomycin A (**29**)[35] were located immediately upstream and/or downstream of the MV pathway gene cluster (**Figures 6** and **14**). These facts suggested that all the actinomycete strains possessing both the MV and MEP pathways produce isoprenoid compounds and the biosynthetic genes of one of these isoprenoids usually exist adjacent to the MV pathway gene cluster.

The biosynthetic gene cluster of **12** located immediately downstream of the MV pathway gene cluster was heterologously expressed in *Streptomyces lividans* TK23 and the production of **12** was confirmed.[31] This cluster contained 10 putative ORFs and the functions of some of them were estimated by comparing their deduced amino acid sequences with their sequence homologues. The ORF1 product presents a significant similarity to the 2-*epi*-5-*epi*-valiolone synthase (50% identity) and it was confirmed to have an estimated enzyme activity with a recombinant enzyme.[36] ORF5 and ORF7 have a similarity to dehydrogenases and oxidoreductases. Taken together, these results indicate that the quinone moiety of **12** can be formed by successive dehydrations of 2-*epi*-5-*epi*-valiolone. The ORF9 product has 30% identity with MenA, 1,4-dihydroxy-2-naphthoate octaprenyltransferases which participates in the biosynthesis of menaquinones and catalyzes the transfer of polyprenyl diphosphate to 1,4-dihydroxy-2-naphthoate. Considering that **12** has a structure composed of the quinone moiety and the C_{15} isoprene unit, ORF9 is strongly suggested to catalyze the transfer reaction of FDP to the quinone moiety (**Figure 7**). This type of prenyltransferases do not show any similarities to CloQ-like enzymes, which also catalyze prenyl transfer reactions (see below). Curiously, any sesquiterpene cyclases that have similarities to other isoprenoid cyclases such as the abovementioned pentalenene synthase and the SCO6073 product did not exist in the cluster, suggesting that other hypothetical ORFs included in the cluster might catalyze the cyclization of C_{15} unit.

(a) BE-40644 **(12)** biosynthetic gene cluster



(b) Terpentecin **(13)** biosynthetic gene cluster



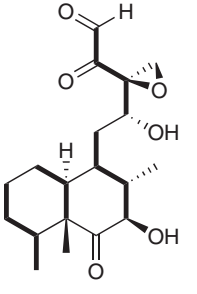


Figure 6 (Continued)

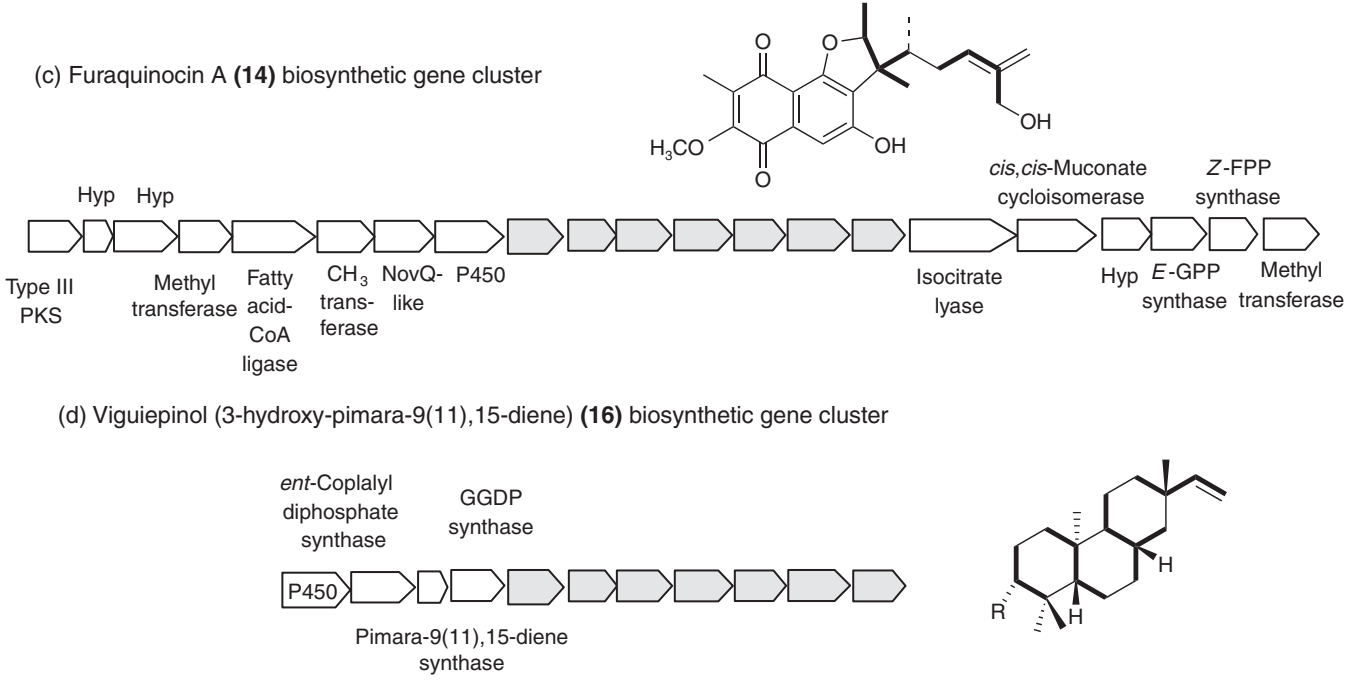


Figure 6 Biosynthetic gene clusters of BE-40644 **(12)**, terpentecin **(13)**, furaquinocin **(14)**, and viguiepinol **(16)**.

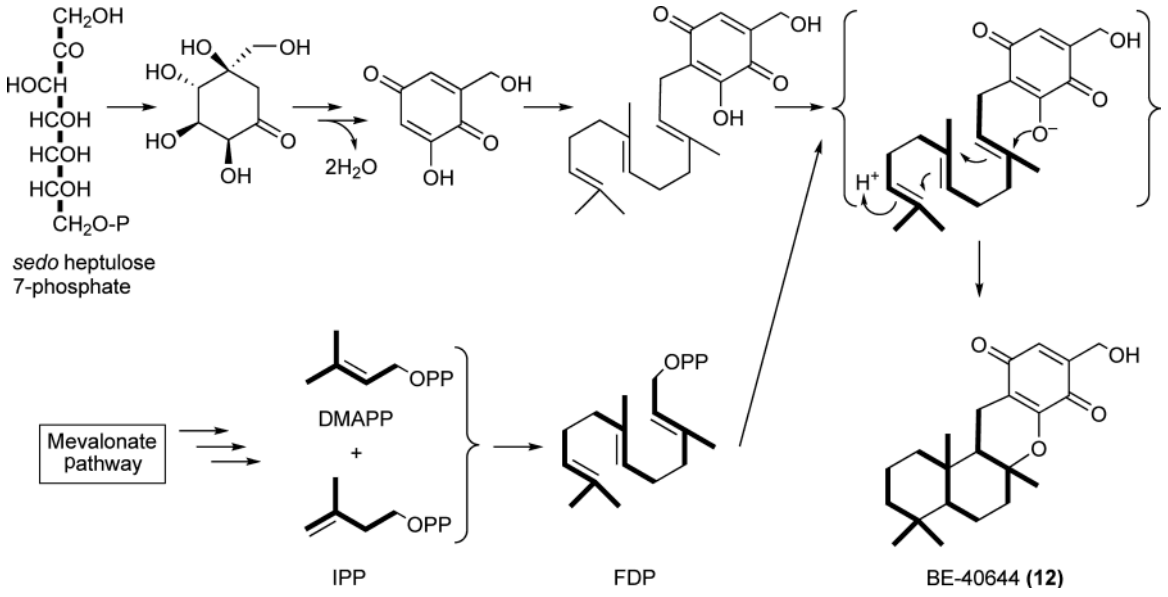


Figure 7 Proposed biosynthetic pathway of BE-40644 **(12)**.

1.22.2.3 Cyclic Diterpene

Among cyclic diterpenes so far identified from actinomycetes, the basic skeletons of **13**,[37,38] UCT4B (**15**, a derivative of **13**),[39] **16**,[33,34] oxaloterpins (**17**, a derivative of **16**),[40] tuberculosinol ((+)-5(6),13-halimadiene-15-ol) (**18**),[41] phenalinolactone (**19**),[42] brasilicardin A (**20**),[43,44] cyslabdan (**21**),[45] and spirocardin A (**22**)[46] are perhaps biosynthesized by type B diterpene cyclases. A sole example of diterpene that would be biosynthesized by type A cyclase is cyclooctatin (**23**)[47] (**Figure 8**).

Biosynthetic gene clusters of **13**[32,48–50] and **16**[33,34] were cloned taking advantage that isoprenoid biosynthetic genes usually exist adjacent to the MV pathway gene cluster in actinomycete strains that possess both the MV and MEP pathways (**Figure 6**). In the former cluster, two diterpene cyclase genes were identified by gene disruption and heterologous expression experiments as the first example of eubacterial diterpene cyclases.[33] One cyclase (terpentedienol diphosphate synthase, type B cyclase), which has a significant sequence similarity with the N-terminal half of diterpene cyclases from plants and fungi (less than 30% identity), was shown to convert GGDP into terpentedienol diphosphate (TDP) with a recombinant enzyme.[51,52] The other cyclase, which shares 25% identical sequences over 331 amino acids of the pentalenene synthase from *Streptomyces* sp. UC5319, catalyzed the conversion of TDP into *ent*-clerod-3,13(16),14-triene (terpentetriene, TTE) (**Figure 9**).[51,52] Therefore, the enzyme initiated the reaction by an ionization of TDP to an allylic carbocation followed by deprotonation to an olefin without cyclization reaction. Interestingly, the enzyme catalyzed the conversion of GGDP into 1,3(20),6,10,14-phytapentaene, α-springene, and 3,7,11,15-tetramethylhexadeca-1,3,6,10,14-pentaene (*E,E,E*). Furthermore, FDP was also converted to 7,11-dimethyl-3-methylenedodeca-1,6,10-triene (*E*), 7,11-trimethyldodeca-1,3,6,10-tetraene (*Z,E*), and 3,7,11-trimethyldodeca-1,3,6,10-tetraene (*E,E*) (**Figure 9**).[51] Enzymatic properties of both enzymes were investigated with the recombinant enzymes and it was found that the enzymes have almost the same enzymatic properties as those of eukaryotic diterpene cyclases in the following ways: requirement of Mg^{2+}; the order of catalytic parameters (K_m and V_{max} values); substrate inhibition (more than 50 μmol l^{-1} of GGDP), and so on.[51]

Compound **17**,[40] a derivative of **16**, was isolated from the culture broth of *Streptomyces* sp. KO-3988, which also produces **14**. The gene cluster for the biosynthesis of **16** was identified immediately upstream of the MV pathway gene cluster though genes encoding for enzymes that modify the 3-position of **16** to yield **17** did not exist in the flanking regions.[33,34] Curiously, another MV pathway gene cluster was also identified in the biosynthetic gene cluster of **14** in the strain KO-3988. This strain, therefore, contained two sets of MV pathway

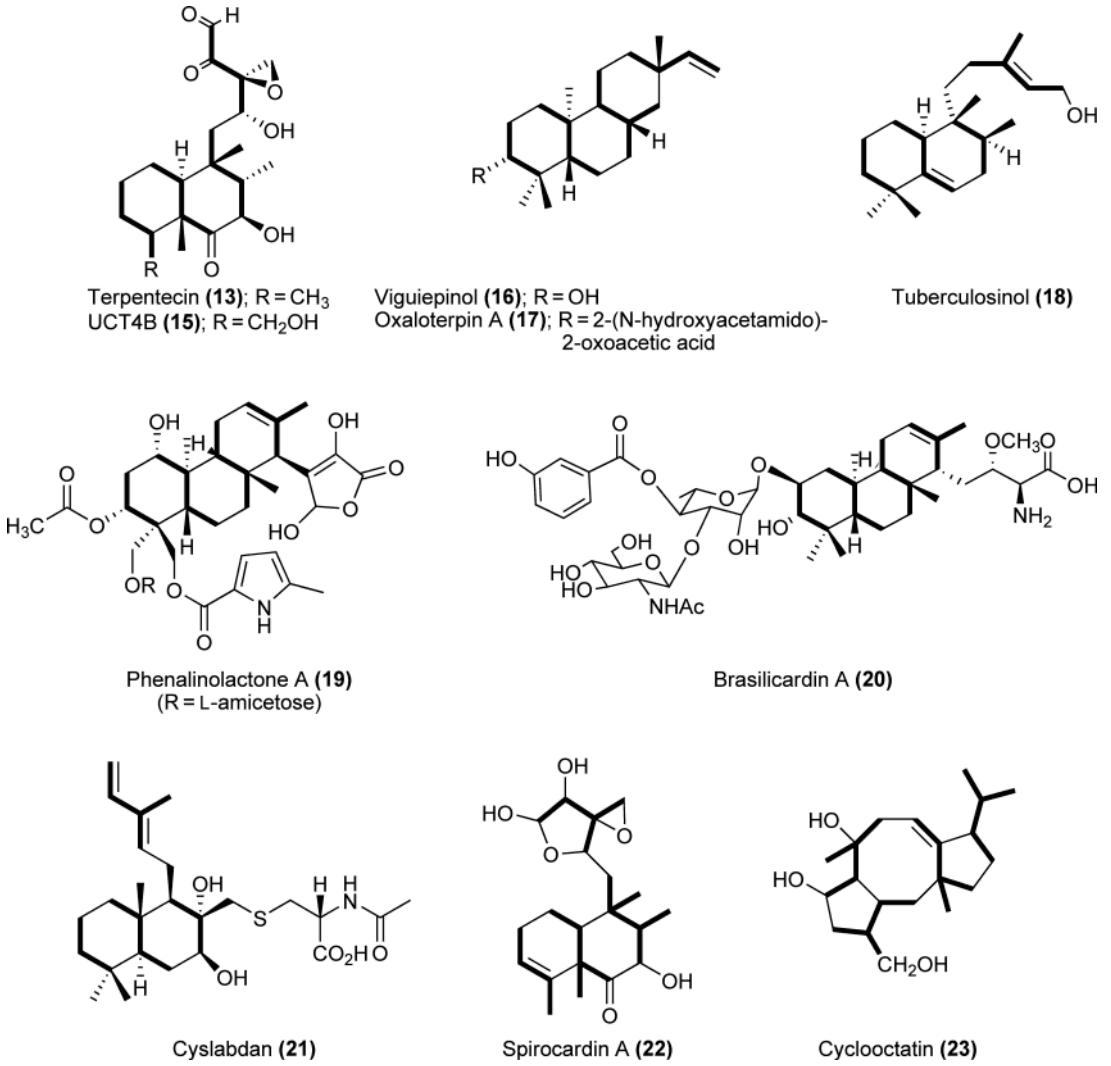


Figure 8 Structures of cyclized diterpenes.

gene clusters (**Figure 6**). In the biosynthetic gene cluster of **16**, a gene homologous to TDP synthase and a gene that showed no significant similarity to any other protein were involved and were confirmed to be *ent*-coplalyl diphosphate (CDP) synthase and pimara-9(11),15-diene (PMD) synthase, respectively, with recombinant enzymes as the first examples of enzymes with these biosynthetic functions from prokaryotes.[53] To date, several eukaryotic *ent*-CDP synthase genes, which are known to participate in gibberellin biosynthesis, have been cloned and characterized. Although the prokaryotic *ent*-CDP synthase had only a low amino acid similarity to these eukaryotic *ent*-CDP synthases (approximately 27% identity), some enzymatic properties such as the K_m and k_{cat} values, suitable concentration and kind of divalent cations, and concentration of substrate inhibition were almost the same as those of eukaryotic enzymes.[53]

To date, several genes encoding type A diterpene cyclases that accepted *ent*- or *syn*-CDP as a substrate have been cloned from plants and fungi. Of these, plant enzymes such as *ent*-cassa-12,15-diene synthase,[54] *syn*-pimara-7,15-diene synthase,[55] *ent*-sandaracopimaradiene synthase,[56] and stemar-13-ene synthase[57] consisted of more than 750 amino acid residues and their primary sequences had more than 40% homology with each other. In contrast, the PMD synthase was composed of only 295 amino acids and had no homology with any other proteins. These findings suggest that eubacterial and eukaryotic diterpene cyclase genes have independently evolved from each other.

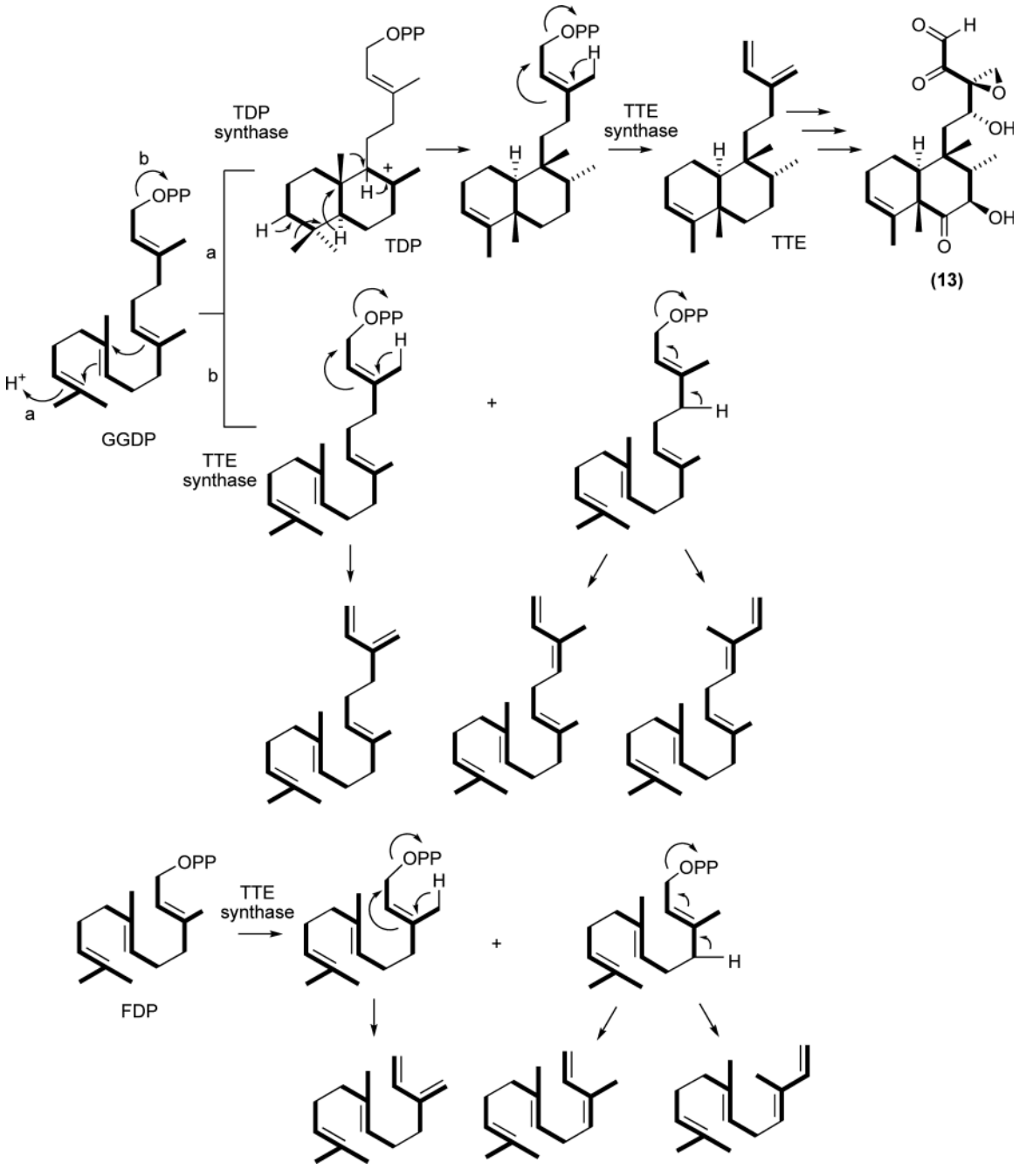


Figure 9 Summary of reactions catalyzed by diterpene cyclases responsible for the biosynthesis of terpentecin **(13)**.

A gene homologous to (oxido)squalene cyclases and some terpene cyclases such as TDP synthase was identified in *Mycobacterium tuberculosis* H37Rv by a bioinformatic screening, and whole-genome sequencing of this was done. The gene product, Rv3377c, was cloned and expressed in *E. coli*. This gene product was shown to convert GGDP into tuberculosinol ((+)-5(6),13-halimadiene-15-ol) (**18**) diphosphate, followed by elimination of diphosphate group by a contaminated phosphatase of *E. coli*.[41]

Compounds **19** and **20** are the so-called 'hybrid isoprenoids'. Compound **19** has a cyclic diterpene skeleton, to which L-amicetose, 5-methylpyrrole-2-carboxylic acid, and γ-butyrolactone moieties are

attached.[42] Compound **20**, produced by *Nocardia brasiliensis* IFM 0406 (currently referred to as *Nocardia terpenica*), has a structure consisting of a cyclic diterpene skeleton with L-rhamnose, *N*-acetylglucosamine, amino acid, and 3-hydroxybenzoate moieties.[43,44] Since producers of **19** and **20** have only the MEP pathway, biosynthetic gene clusters of the former[42] and the latter[58] were cloned with an L-amicetose biosynthetic gene and a GGDP synthase gene as probes. In both cases, gene disruption experiments were employed to confirm that these gene clusters were indeed responsible for biosyntheses of **19** and **20**.[42,58]

In the biosynthetic gene cluster of **19**, the *plaT2* gene, which has a significant similarity to isoprenoid cyclases, is confirmed to be essential for the biosynthesis of **19** by a gene disruption experiment.[42] Moreover, the *plaT2* product was suggested to use epoxy-GGPP as a substrate based on the following observations: The *plaT1* gene, which is located immediately downstream of the *plaT2* gene, had a significant similarity to eukaryotic squalene epoxidase genes and the *plaT2* product lacked a DXDD motif, in a manner similar to that of oxidosqualene cyclases that use epoxy-squalene as a substrate. The *plaT3* product has a significant similarity to prenyltransferases responsible for biosynthesis of ubiquinone (UbiA) but not CloQ-like prenyltransferases. Therefore, the enzyme is perhaps thought to catalyze the attachment of phosphoenolpyruvate to a cyclized GGDP intermediate in a manner similar to ORF9 in **12** biosynthesis (**Figure 6(a)**). Based on these observations and structures of intermediates accumulated in the culture broth of the disruptants, a biosynthetic pathway of **19** was proposed (**Figure 10**).

The biosynthetic gene cluster of **20** contained 11 putative genes though genes related to L-rhamnose and *N*-acetylglucosamine biosyntheses did not exist in the flanking regions.[58] Among them, the products of *bra3*, *bra4*, and *bra5* had high amino acid identities to those of *plaT3* (47% identity), *plaT2* (43% identity), and *plaT1* (48% identity), respectively. Since the early step of the proposed biosynthetic pathway of **20** was thought to be identical of that of **19** (**Figure 10**), high identities of amino acid sequences between the Bra3–5 and the PlaT3–1 are reasonable. By ClustalW analysis, the Bra4 was also shown to have no typical DXDD motif, in contrast to TDP synthase,[32,52] tuberculosinol diphosphate synthase,[42] and *ent*-CDP synthase.[33,53] The *bra4* gene product was expressed as a recombinant enzyme and used in an *in vitro* assay using GGPP as a substrate. However, no reaction products were formed under various conditions. Taken together, these results indicate that Bra4 and PlaT2 might use epoxy-GGPP as a substrate.

Interestingly, two MEP pathway genes, 1-deoxy-D-xylulose-5-phosphate synthase gene (*dxs*) and 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (*isp*G/*gcpE*) gene, were involved in the biosynthetic gene cluster of **19**[42] though no MEP genes were involved in the biosynthetic gene cluster of **20**.[58] The producer of **13**, which possesses both the MV and the MEP pathway, 60% of the IPP needed for the biosynthesis of **13** was shown to be supplied via the MV pathway and the MEP pathway contributed 40%.[59] Considering that 1-deoxy-D-xylulose-5-phosphate synthase was previously shown to be a rate-limiting enzyme in the MEP pathway,[60] the existence of additional MEP genes might be reasonable for full production of isoprenoids in producers possessing only the MEP pathway. Indeed, both *S. coelicolor* A3(2) and *S. avermitilis* were found to have two 1-deoxy-D-xylulose-5-phosphate synthase genes, SCO6768 and SCO6013 in the former strain and SAV1646 and SAV2244 in the latter strain. Two distinct 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase genes also exist in the former (SCO5696 and SCO6767) and the latter strain (SAV1647 and SAV2561). One set of the genes (SCO6768/6767 and SAV1646/1647) constitutes a cluster with the genes responsible for biosynthesis of squalene-hopene and phytoene (SCO6759–SCO6768 and SAV1646–SAV1654).

1.22.2.4 Other Cyclic Isoprenoids

To date, cyclic sesterterpenes and cyclic triterpenes, which are formed by geranylfarnesyl diphosphate and farnesylfarnesyl diphosphate, respectively, were not isolated from actinomycetes. However, many actinomycetes are known to produce squalene/hopanoid triterpenes. The first biosynthetic step of these compounds is initiated by a tail-to-tail condensation of FDP to form squalene by squalene synthases. In actinomycetes whose genome sequences have been analyzed, the genes encoding squalene-hopene cyclase were identified, such as SCO6764 in *S. coelicolor* and SAV1650 in *S. avermitilis*.

The most common cyclic tetraterpenes are carotenoid-related compounds. Carotenoid biosynthesis is initiated by phytoene synthase, which catalyzes a tail-to-tail condensation of GGDP to form phytoene probably via lycopersene. After successive desaturation reactions, phytoene is converted into lycopene. In the last step,

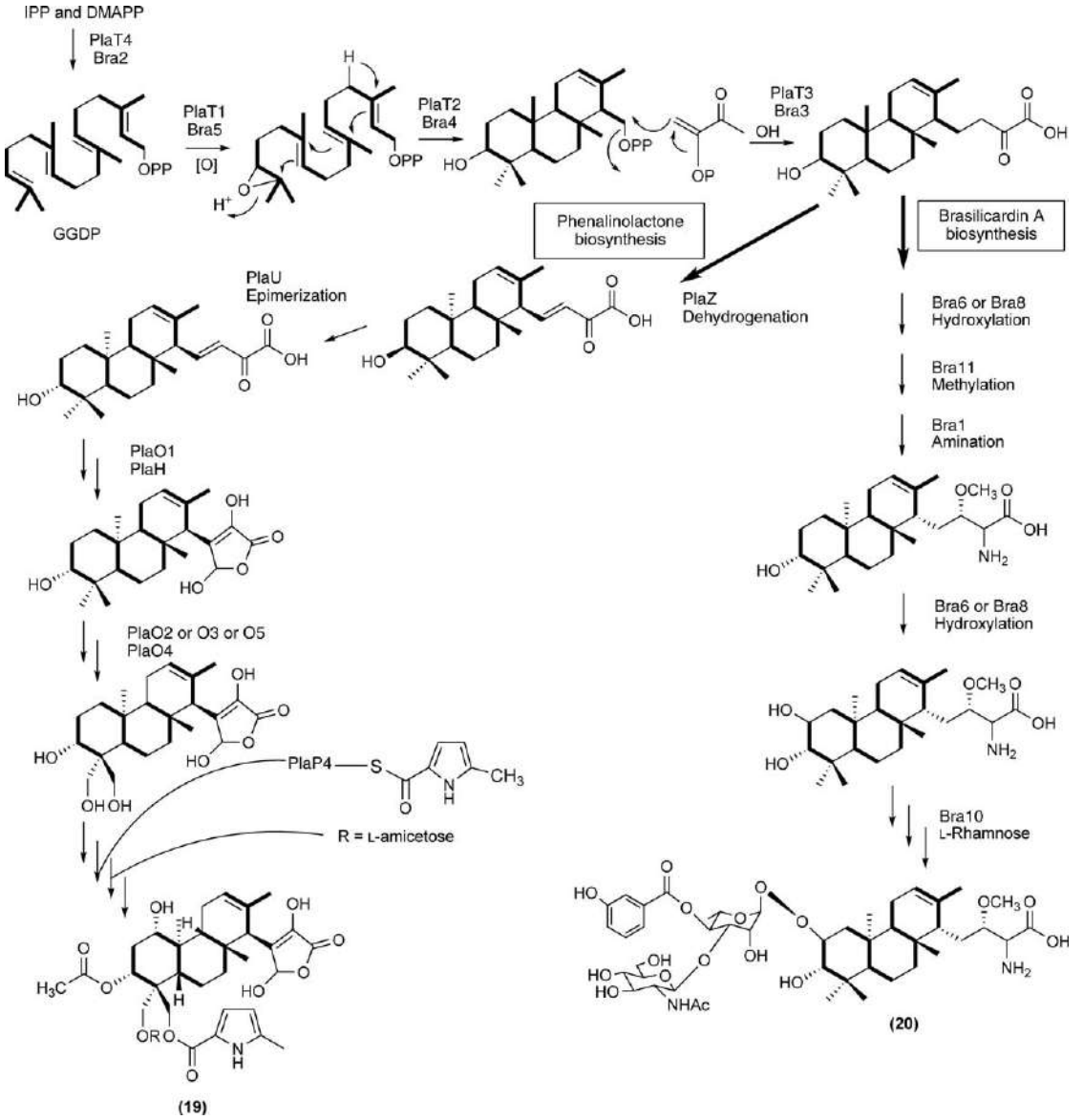


Figure 10 Proposed biosynthetic pathways of phenalinolactone A **(19)** and brasilicardin A **(20)**.

carotene is formed by terminal cyclization of neurosporene by lycopene cyclases. Biosynthetic gene clusters coding for these biosynthetic enzymes were identified in *S. coelicolor* (SCO0185–SCO0191) and *S. avermitilis* (SAV1019–SAV1025). In many species of actinomycetes, the biosynthesis of carotenoid has been known to be photoinduced.[61–63] Recently, this light regulation in *S. coelicolor* A3(2) was shown to be mediated by σLitS, a light-induced sigma factor, which directs transcription of the carotenoid biosynthesis gene cluster.[64]

A sole tetraterpene compound biosynthesized from octaprenyl diphosphate (OPD)[65] is KS-505a (longestin) (**24**) produced by *Streptomyces argenteolus*.[66] Compound **24** has a unique structure that consists of a tetraterpene (C_{40}) skeleton, to which a 2-*O*-methylglucuronic acid and an *o*-succinylbenzoate moiety are attached (**Figure 11**). Recently, a KS-505a biosynthetic gene cluster was cloned by using OPD synthase gene as a probe, which yielded a C_{40} precursor.[67] A gene disruption experiment was employed to confirm that the cluster indeed participated in KS-505a biosynthesis. The gene cluster consists of 24 ORFs though any genes related to glucuronic acid biosyntheses did not exist in the flanking regions. The cluster contained *o*-succinylbenzoate synthase gene and isochorismate synthase gene for *o*-succinylbenzoate biosynthesis, glycosyltransferase gene

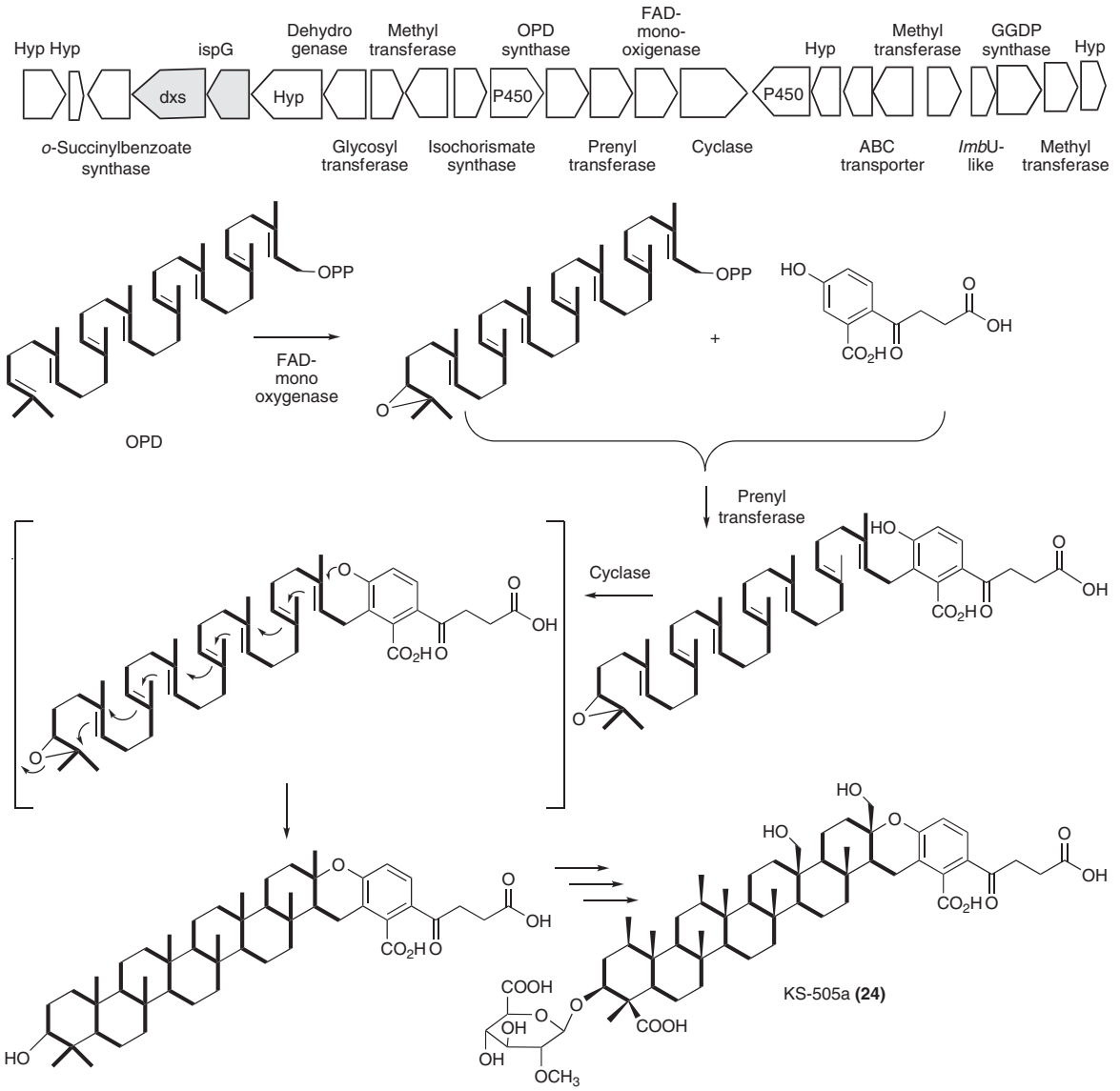


Figure 11 Biosynthetic gene cluster and proposed biosynthetic pathway of KS-505a (longestin) **(24)**.

for the attachment of glucuronic acid, two P-450 genes for hydroxylation, and three genes for methylation, in addition to OPD synthase, prenyltransferase, and isoprenoid cyclase genes.[67] The last two gene products have weak similarities to PlaT3[42]/Bra3[58] and PlaT2/Bra4, suggesting that KS-505a might also be synthesized from an epoxied intermediate (**Figure 11**). In the cluster, two MEP pathway genes, a 1-deoxy-D-xylulose-5-phosphate synthase gene and 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase gene, were also included similar to the biosynthetic gene cluster of **19**.[42]

1.22.3 Hybrid Isoprenoids

As described above, examples of cyclic isoprenoids that are formed by cyclization of GDP, FDP, or GGDP from actinomycete origin are limited. In many cases, isoprenoid moieties of compounds produced by actinomycetes are attached to other moieties, such as a polyketide, an aromatic ring, and an amino acid, that have

been synthesized via pathways independent of isoprenoid synthesis, to give the so-called 'isoprenoid hybrid compounds'. Therefore, the structural diversity of isoprenoid hybrid compounds produced by actinomycetes is summarized in this section.

1.22.3.1 Novobiocin/Clorobiocin

Prenyltransferases play a significant role in the structural diversity of isoprenoid hybrid compounds. Pioneering studies of a prenyltransferase, which does not show any similarities to the known prenyltransferases such as MenA/UbiA responsible for menaquinone/ubiquinone biosynthesis, have been performed by Heide and coworkers.[68,69] They found a new type of prenyltransferase gene (*clo*Q) in the biosynthetic gene cluster of clorobiocin (**25**).

Clorobiocin (**25**) and novobiocin (**26**), produced by *Streptomyces roseochromogenes* and *Streptomyces spheroids*/*Streptomyces niveus*, respectively, are composed of three moieties, a noviose sugar, a substituted coumarin, and a prenylated 4-hydroxybenzoic acid, and these rings are linked by glycosidic and amide bonds (**Figure 12**). Compound **26** was isolated in the 1960s and many studies have been carried out after that. The biosynthetic studies of these compounds were extensively studied by Heide and coworkers.[70] First, they cloned biosynthetic gene clusters of **26** with a PCR-amplified (dNDP)-glucose 4,6-dehydratase gene as a probe, followed by cloning of a gene cluster of **25** with **26** biosynthetic genes as probes.[71] Since both compounds have 3-dimethylallyl-4-hydroxybenzoic acid as a structural moiety and 3-prenylated 4-hydroxybenzoic acid moieties are known as intermediates in the biosynthesis of ubiquinones, they expected the presence of genes encoding UbiA-like enzymes in the biosynthetic gene clusters of **25** and **26**. However, they did not find any genes with similarity to known MenA/UbiA type prenyltransferases, nor genes containing the typical prenyl diphosphate binding site (N/D)DXXD. Therefore, they searched for a gene that is present in the biosynthetic gene clusters of **25** and **26** but absent in the biosynthetic gene cluster of coumermycin A_1,[72] which has a similar structure to that of **25** and **26**, but does not have a prenylated moiety. Eventually, a candidate gene (*clo*Q) was selected and confirmed to be the gene being sought by a gene disruption experiment.[69] A prenylation activity was also detected with a recombinant CloQ and 4-hydroxyphenylpyruvate and DMAPP as substrates. In the **26** producer, IPP was shown to be supplied via the MEP pathway by a tracer experiment.[73]

1.22.3.2 Polyketide-Isoprenoid Hybrid Compound

Since there are several kinds of isoprenoid hybrid compounds, compounds composed of polyketide and isoprenoid are described first. These are synthesized by attachments of C_5–C_{15} isoprene units by CloQ type prenyltransferases to polyketide moieties supplied by type III polyketide synthases. Furaquinocin A (**14**),[74] naphterpin (**27**),[75] furanonaphthoquinone I (**28**),[76] napyradiomycin A (**29**),[77] marinone (**30**),[78] neomarinone (**31**),[79] fumaquinone (**32**),[80] and naphthablin (**33**)[81] are involved in this type of compounds (**Figure 13**). Of these, many **14**-, **27**-, and **29**-related compounds have been isolated. For example, more than ten **29**-related compounds, such as napyradiomycin A1, 18-hydroxy A1, 18-oxo A1, A2a, A2b, 16-oxo A2, B1–B4, C1, C2 and SR, A80915A–D and G, SF2415A1–A3 and B1–B3, and naphthomevalin were isolated from the culture broth of

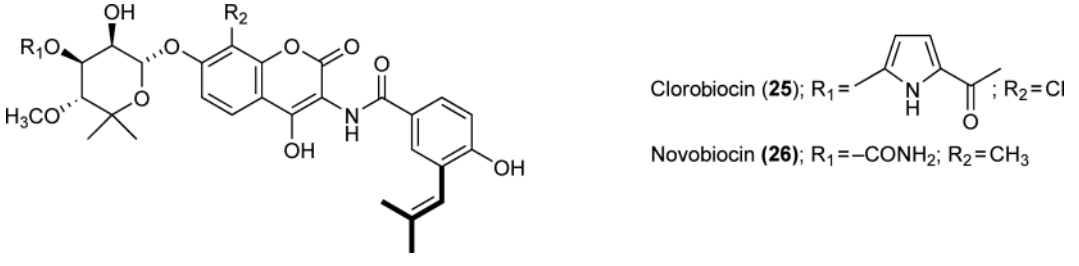


Figure 12 Structures of clorobiocin **(25)** and novobiocin **(26)**.

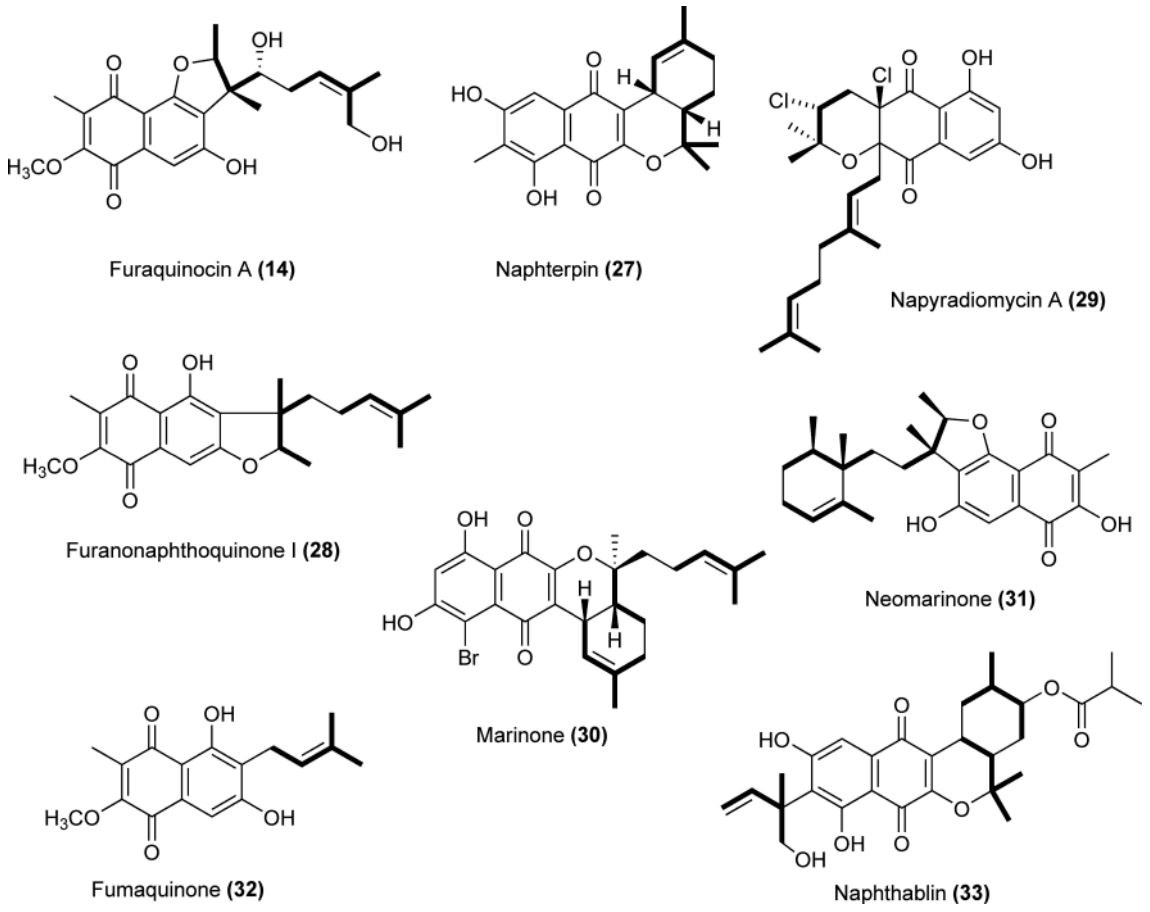


Figure 13 Structures of polyketide-isoprenoid hybrid compounds.

several actinomycetes including *Chainia rubra* and *Streptomyces aculeolatus*.[82–89] Moreover, eight **14**-related compounds (A–H)[90,91] and four **27**-related compounds (A–C and 7-demethyl)[92] were also isolated from the producers.

These compounds were reported to show biological activities and act as antioxidative agents, nonsteroidal estrogen-receptor antagonists, and antitumor and anticancer drugs. Considering that the structures of polyketide moieties, which are derived from 1,3,6,8-tetrahydroxynaphthalene (THN) biosynthesized by type III polyketide synthase,[33,35,93,94] are almost the same in these compounds, the prenyl moieties are suggested to play an important roles in exhibiting a diversity in the biological activities of these compounds. Therefore, a gene encoding this key enzyme (CloQ type) catalyzing the transfer of prenyl side chain to polyketide moiety was cloned from **27** producer and excellent biosynthetic studies with recombinant enzymes were carried out.[95]

The gene cluster for biosynthesis of **14** was cloned by taking advantage that an isoprenoid biosynthetic gene cluster generally exists in the flanking regions of the MV pathway gene cluster in actinomycetes possessing both the MEP and MV pathways as the first example of a gene cluster responsible for the biosynthesis of polyketide–isoprenoid hybrid compounds.[33] The cluster was heterologously expressed in *S. lividans* TK23 and the production of **14** was confirmed. The cluster contained 21 ORFs (**Figure 14**), in which THN synthases (*fur*1), *momA*-like monooxygenase catalyzing the oxidation of THN to flaviolin (*fur*2), methyltransferases for methylation of naphthalene moiety (*fur*4 and 6), CloQ-like prenyltransferase that catalyzes a reverse prenylation to flaviolin (*fur*7), P-450 for hydroxylation of prenyl side chain (*fur*8), and GDP synthase (*fur*19) were included.

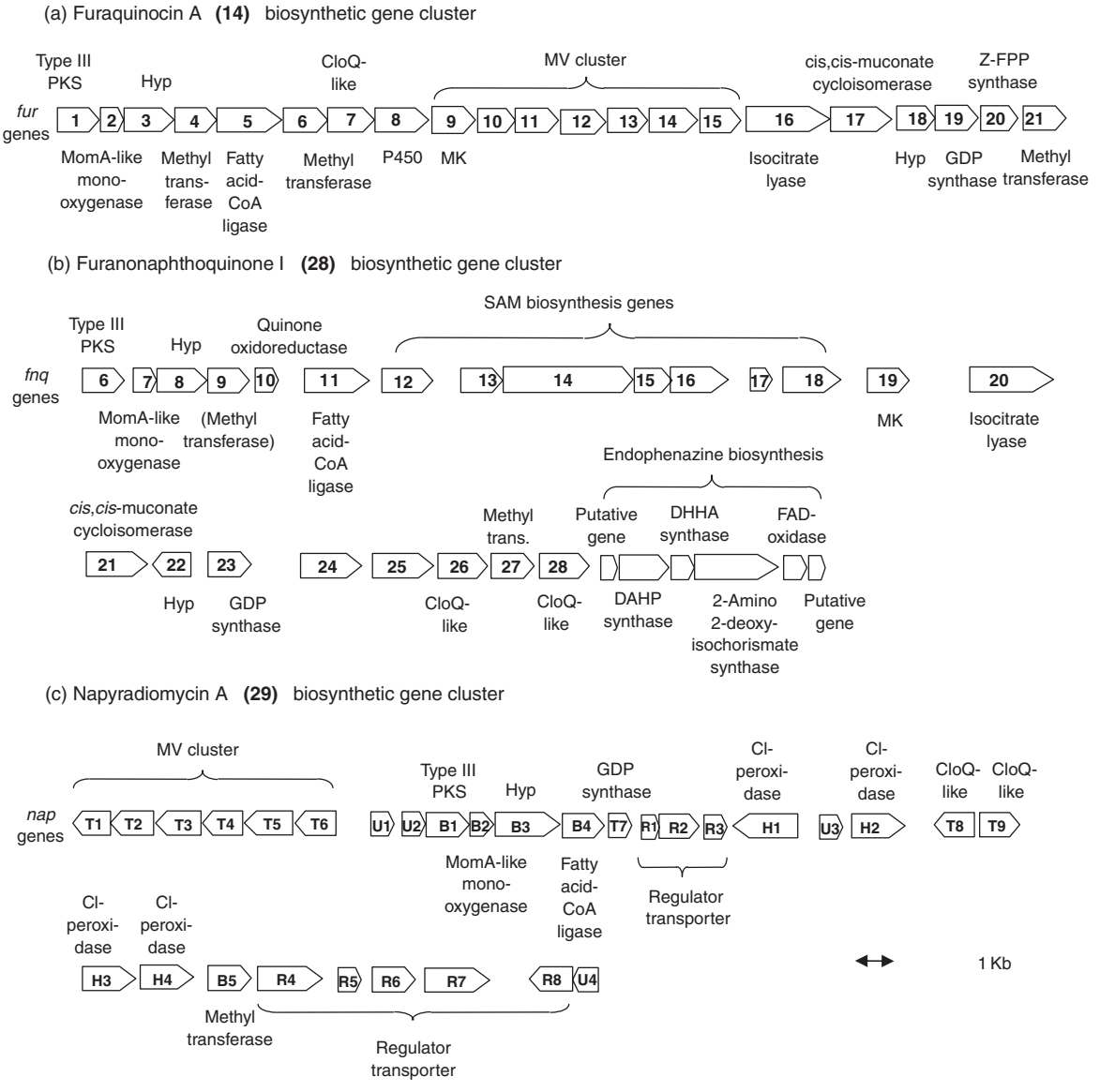


Figure 14 Biosynthetic gene clusters of furaquinocin A **(14)**, furanonaphthoquinone I **(28)**, and napyradiomycin A **(29)**.

Immediately after that, gene clusters of **28**[94] and **29**[35] were also cloned from *Streptomyces cinnamonensis* DSM 1042 and *Streptomyces aculeolatus* NRRL 18422, respectively (**Figure 14**). The former cluster was identified at just flanking region of endophenazine A (**40**),[96] a prenylated phenazine, biosynthetic genes. Of 28 *fnq* genes identified in the cluster, 13 showed high similarity to the *fur* genes. Besides the abovementioned genes (*fur*1/*fnq*6, *fur*2/*fnq*7, *fur*4/*fnq*9, *fur*6/*fnq*27, *fur*7/*fnq*26, and *fur*19/*fnq*23), the following genes commonly existed in both the clusters: *fur*3/*fnq*8, weakly homologous to aminotransferase; *fur*5/*fnq*11, fatty acid-CoA ligase; *fur*9/*fnq*19, mevalonate kinase; *fur*16/*fnq*20, hypothetical protein; *fur*17/*fnq*21, 3-carboxy-*cis*, *cis*-muconate cycloisomerase that is involved in the cyclization of the isoprenoid side chain with the 7-hydroxy group of flaviolin; *fur*18/*fnq*22, hypothetical protein; *fur*22/*fnq*4, SARP family regulator. Considering the structural similarity between the two compounds, the existence of common genes is quite reasonable. Curiously, the cluster contained two CloQ-like prenyltransferase genes, *fnq*26 and *fnq*28. By gene disruption experiments, the former was confirmed to be involved in **28** biosynthesis though disruption of the *fnq*28 gene showed no effects on the

production of both **28** and **40**. Substrate specificities of the two methyltransferases FNQ9 and FNQ27 were also clarified by gene disruptions, that is, the former is responsible for the 7-O-methylation and the latter for the 6-C-methylation reaction in **28** biosynthesis.

Very interestingly, the cluster of **28** contained only the mevalonate kinase (MK) genes (*fnq*19) in contrast to **14** and **29** producers, in which an MK gene constitutes an operon with other MV pathway genes (*fur*9–14 and T1–T6 in **Figure 14**). Therefore, a tracer experiment was performed to examine if the isoprenoid moieties of **28** and **40** were supplied via the MV and/or MEP pathway(s). Eventually, both compounds were mainly formed via the MV pathway (approximately 80%), showing that other MV pathway genes would be scattered at other loci in the strain DSM 1042.[97] In contrast to MV pathway genes, a complete set of genes that encode enzymes for recycling of *S*-adenosylhomocysteine to SAM, which is used as a substrate for methylation catalyzed by FNQ9 and FNQ27, is involved in the cluster (*fnq*12–16).[94] This is the first example that a complete set of SAM biosynthetic genes has been found as part of a secondary metabolic gene cluster.

The gene cluster for **29** biosynthesis was identified in two distinct strains, *S. aculeolatus* NRRL 18422 and the marine sediment-derived *Streptomyces* sp. CNQ-525, by genome screening approach and PCR-amplified THN synthase and prenyltransferase as probes, respectively.[35] The two clusters are similarly organized and 97% identical at the nucleotide level. The cluster was heterologously expressed in *Streptomyces albus* and the production of **29**-related compounds was confirmed. The cluster contained 33 ORFs including two transposases (**Figure 14**). Other 31 ORFs were categorized into five groups as follows: *nap*B1–B5, naphthoquinone polyketide synthesis; *nap*T1–T9, biosynthesis and attachment of isoprenoid; *nap*H1–H4, haloperoxidases; *nap*R1–R9, putative regulators and transporters; and *nap*U1–U4, whose function remains unknown. Of these, the *nap*B genes show high similarities to the corresponding orthologues in **14** and **28** biosynthetic gene clusters (*nap*B1, B2, B3, B4, and B5 to *fur*1/*fnq*6, *fur*2/*fnq*7, *fur*3/*fnq*8, *fur*5/*fnq*11, and *fur*4/*fnq*9, respectively). The *nap*T6, *nap*T5, *nap*T4, *nap*T3, *nap*T2, and *nap*T1, which encode the MV pathway enzymes, have high identities to *fur*9, *fur*10, *fur*11, *fur*12, *fur*13, and *fur*14, respectively. The *nap*T7 and *nap*T8 genes are also similar to *fur*19/*fnq*23 and *fur*7/*fnq*26, respectively. The most striking feature of this cluster is the presence of four haloperoxidase genes. Of these, one (*nap*H2) is an $FADH_2$-dependent and the other three (*nap*H1, H3, and H4) are vanadium-dependent haloperoxidases. Although the specificities of these four haloperoxidases remained unclear, the function of each enzyme in **29** biosynthesis was estimated based on phylogenetic analysis and comparison of amino acid sequences in the active site of the enzymes. A chloronium-induced cyclization mechanism of the two isoprene units was proposed.

1.22.3.3 Moenomycin A/Pholipomycin/AC326-Alpha

Moenomycin (**34**) is a phosphoglycolipid antibiotic with a unique C_{25} prenyl moiety (**Figure 15**). It was isolated in the 1960s and later its structurally related compounds pholipomycin[98] and AC326-alpha[99] were isolated. Compound **34** is known to be the sole antibiotic that prevents peptidoglycan biosynthesis by directly binding to the peptidoglycan glycosyltransferases. However, its development into practical therapeutic use has been prevented probably because of the physicochemical properties. Therefore, as the first step to develop **34** derivatives with better physicochemical properties, the biosynthetic machinery of this unique compound was studied. First, the biosynthetic mechanism of the unique C_{25} prenyl moiety of **34** was investigated by a feeding experiment with [2-^{13}C,4-^{2}H]-1-deoxy-D-xylulose and it was confirmed to be derived from C_{10} and C_{15} isoprene units in the mechanism described in **Figure 16**.[100–102] Moreover, recently, the biosynthetic gene clusters of **34** have been cloned from *Streptomyces ghanaensis* (ATCC14672) (**Figure 15**).[103] Three genes, *moe*A4, *moe*B4, and *moe*C4, which encode acyl-CoA ligase, amide synthetase, and aminolevulinate synthase, respectively, and participate in the biosynthesis of the C_5N subunit attached to the terminal glycoside, were cloned with PCR-amplified aminolevulinate synthase gene as a probe (cluster 2). Although these genes were confirmed to be essential for the biosynthesis of **34** by a gene disruption experiment, none of the other genes related to **34** biosynthesis existed in the flanking regions of these three genes. After several attempts to clone other biosynthetic genes, the main gene cluster for **34** biosynthesis (cluster 1) was identified by shotgun sequencing of the genome of the producer. The gene cluster was confirmed to be essential for the production of **34** by gene disruption and heterologous expression experiments. The cluster contained two genes, *moe*N5 and *moe*O5, which are related to biosynthesis of prenyl moiety. MoeN5 shows homology to other polyprenyl diphosphate synthases such as FDP synthase and GGDP synthase and is suggested to be involved in the

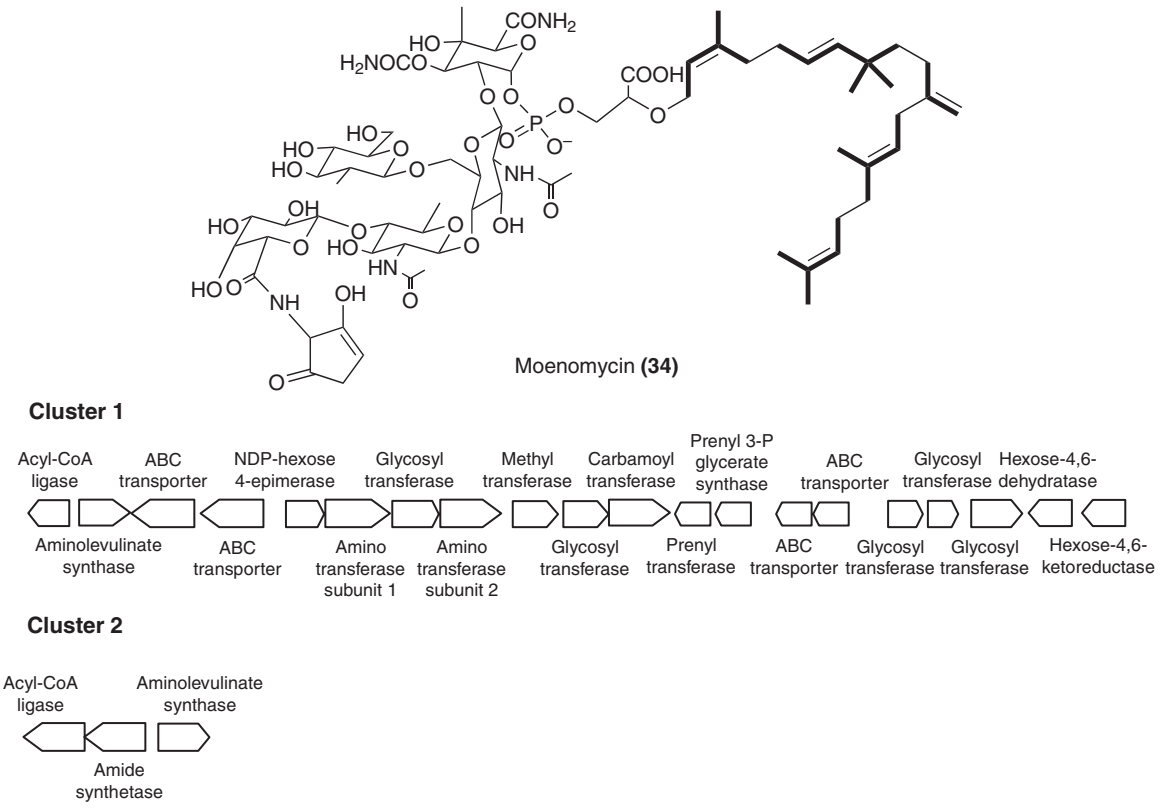


Figure 15 Structure and biosynthetic gene cluster of moenomycin **(34)**.

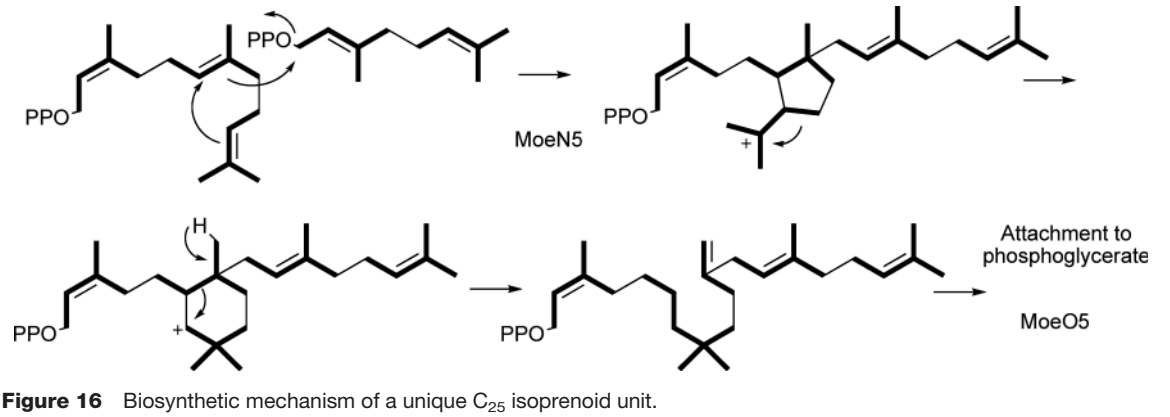


Figure 16 Biosynthetic mechanism of a unique C_{25} isoprenoid unit.

synthesis of the C_{25} lipid chain (**Figure 16**). MoeO5, which shows weak similarity to the archaeal enzymes that form the first ether linkage between isoprenoid precursors and glycerol phosphate in the biosynthesis of the unusual membrane lipids, is predicted to catalyze the formation of ether linkage between 3-*R*-phosphoglycerate and isoprenoid chain.

1.22.3.4 Phenazine-Isoprenoid Hybrid Compounds

To date, several phenazine-isoprenoid hybrid compounds have been isolated. Lavanducyanin (**35**),[104] WS-9659 B (**36**),[105] benthocyanin A (**37**),[106] benthophoenin (**38**),[107] aestivophoenins A (**39**),[108] and endophenazine A

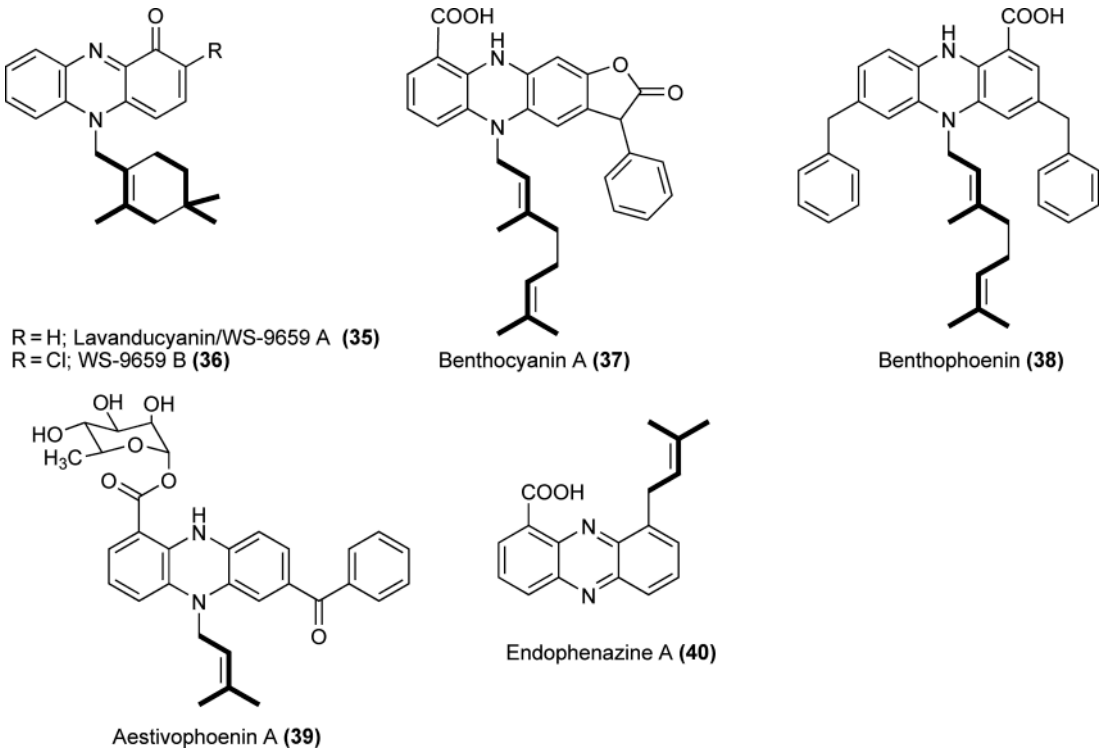


Figure 17 Structures of phenazine-isoprenoid hybrid compounds.

(**40**)[96] were isolated as antitumor antibiotic, testosterone 5α-reductase inhibitor, potent radical scavenger, free radical scavenger, neuronal cell protecting substances, and antibiotic against Gram-positive bacteria, respectively (**Figure 17**). As described above, six genes, all of which have significant similarity to phenazine biosynthetic genes from *Pseudomonas,* were identified in the region immediately next to the biosynthetic gene cluster of **28** and confirmed to be responsible for the biosynthesis of **40**.[94] As described, the clusters for biosyntheses of **28** and **40** contained two CloQ-like prenyltransferase genes, *fnq*26 and *fnq*28. By gene disruption experiments, the former was confirmed to be involved in **28** biosynthesis. However, disruption of the *fnq*28 gene showed no effects on the production of both **28** and **40**. Therefore, a prenylation gene responsible for the biosynthesis of **40** remains unclear.

1.22.3.5 Carbazole-Isoprenoid Hybrid Compounds

Neocarazostatins (**41**),[109] carquinostatins (**42**),[110,111] and lavanduquinocin (**43**)[112] were isolated from *Streptomyces* sp. strain GP 38, *Streptomyces exfoliates*, and *Streptomyces viridochromogenes*, respectively (**Figure 18**). Compound **41** had free radical scavenging activities and the last two compounds showed neuronal cell protecting activities. There are no biosynthetic studies on these compounds except that the prenyl moiety of **42** was shown to be synthesized via the MEP pathway by tracer experiments.[113]

1.22.3.6 Other Isoprenoid Hybrid Compounds

Teleocidin (**44**) is a well-known naturally occurring tumor promoter. Since the isolation of **44** in the early 1960s, many **44**-related compounds have been isolated.[114–128] For instance, teleocidin A-1, A-2, B-1, B-2, B-3, B-4, and B-18, des-*O*-methylolivoretin C, and des-*N*-methylteleocidin B-4 were isolated from *Streptomyces mediocidicus*. Blastmycetin A, B, C, D, E, and F, (−)-indolactam-V, (−)-14-*O*-malonylindolactam-V, (−)-14-*O*-acetylindolactam-V, (−)-7-geranylindolactam-V, *N*13-desmethylteleocidin A-1, *N*13-desmethylteleocidin B-4, (−)-2-oxy-indolactam, olivoretin A (14-*O*-methylteleocidin B), olivoretin B, and olivoretin C were

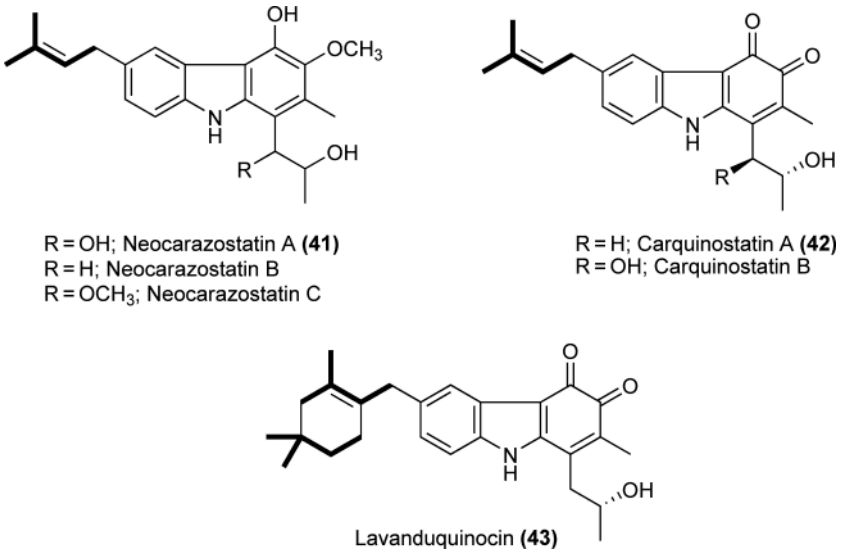


Figure 18 Structures of carbazole-isoprenoid hybrid compounds.

isolated from *Streptoverticillium blastmyceticum* NA34-17. Olivoretin A, olivoretin B, olivoretin C, olivoretin D (teleocidin B), olivoretin E, and de-*O*-methylolivoretin C were isolated from *Streptoverticillium olivoreticulus*. Moreover, pendolmycin and 14-*O*-(*N*-acetylglucosaminyl) teleocidin A were reported to be produced by *Nocardiopsis* and *Streptomyces*, respectively. In many cases, these compounds have a monoterpene moiety (**Figure 19**).

To date, there are many reports on the mechanisms of the action of **44**-related compounds. In contrast, limited studies on the biosynthesis of these compounds have been reported. (−)-Indolactam V, the core skeleton of **44**, was shown to be biosynthesized from L-tryptophan, L-valine, and L-methionine via *N*-methyl-L-valyl-L-tryptophanol by feeding experiments with *S. blastmyceticum*.[129] In addition, the monoterpenoid moiety of teleocidin B-4 was shown to be biosynthesized via the MEP pathway in *S. blastmyceticum* NA34-17.[130,131]

Virantmycin (**45**) is composed of tetrahydroquinoline moiety and monoterpene moiety (**Figure 19**). It was isolated from *Streptomyces nitrosporeus* as an antiviral antibiotic.[132–134] After that, **45**-related compounds,

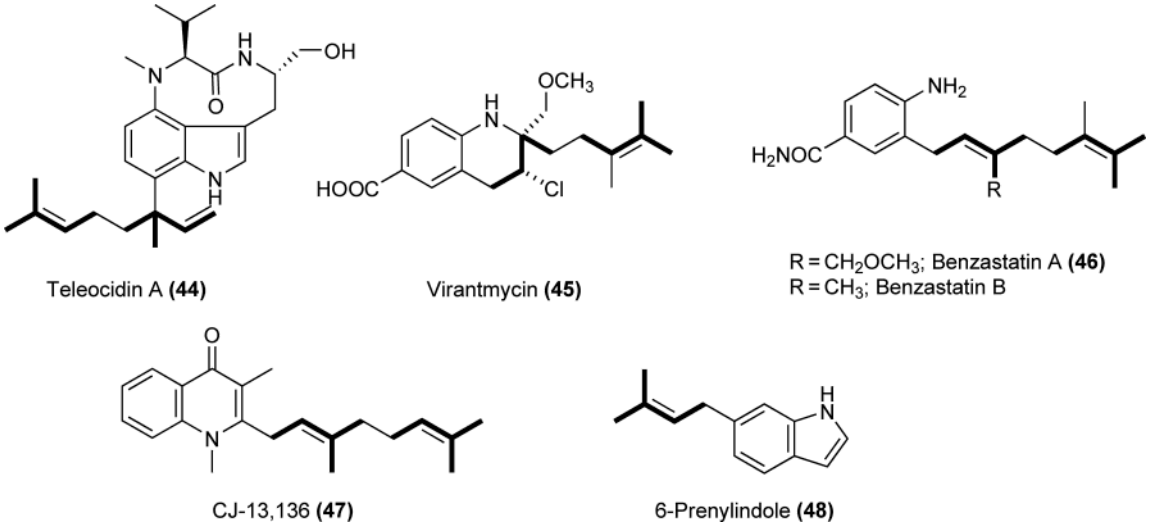


Figure 19 Structures of other isoprenoid hybrid compounds.

benzastatin A (**46**), B, C, and D, were isolated as free radical scavengers from *S. nitrosporeus*.[135–137] Furthermore, benzastatin E–I were also isolated as neuronal cell protecting substances. The compound CJ-13,136 (**47**) was isolated from *Pseudonocardia* sp. CL38489 as a potent anti-*Helicobacter pylori* activity and its semisynthetic epoxy derivative, CJ-13,564, showed a high activity.[138] 6-Prenylindole (**48**), a simple compound, was isolated from *Streptomyces* sp. TP-A0595 and was shown to have antifungal activity.[139]

Very unique isoprenoid hybrid compounds have been isolated from marine actinomycetes, especially from marine sediments (**Figure 20**): azamerone (**49**)[140] isolated from a marine-derived bacterium related to the genus *Streptomyces* has a unique structure composed of a phthalazinone core with chlorinated C_{10} and C_5 isoprenoid units; glaciapyrroles (**50**),[141] an unprecedented unique compound, was also produced by a *Streptomyces* sp. (NPS008187) isolated from a marine sediment collected in Alaska; antibiotic BU-4664L (**51**)[142] was isolated from *Micromonospora* sp. M990-6 and its structure was recently revised.

Recently, platensimycin (**52**)[143] and platencin (**53**),[144] which were produced by *Streptomyces platensis* MA7327 and *S. platensis* MA7339, respectively, were isolated as antibiotics that inhibit cellular lipid biosynthesis. Both compounds consist of a 3-amino-2,4-dihydroxy benzoic acid moiety and an isoprenoid side chain (**Figure 21**). By a tracer experiment, the isoprenoid moiety of **52** was proposed to be derived from GGDP via complicated reactions similar to those in *ent*-kaurene biosynthesis, in which cyclization of GGDP might be involved.[145]

In some microorganisms including actinomycetes, menaquinone (**54**) is an obligatory component of the electron transfer pathway. Compound **54** is composed of a naphthoquinone moiety and an isoprenoid side chain. The naphthoquinone moiety is derived from chorismate by seven enzymes (MenA–G) in *E. coli*. However, very recently, a novel pathway was demonstrated to be operated in a *Streptomyces* strain by tracer experiments (**Figure 22**).[146] The results showed that an alternative biosynthetic pathway might be operating in *Streptomyces* strains.[147]

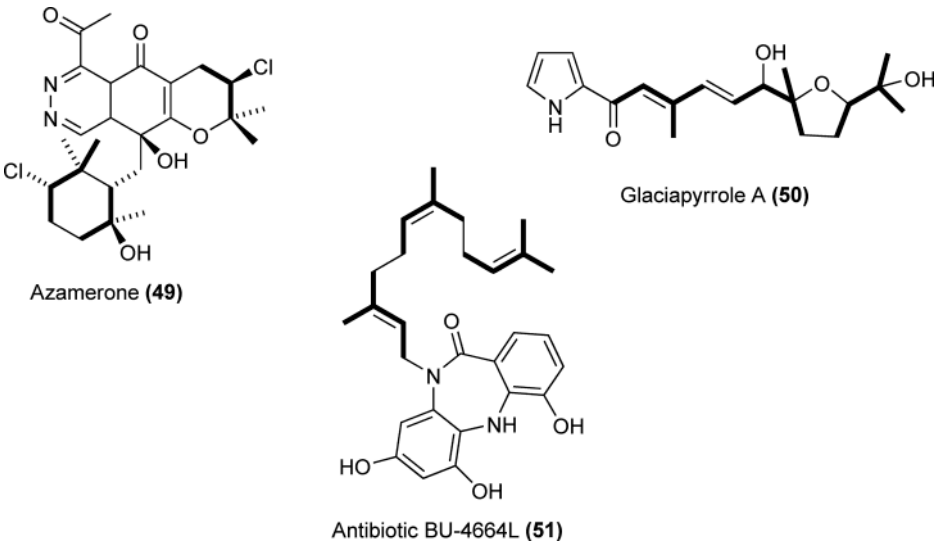


Figure 20 Structures of isoprenoid hybrid compounds isolated from marine actinomycetes.

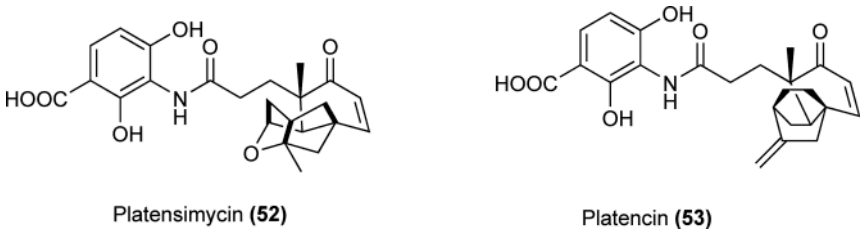


Figure 21 Structures of platensimycin **(52)** and platencin **(53)**.

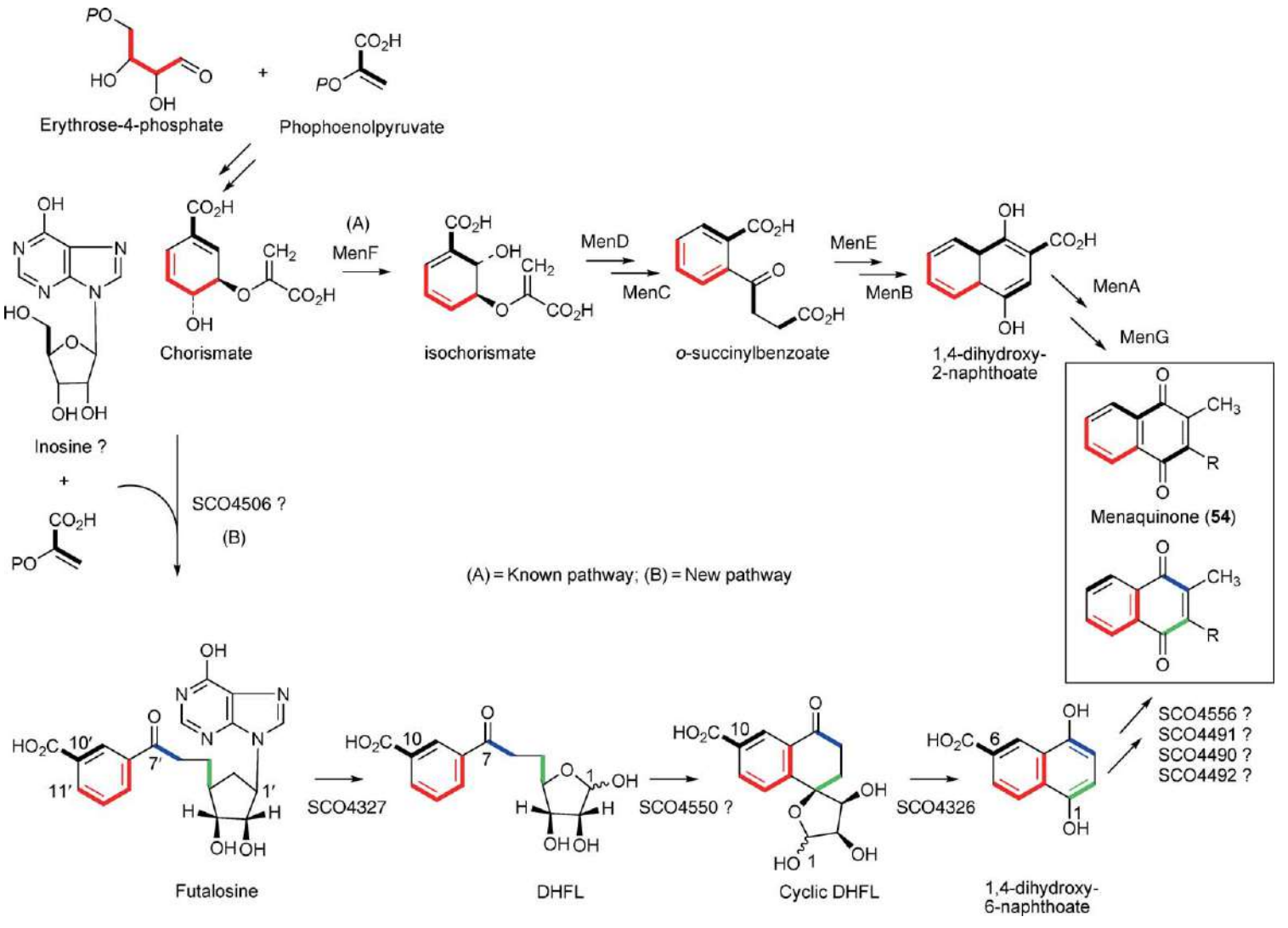


Figure 22 MK biosynthetic pathways. (a) Classical pathway. Chorismate, which is derived from the shikimate pathway, is initially converted into isochorismate by MenF (isochorismate synthase) and then into 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate by MenD, which is a thiamine-dependent enzyme. This compound is dehydrated by MenC to give an aromatic compound, *o*-succinylbenzoate, followed by the attachment of coenzyme A to yield *o*-succinylbenzoyl-CoA by MenE. *o*-Succinylbenzoyl-CoA is then converted into 1,4-dihydroxy-2-naphthoate by MenB. In the last two steps of the pathway, MK is synthesized by MenA and MenG, which catalyze prenylation and *S*-adenosylmethionine-dependent methylation, respectively. (b) Alternative pathway. Green and blue bold lines indicate two carbon units derived from C-5 and C-6 of glucose via different metabolic pathways. Red and black bold lines show carbons derived from erythrose-4-phosphate and phosphoenolpyruvate, respectively.

Abbreviations

ACP	3-oxoacyl-(acyl-carrier-protein)
CDP	coplalyl diphosphate
DMAPP	dimethylallyl diphosphate
FDP	farnesyl diphosphate
GDP	geranyl diphosphate
GGDP	geranylgeranyl diphosphate
HMG-CoA	3-hydroxy-3-methylglutaryl coenzyme A
HMGR	HMG-CoA reductase
HMGS	HMG-CoA synthase
IPP	isopentenyl diphosphate
MDPD	mevalonate diphosphate decarboxylase
MEP	2*C*-methyl-D-erythritol-4-phosphate
MK	mevalonate kinase
MV	mevalonate
OPD	octaprenyl diphosphate
ORF	open reading frame
PCR	polymerase chain reaction
PMD	pimara-9(11),15-diene
PMK	phosphomevalonate kinase
TDP	terpentedienol diphosphate
THN	1,3,6,8-tetrahydroxynaphthalene
TTE	terpentetriene
UbiA	ubiquinone

References

1. D. E. Cane; J. K. Sohng; C. R. Lamberson; S. M. Rudnicki; Z. Wu; M. D. Lloyd; J. S. Oliver; B. R. Hubbard, *Biochemistry* **1994**, *33*, 5846.
2. N. N. Gerber, *J. Antibiot.* **1969**, *22*, 508.
3. M. Komatsu; M. Tsuda; S. Omura; H. Oikawa; H. Ikeda, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 7422.
4. H. Seto; H. Yonehara, *J. Antibiot.* **1980**, *33*, 92.
5. B. K. Koe; B. A. Sobin; W. D. Celmer, *Antibiot. Ann.* **1956–1957**, 672–675.
6. S. Takeuchi; Y. Ogawa; H. Yonehara, *Tetrahedron Lett.* **1969**, *32*, 2737.
7. Y. Tsuchiya; A. Matsumoto; K. Shudo; T. Okamoto, *J. Pharm. Soc. Jpn.* **1980**, *100*, 468.
8. N. N. Gerber; H. A. Lechevalier, *Appl. microbiol.* **1965**, *13*, 935.
9. X. Lin; R. Hopson; D. E. Cane, *J. Am. Chem. Soc.* **2006**, *128*, 6022.
10. H. Gurtler; R. Pedersen; U. Anthoni; C. Christophersen; P. H. Nielsen; E. M. H. Wellington; M. H. Elizabeth; C. Pedersen; K. Bock, *J. Antibiot.* **1994**, *47*, 434.
11. S. J. Wu; S. Fotso; F. Li; S. Qin; H. Laatsch, *J. Nat. Prod.* **2007**, *70*, 304.
12. N. N. Gerber, *Phytochemistry* **1972**, *11*, 385.
13. S. J. Wu; S. Fotso; F. Li; S. Qin; G. Kelter; H. H. Fiebig; H. Laatsch, *J. Antibiot.* **2006**, *59*, 331.
14. J.-F. Hu; D. Wunderlich; R. Thiericke; H.-M. Dahse; S. Grabley; X.-Z. Feng; I. Sattler, *J. Antibiot.* **2003**, *56*, 747.
15. K. Torigoe; N. Wakasugi; N. Sakaizumi; T. Ikejima; H. Suzuki; K. Kojiri; H. Suda, *J. Antibiot.* **1996**, *49*, 314.
16. M. Seemann; G. Zhai; K. Umezawa; D. E. Cane, *J. Am. Chem. Soc.* **1999**, *121*, 591.
17. M. Seemann; G. Zhai; J.-W. de Kraker; C. M. Paschall; D. W. Christianson; D. E. Cane, *J. Am. Chem. Soc.* **2002**, *124*, 7681.
18. C. N. Tetzlaff; Y. Z. You; D. E. Cane; S. Takamatsu; S. Omura; H. Ikeda, *Biochemistry* **2006**, *45*, 6179.
19. R. Quaderer; S. Omura; H. Ikeda; D. E. Cane, *J. Am. Chem. Soc.* **2006**, *128*, 13036.
20. Z. You; S. Omura; H. Ikeda; D. E. Cane, *J. Am. Chem. Soc.* **2006**, *128*, 6566.
21. Z. You; S. Omura; H. Ikeda; D. E. Cane, *Arch. Biochem. Biophys.* **2007**, *459*, 233.
22. D. E. Cane; R. M. Watt, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1547.
23. X. He; D. E. Cane, *J. Am. Chem. Soc.* **2004**, *126*, 2678.
24. J. Jiang; X. He; D. E. Cane, *J. Am. Chem. Soc.* **2006**, *128*, 8128.
25. D. E. Cane; X. He; S. Kobayashi; S. Omura; H. Ikeda, *J. Antibiot.* **2006**, *59*, 471.
26. J. Jiang; X. He; D. E. Cane, *Nat. Chem. Biol.* **2007**, *3*, 711.
27. J. Jiang; D. E. Cane, *J. Am. Chem. Soc.* **2008**, *130*, 428.

28. T. Nawrath; J. S. Dickschat; R. Mueller; J. Jiang; D. E. Cane; S. Schulz, *J. Am. Chem. Soc.* **2008**, *130*, 430.
29. B. Zhao; X. Lin; L. Lei; D. C. Lamb; S. L. Kelly; M. R. Waterman; D. E. Cane, *J. Biol. Chem.* **2008**, *283*, 8183.
30. H. Seto; N. Orihara; K. Furihata, *Tetrahedron Lett.* **1998**, *39*, 9497.
31. T. Kawasaki; T. Kuzuyama; K. Furihata; N. Itoh; H. Seto; T. Dairi, *J. Antibiot.* **2003**, *56*, 957.
32. T. Dairi; Y. Hamano; T. Kuzuyama; N. Itoh; K. Furihata; H. Seto, *J. Bacteriol.* **2001**, *183*, 6085.
33. T. Kawasaki; Y. Hayashi; T. Kuzuyama; K. Furihata; N. Itoh; H. Seto; T. Dairi, *J. Bacteriol.* **2006**, *188*, 1236.
34. T. Kawasaki; T. Kuzuyama; Y. Kuwamori; N. Matsuura; N. Itoh; K. Furihata; H. Seto; T. Dairi, *J. Antibiot.* **2004**, *57*, 739.
35. J. M. Winter; M. C. Moffitt; E. Zazopoulos; J. B. McAlpine; P. C. Dorrestein; B. S. Moore, *J. Biol. Chem.* **2007**, *282*, 16362.
36. X. Wu; P. M. Flatt; O. Schlorke; A. Zeeck; T. Dairi; T. Mahmud, *ChemBioChem* **2007**, *8*, 239.
37. T. Tamamura; T. Sawa; K. Isshiki; T. Masuda; Y. Homma; H. Iinuma; H. Naganawa; M. Hamada; T. Takeuchi; H. Umezawa, *J. Antibiot.* **1985**, *38*, 1664.
38. K. Isshiki; T. Tamamura; Y. Takahashi; T. Sawa; H. Naganawa; T. Takeuchi; H. Umezawa, *J. Antibiot.* **1985**, *38*, 1819.
39. Y. Uosaki; S. Kawada; H. Nakano; Y. Saitoh; H. Sano, *J. Antibiot.* **1993**, *46*, 235.
40. K. Motohashi; R. Ueno; M. Sue; K. Furihata; T. Matsumoto; T. Dairi; S. Omura; H. Seto, *J. Nat. Prod.* **2007**, *70*, 1712.
41. C. Nakano; T. Okamura; T. Sato; T. Dairi; T. Hoshino, *Chem. Commun.* **2005**, (8), 1016.
42. C. Dürr; H. J. Schnell; A. Luzhetskyy; R. Murillo; M. Weber; K. Welzel; A. Vente; A. Bechthold, *Chem. Biol.* **2006**, *13*, 365.
43. H. Komaki; A. Nemoto; Y. Tanaka; H. Takagi; K. Yazawa; Y. Mikami; H. Shigemori; J. Kobayashi; A. Ando; Y. Nagata, *J. Antibiot.* **1999**, *52*, 13.
44. H. Komaki; Y. Tanaka; K. Yazawa; H. Takagi; A. Ando; Y. Nagata; Y. Mikami, *J. Antibiot.* **2000**, *53*, 75.
45. A. Fukumoto; Y. P. Kim; A. Matsumoto; Y. Takahashi; K. Shiomi; H. Tomoda; S. Omura, *J. Antibiot.* **2008**, *61*, 1.
46. M. Nakajima; T. Okazaki; S. Iwado; T. Kinoshita; T. Haneishi, *J. Antibiot.* **1989**, *42*, 1741.
47. T. Aoyama; H. Naganawa; Y. Muraoka; T. Aoyagi; T. Takeuchi, *J. Antibiot.* **1992**, *45*, 703.
48. K. Isshiki; T. Tamamura; T. Sawa; H. Naganawa; T. Takeuchi; H. Umezawa, *J. Antibiot.* **1986**, *39*, 1634.
49. T. Dairi; Y. Motohira; T. Kuzuyama; S. Takahashi; N. Itoh; H. Seto, *Mol. Gen. Genet.* **2000**, *262*, 957.
50. Y. Hamano; T. Dairi; M. Yamamoto; T. Kawasaki; K. Kaneda; T. Kuzuyama; N. Itoh; H. Seto, *Biosci. Biotechnol. Biochem.* **2001**, *65*, 1627.
51. Y. Hamano; T. Kuzuyama; N. Itoh; K. Furihata; H. Seto; T. Dairi, *J. Biol. Chem.* **2002**, *277*, 37098.
52. T. Eguchi; Y. Dekishima; Y. Hamano; T. Dairi; H. Seto; K. Kakinuma, *J. Org. Chem.* **2003**, *68*, 5433.
53. C. Ikeda; Y. Hayashi; N. Itoh; H. Seto; T. Dairi, *J. Biochem.* **2007**, *141*, 37.
54. M. Cho; A. Okada; H. Kenmoku; K. Otomo; T. Toyomasu; W. Mitsuhashi; T. Sassa; A. Yajima; G. Yabuta; K. Mori; H. Oikawa; H. Toshima; N. Shibuya; H. Nojiri; T. Omori; M. Nishiyama; H. Yamane, *Plant J.* **2004**, *37*, 1.
55. P. R. Wilderman; M. Xu; Y. Jin; R. M. Coates; R. J. Peters, *Plant Physiol.* **2004**, *135*, 2098.
56. K. Otomo; Y. Kanno; A. Motegi; H. Kenmoku; H. Yamane; W. Mitsuhashi; H. Oikawa; H. Toshima; H. Itoh; M. Matsuoka; T. Sassa; T. Toyomasu, *Biosci. Biotechnol. Biochem.* **2004**, *68*, 2001.
57. T. Nemoto; E. M. Cho; A. Okada; K. Okada; K. Otomo; Y. Kannno; T. Toyomasu; W. Mitsuhashi; T. Sassa; E. Minami; N. Shibuya; M. Nishiyama; H. Nojiri; H. Yamane, *FEBS Lett.* **2004**, *571*, 182.
58. Y. Hayashi; N. Matsuura; H. Toshima; N. Itoh; J. Ishikawa; Y. Mikami; T. Dairi, *J. Antibiot.* **2008**, *61*, 164.
59. Y. Hamano; T. Dairi; M. Yamamoto; T. Kuzuyama; N. Itoh; H. Seto, *Biosci. Biotechnol. Biochem.* **2002**, *66*, 808.
60. J. M. Esteves; A. Cantero; A. Reindl; S. Reichler; P. Leon, *J. Biol. Chem.* **2001**, *276*, 22901.
61. F. Kato; T. Hino; A. Nakaji; M. Tanaka; Y. Koyama, *Mol. Gen. Genet.* **1995**, *247*, 387.
62. G. Schumann; H. Nürnberger; G. Sandmann; H. Krügel, *Mol. Gen. Genet.* **1996**, *252*, 658.
63. H. Krügel; P. Krubasik; K. Weber; H. P. Saluz; G. Sandmann, *Biochim. Biophys. Acta* **1999**, *1439*, 57.
64. H. Takano; S. Obitsu; T. Beppu; K. Ueda, *J. Bacteriol.* **2005**, *187*, 1825.
65. H. Seto; H. Watanabe; N. Orihara; K. Furihata, Tennen Yuki Kagobutsu Toronkai Koen Yoshishu. In *Proceedings of the 38th Symposium on the Chemistry of Natural Products, Sendai, Japan*, 1996, 19.
66. S. Nakanishi; K. Osawa; Y. Saito; I. Kawamoto; K. Kuroda; H. Kase, *J. Antibiot.* **1992**, *45*, 341.
67. Y. Hayashi; H. Onaka; N. Itoh; H. Seto; T. Dairi, *Biosci. Biotechnol. Biochem.* **2007**, *71*, 3072.
68. M. Steffensky; S.-M. Li; B. Vogler; L. Heide, *FEMS Microbiol. Lett.* **1998**, *161*, 69.
69. F. Pojer; E. Wemakor; B. Kammerer; H. Chen; C. T. Walsh; S.-M. Li; L. Heide, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 2316.
70. M. Steffensky; A. Mühlenweg; Z.-X. Wang; S.-M. Li; L. Heide, *Antimicrob. Agents Chemother.* **2000**, *44*, 1214.
71. F. Pojer; S.-M. Li; L. Heide, *Microbiology* **2002**, *148*, 3901.
72. Z.-X. Wang; S.-M. Li; L. Heide, *Antimicrob. Agents Chemother.* **2000**, *44*, 3040.
73. N. Orihara; T. Kuzuyama; S. Takahasi; K. Furihata; H. Seto, *J. Antibiot.* **1998**, *51*, 676.
74. S. Funayama; M. Ishibashi; Y. Anraku; K. Komiyama; S. Omura, *Tetrahedron Lett.* **1989**, *30*, 7427.
75. K. Shinya; S. Imai; K. Furihata; Y. Hayakawa; Y. Kato; G. D. Vanduyne; J. Clardy; H. Seto, *J. Antibiot.* **1990**, *43*, 444.
76. P. Sedmera; S. Pospisil; J. Novak, *J. Nat. Prod.* **1991**, *54*, 870.
77. K. Shiomi; H. Iinuma; M. Hamada; H. Naganawa; M. Manabe; C. Matsuki; T. Takeuchi; H. Umezawa, *J. Antibiot.* **1986**, *39*, 487.
78. C. Pathirana; P. R. Jensen; W. Fenical, *Tetrahedron Lett.* **1992**, *33*, 7663.
79. I. H. Hardt; P. R. Jensen; W. Fenical, *Tetrahedron Lett.* **2000**, *41*, 2073.
80. R. D. Charan; G. Schlingmann; V. S. Bernan; X. Feng; G. T. Carter, *J. Antibiot.* **2005**, *58*, 271.
81. K. Umezawa; S. Masuoka; T. Ohse; H. Naganawa; S. Kondo; Y. Ikeda; N. Kinoshita; M. Hamada; T. Sawa; T. Takeuchi, *J. Antibiot.* **1995**, *48*, 604.
82. K. Shiomi; H. Nakamura; H. Iinuma; H. Naganawa; K. Isshiki; T. Takeuchi; H. Umezawa; Y. Iitaka, *J. Antibiot.* **1986**, *39*, 494.
83. S. Gomi; S. Ohuchi; T. Sasaki; J. Itoh; M. Sezaki, *J. Antibiot.* **1987**, *40*, 740.
84. K. Shiomi; H. Nakamura; H. Iinuma; H. Naganawa; T. Takeuchi; H. Umezawa; Y. Iitaka, *J. Antibiot.* **1987**, *40*, 1213.
85. D. S. Fukuda; J. S. Mynderse; P. J. Baker; D. M. Berry; L. D. Boeck; R. C. Yao; F. P. Mertz; W. M. Nakatsukasa; J. Mabe; J. Ott; T. Counter; P. W. Ensminger; N. E. Allen; W. E. Alborn; J. N. Hobbs *J. Antibiot.* **1990**, *43*, 623.
86. T. Henkel; A. Zeeck, *J. Antibiot.* **1991**, *44*, 665.
87. Y. Hori; Y. Abe; N. Shigematsu; T. Goto; M. Okuhara; M. Kohsaka, *J. Antibiot.* **1993**, *46*, 1890.

88. I. E. Soria-Mercado; A. Prieto-Davo; P. R. Jensen; W. Fenical, *J. Nat. Prod.* **2005**, *68*, 904.
89. K. Motohashi; M. Sue; K. Furihata; S. Ito; H. Seto, *J. Nat. Prod.* **2008**, *71*, 595.
90. M. Ishibashi; S. Funayama; Y. Anraku; K. Komiyama; S. Omura, *J. Antibiot.* **1991**, *44*, 390.
91. P. G. Dormer; A. B. Smith; S. Funayama; S. Omura, *Tetrahedron Lett.* **1992**, *33*, 1717.
92. K. Shinya; A. Shimazu; Y. Hayakawa; H. Seto, *J. Antibiot.* **1992**, *45*, 124.
93. N. Funa; Y. Ohnishi; I. Fijii; M. Shibuya; Y. Ebizuka; S. Horinouchi, *Nature* **1999**, *400*, 897.
94. Y. Haagen; K. Glueck; K. Fay; B. Kammerer; B. Gust; L. Heide, *ChemBioChem* **2006**, *7*, 2016.
95. T. Kuzuyama; L. P. Noel; S. B. Richard, *Nature* **2005**, *435*, 983.
96. P. Krastel; A. Zeeck; K. Gebhardt; H.-P. Fiedler; J. Rheinheimer, *J. Antibiot.* **2002**, *55*, 801.
97. G. Bringmann; Y. Haagen; T. A. M. Gulder; T. Gulder; L. Heide, *J. Org. Chem.* **2007**, *72*, 4198.
98. M. Arai; A. Torikata; R. Enokita; H. Fukatsu; R. Nakayama; K. Yoshida, *J. Antibiot.* **1977**, *30*, 1049.
99. H. He; B. Shen; J. Korshalla; M. M. Siegel; G. T. Carter, *J. Antibiot.* **2000**, *53*, 191.
100. U. Schuricht; L. Hennig; M. Findeisen; P. Welzel, *Tetrahedron Lett.* **2000**, *41*, 3047.
101. U. Schuricht; L. Hennig; M. Findeisen; P. Welzel; D. Arigoni, *Tetrahedron Lett.* **2001**, *42*, 3835.
102. I. Neundorf; C. Koehler; L. Hennig; M. Findeisen; D. Arigoni; P. Welzel, *ChemBioChem* **2003**, *4*, 1201.
103. B. Ostash; A. Saghatelian; S. Walker, *Chem. Biol.* **2007**, *14*, 257.
104. S. Imai; K. Furihata; Y. Hayakawa; T. Noguchi; H. Seto, *J. Antibiot.* **1989**, *42*, 1196.
105. O. Nakayama; M. Yagi; M. Tanaka; S. Kiyoto; M. Okuhara; M. Kohsaka, *J. Antibiot.* **1989**, *42*, 1221.
106. K. Shinya; K. Furihata; Y. Hayakawa; H. Seto; Y. Kato; J. Clardy, *Tetrahedron Lett.* **1991**, *32*, 943.
107. K. Shinya; Y. Hayakawa; H. Seto, *J. Nat. Prod.* **1993**, *56*, 1255.
108. K. Shinya; S. Shimizu; T. Kunigami; K. Furihata; Y. Hayakawa; H. Seto, *J. Antibiot.* **1995**, *48*, 1378.
109. S. Kato; K. Shindo; Y. Kataoka; Y. Yamagishi; J. Mochizuki, *J. Antibiot.* **1991**, *44*, 903.
110. K. Shinya; M. Tanaka; K. Furihata; Y. Hayakawa; H. Seto, *Tetrahedron Lett.* **1993**, *34*, 4943.
111. K. Shinya; T. Kunigami; J.-S. Kim; K. Furihata; Y. Hayakawa; H. Seto, *Biosci. Biotechnol. Biochem.* **1997**, *61*, 1768.
112. K. Shinya; S. Shimizu; T. Kunigami; K. Furihata; K. Furihata; H. Seto, *J. Antibiot.* **1995**, *48*, 574.
113. N. Orihara; K. Furihata; H. Seto, *J. Antibiot.* **1997**, *50*, 979.
114. M. Takashima; H. Sakai, *Agric. Biol. Chem.* **1960**, *24*, 647.
115. M. Takashima; H. Sakai; K. Arima, *Agric. Biol. Chem.* **1962**, *26*, 660.
116. H. Nakata; H. Harada; Y. Hirata, *Tetrahedron Lett.* **1966**, *23*, 2515.
117. H. Fujiki; M. Mori; M. Nakayasu; M. Terada; T. Sugimura; R. E. Moore, *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 3872.
118. Y. Hitotsuyanagi; H. Fujiki; M. Suganuma; N. Aimi; S. Sakai; Y. Endo; K. Shudo; T. Sugimura, *Chem. Pharm. Bull.* **1984**, *32*, 4233.
119. K. Irie; N. Hagiwara; K. Koshimizu; H. Hayashi; S. Murao; H. Tokuda; Y. Ito, *Agric. Biol. Chem.* **1985**, *49*, 221.
120. N. Hagiwara; K. Irie; K. Koshimizu; H. Hayashi; S. Murao; H. Tokuda; Y. Ito, *Agric. Biol. Chem.* **1985**, *49*, 2529.
121. K. Irie; N. Hagiwara; T. Kurome; H. Hayashi; M. Arai; K. Koshimizu, *Agric. Biol. Chem.* **1987**, *51*, 285.
122. S. Sakai; Y. Hitotsuyanagi; K. Yamaguchi; N. Aimi; K. Ogata; T. Kuramochi; H. Seki; R. Hara; H. Fujiki, *et al.*, *Chem. Pharm. Bull.* **1986**, *34*, 4883.
123. K. Irie; N. Hagiwara; A. Funaki; H. Hayashi; M. Arai; H. Tokuda; K. Koshimizu, *Agric. Biol. Chem.* **1988**, *52*, 3193.
124. K. Irie; A. Funaki; K. Koshimizu; H. Hayashi; M. Arai, *Tetrahedron Lett.* **1989**, *30*, 2113.
125. S. Nishiwaki; H. Fujiki; S. Yoshizawa; M. Suganuma; H. Furuya-Suguri; S. Okabe; M. Nakayasu; K. Okabe; H. Muratake; M. Natsume; K. Umezawa; S. Sakai; T. Sugimura *Jpn. J. Cancer Res.* **1991**, *82*, 779.
126. K. Irie; S. Kajiyama; S. Okuno; M. Kondo; K. Koshimizu; H. Hayashi; M. Arai; H. Nishino; A. Iwashima, *J. Nat. Prod.* **1994**, *57*, 363.
127. K. Irie; S. Tomimatsu; Y. Nakagawa; H. Ohigashi; H. Hayashi, *Biosci. Biotechnol. Biochem.* **1999**, *63*, 1669.
128. K. Nakae; N. Hosokawa; R. Sawa; Y. Kubota; T. Masuda; S. Ohba; M. Igarashi; N. Nakagawa; Y. Nishimura; Y. Akamatsu, *J. Antibiot.* **2006**, *59*, 11.
129. K. Irie; S. Kajiyama; A. Funaki; K. Koshimizu; H. Hayashi; M. Arai, Tennen Yuki Kagobutsu Toronkai Koen Yoshishu. In *Proceedings of the 31st symposium on the chemistry Natural Products,* Nagoya, Japan, 1989, 308.
130. K. Irie; Y. Nakagawa; S. Tomimatsu; H. Ohigashi, *Tetrahedron Lett.* **1998**, *39*, 7929.
131. K. Irie; S. Kajiyama; A. Funaki; K. Koshimizu; H. Hayashi; M. Arai, *Tetrahedron* **1990**, *46*, 2773.
132. S. Omura; A. Nakagawa; H. Hashimoto; R. Oiwa; Y. Iwai; A. Hirano; N. Shibukawa; Y. Kojima, *J. Antibiot.* **1980**, *33*, 1395.
133. A. Nakagawa; Y. Iwai; H. Hashimoto; N. Miyazaki; R. Oiwa; Y. Takahashi; A. Hirano; N. Shibukawa; Y. Kojima; S. Omura, *J. Antibiot.* **1981**, *34*, 1408.
134. S. Omura; A. Nakagawa, *Tetrahedron Lett.* **1981**, *22*, 2199.
135. W.-G. Kim; J.-P. Kim; C.-J. Kim; K.-H. Lee; I.-D. Yoo, *J. Antibiot.* **1996**, *49*, 20.
136. W.-G. Kim; J.-P. Kim; H. Koshino; K. Shin-Ya; H. Seto; I.-D. Yoo, *Tetrahedron* **1997**, *53*, 4309.
137. W.-G. Kim; I.-J. Ryoo; J.-S. Park; I.-D. Yoo, *J. Antibiot.* **2001**, *54*, 513.
138. K. A. Dekker; T. Inagaki; T. D. Gootz; L. H. Huang; Y. Kojima; W. E. Kohlbrenner; Y. Matsunaga; P. R. Mcguirk; E. Nomura; T. Sakakibara; S. Sakemi; Y. Suzuki; Y. Yamauchi; N. Kojima, *J. Antibiot.* **1998**, *51*, 145.
139. T. Sasaki; Y. Igarashi; M. Ogawa; T. Furumai, *J. Antibiot.* **2002**, *55*, 1009.
140. J. Y. Cho; H. C. Kwon; P. G. Williams; P. R. Jensen; W. Fenical, *Org. Lett.* **2006**, *8*, 2471.
141. V. R. Macherla; J. Liu; C. Bellows; S. Teisan; B. Nicholson; K. S. Lam; B. C. M. Potts, *J. Nat. Prod.* **2005**, *68*, 780.
142. Y. Igarashi; S. Miyanaga; H. Onaka; M. Takahashi; T. Furumai, *J. Antibiot.* **2005**, *58*, C-2.
143. J. Wang; S. M. Soisson; K. Young; W. Shoop; S. Kodali; A. Galgoci; R. Painter; G. Parthasarathy; Y. S. Tang; R. Cummings; S. Ha; K. Dorso; M. Motyl; H. Jayasuriya; J. Ondeyka; K. Herath; C. Zhang; L. Hernandez; L. Allocco; A. Basilio; R. Tormo; O. Genilloud; F. Vicente; F. Pelaez; L. Colwell; S. H. Lee; B. Michael; T. Felcetto; C. Gill; L. L. Silver; J. D. Hermes; K. Bartizal; J. Barrett; D. Schmatz; J. W. Becker; D. Cully; S. B. Singh, *Nature* **2006**, *441*, 358.
144. J. Wang; S. Kodali; S. H. Lee; A. Galgoci; R. Painter; K. Dorso; F. Racine; M. Motyl; L. Hernandez; E. Tinney; S. L. Colletti; K. Herath; R. Cummings; O. Salazar; I. González; A. Basilio; F. Vicente; O. Genilloud; F. Pelaez; H. Jayasuriya; K. Young; D. F. Cully; S. B. Singh, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 7612.

145. K. B. Herath; A. B. Attygalle; S. B. Singh, *J. Am. Chem. Soc.* **2007**, *129*, 15422.
146. H. Seto; Y. Jinnai; T. Hiratsuka; M. Fukawa; K. Furihata; N. Itoh; T. Dairi, *J. Am. Chem. Soc.* **2008**, *130*, 5614.
147. T. Hiratsuka; K. Furihata; J. Ishikawa; H. Yamashita; N. Itohi; N. Seto; T. Dairi, *Science* **2008**, *321*, 1670.

Biographical Sketch

Tohru Dairi received his B.S. (1983) and M.S. (1985) from Nagoya University, and later joined Kyowa Hakko Kogyo Co. Ltd. (1985–94). After receiving Ph.D. from The University of Tokyo (1992), he was appointed assistant professor at Toyama Prefectural University and is currently associate professor in the Department of Bioengineering. He has authored and coauthored ~70 publications. The current research on natural products includes studies on biosynthetic genes and enzymes responsible for terpenoids produced by bacteria and fungi and finding a novel pathway for primary metabolism by bioinformatics and biochemical experiments.

1.23 Lignans (Neolignans) and Allyl/Propenyl Phenols: Biogenesis, Structural Biology, and Biological/Human Health Considerations

Daniel G. Vassão, Kye-Won Kim, Laurence B. Davin, and Norman G. Lewis, Washington State University, Pullman, WA, USA

© 2010 Elsevier Ltd. All rights reserved.

1.23.1	**Introduction**	817
1.23.2	**Definition and Nomenclature**	818
1.23.2.1	Allyl-/Propenylphenols	818
1.23.2.2	Inconsistencies in Current Nomenclature of Lignans and Neolignans	818
1.23.3	**Chemotaxonomical Diversity: Evolutionary Considerations**	821
1.23.3.1	Allyl-/Propenylphenols and Their Derivatives	821
1.23.3.1.1	Algae	821
1.23.3.1.2	Bryophytes: liverworts, hornworts, and mosses	825
1.23.3.1.3	Pteridophytes: lycophytes, horsetails, and ferns	826
1.23.3.1.4	Spermatophytes: gymnosperms and angiosperms	826
1.23.3.2	Lignans	830
1.23.3.2.1	Bryophytes: liverworts, hornworts, and mosses	830
1.23.3.2.2	Pteridophytes: lycophytes, horsetails, and ferns	833
1.23.3.2.3	Spermatophytes: gymnosperms and angiosperms	836
1.23.3.3	Evolution of Biochemical Pathways to Allyl-/Propenylphenols and Lignans: Observations on Co-occurrence	845
1.23.4	**Lignan Early Biosynthetic Steps: 8–8′ Phenylpropanoid Coupling**	847
1.23.4.1	Discovery of the (+)-Pinoresinol-Forming Dirigent Protein and Encoding Gene	847
1.23.4.2	Western Red Cedar Dirigent Proteins	849
1.23.4.3	Structural and Mechanistic Studies	849
1.23.4.4	Discovery of the (−)-Pinoresinol-Forming Dirigent Protein and Encoding Gene	849
1.23.4.5	Dirigent Protein Tissue Localization and Metabolic Networks	851
1.23.4.5.1	mRNA tissue localization	851
1.23.4.5.2	Dirigent protein tissue localization and proposed proteins harboring arrays of dirigent sites	852
1.23.4.5.3	Proposed dirigent protein metabolic networks	853
1.23.4.6	Other Examples of 8–8′ Phenylpropanoid Coupling: Hydroxycinnamic Acid and Allyl-/Propenylphenol-Derived Lignans in Liverworts and the Creosote Bush	855
1.23.5	**Downstream Lignan Metabolism**	856
1.23.5.1	Furofuran Lignans in Sesame	856
1.23.5.1.1	Methylenedioxy bridge formation	856
1.23.5.1.2	Glucosylation	859
1.23.5.1.3	Oxygen insertion	860
1.23.5.2	Pinoresinol/Lariciresinol Reductases and Pinoresinol Reductase	860
1.23.5.2.1	Forsythia PLR: discovery of (+)-pinoresinol/lariciresinol reductase	860
1.23.5.2.2	Gymnosperm PLR/PLR homologs: discovery of PLR/PLR homologs of differing PLR and pinoresinol reductase enantiospecificities	862
1.23.5.2.3	*Linum* species PLR: additional discovery of genes encoding (−)-PLR activity	862
1.23.5.2.4	*Arabidopsis* PLR homologs: Pinoresinol reductases	865
1.23.5.2.5	Tissue localization of PLRs and PRs	866
1.23.5.2.6	Structural biology studies: PLR and PLR homolog	867

1.23.5.3	Secoisolariciresinol Dehydrogenase	873
1.23.5.3.1	Discovery of SDH and encoding gene	873
1.23.5.3.2	Structural biology studies	874
1.23.5.4	Creosote Bush Lignan Metabolism: Enantiospecific Polyphenol Oxidase	876
1.23.5.5	Additional (Preliminary) Studies Toward Justicidin B, Hinokinin, and Podophyllotoxin/6-Methoxypodophyllotoxin Biosynthesis	876
1.23.5.5.1	Justicidin B	876
1.23.5.5.2	Hinokinin	876
1.23.5.5.3	Podophyllotoxin/6-Methoxypodophyllotoxin	877
1.23.6	**Other Phenylpropanoid Coupling Modes: 8–2′, 8–3′ (8–5′), and 8–*O*–4′-Linked Lignans**	878
1.23.6.1	8–2′ Coupling	879
1.23.6.2	8–3′ (8–5′) Coupling	879
1.23.6.3	8–*O*–4′ Coupling	880
1.23.7	**Allylic (Phenylpropenal) Double Bond Reductases and Phenylcoumaran Benzylic Ether Reductases**	880
1.23.7.1	Allylic (Phenylpropenal) Double Bond Reductases: Biosynthesis of Dihydrolignans and Dihydromonolignols	880
1.23.7.1.1	Discovery of allylic (phenylpropenal) double bond reductases and gene cloning: Loblolly pine (*Pinus taeda*)	882
1.23.7.1.2	mRNA tissue localization of PtPPDBR in loblolly pine	883
1.23.7.1.3	Allylic double bond reductase homologs: eleven-membered multigene family in *Arabidopsis*	883
1.23.7.1.4	Structural biology studies: *Arabidopsis* DBR1	884
1.23.7.2	PLR Homologs: Phenylcoumaran Benzylic Ether Reductases, Isoflavone Reductases, and Pterocarpan Reductases	887
1.23.7.2.1	Tissue localization	888
1.23.7.2.2	Structural biology studies of PLR homologs: PCBER, IFR, and pterocarpan reductases	888
1.23.8	**Norlignan Biosynthesis**	889
1.23.8.1	Hinokiresinol: Discovery of Biochemical Pathway, Encoding Genes, and Enzymes	889
1.23.8.2	Agatharesinol	891
1.23.8.3	Acetylenic Norlignans	891
1.23.9	**Allyl-/Propenylphenol Biosynthesis**	892
1.23.9.1	Deduction of Allyl-/Propenylphenol (Monomeric and Dimeric) Biosynthetic Pathways	893
1.23.9.1.1	Radiolabel tracer studies: controversy over intact incorporation of monolignol pathway intermediates and scientific judgment?	893
1.23.9.1.2	Intermediacy of monolignol esters in allyl-/propenylphenol biosynthesis: clues from norlignans?	895
1.23.9.2	Allyl-/Propenylphenol Synthases	898
1.23.9.2.1	Bifunctional chavicol/eugenol and *p*-anol/isoeugenol synthases (CES and AIS): The twists and turns to biochemical clarity	898
1.23.9.2.2	Chemotaxonomy, kinetic properties, and homology comparisons of CES/AIS with PCBER, PLR, IFR (-like) annotations in the plant kingdom: caveats on incomplete analyses	900
1.23.9.2.3	CES (AIS) structural and mechanistic studies: comparison to PLRs, PCBERs, and IFRs	903
1.23.9.2.4	Allyl-/propenylphenol downstream metabolism	904
1.23.9.3	Monolignol Acyltransferases: Incomplete Characterization and Substrate Degeneracy	905
1.23.10	**Biological Properties *in Planta* and in Human Usage**	906
1.23.10.1	Allyl-/Propenylphenols	907
1.23.10.1.1	Antimicrobial properties	907
1.23.10.1.2	Anesthetic properties	908

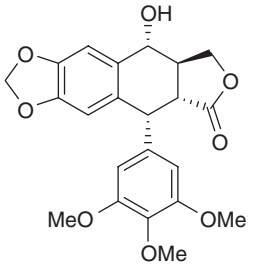
1.23.10.1.3	Other reported activities	908
1.23.10.1.4	Effects *in planta*	908
1.23.10.1.5	Mutagenicity	908
1.23.10.1.6	Potential future uses as commodity chemicals/biofuels	909
1.23.10.2	Lignans	909
1.23.10.2.1	Lignans in cancer chemotherapy and cancer prevention	909
1.23.10.2.2	Antiviral lignans	913
1.23.10.2.3	Nutraceutical lignans: sesame	913
1.23.10.2.4	Antichagasic lignans	915
1.23.10.2.5	Properties *in planta*	915
References		916

1.23.1 Introduction

Allyl-/propenylphenols are generally volatile substances intimately associated with essential oils of scents and flavors of flowers, herbs, and spices. Many are of great historical importance, for example, as spice constituents. Lignans (and neolignans) are also very important natural products of vast structural diversity, these being apparently found ubiquitously throughout the vascular plant kingdom. Many have roles in plant defense against herbivores/pathogens, and many contribute positively to desirable heartwood properties of various woody species, such as in enhancing durability, color, texture, and so on. Others are beneficial in foodstuffs and oils, such as in sesame, due in part to their antioxidant properties. Some are also medicinal, such as either the antiviral *Podophyllum* lignan, (−)-podophyllotoxin (**1b**, **Figure 1**), or its related derivatives widely used in conventional cancer chemotherapy.

The primary focus herein is on the advances made in dimeric/oligomeric lignan (neolignan) and allyl-/propenylphenol research since 1999.[1] Specifically, various developments made over the past decade are critically discussed, with regard to both metabolic pathways and the associated structural biology, as well as what is known of their corresponding biological/health-protecting properties (both established and potential). The addition of allyl-/propenylphenols is timely as the corresponding biosynthetic pathways are only now coming to light.[2–5] By contrast, the earlier treatise[1] in *Comprehensive Natural Products Chemistry* placed substantial emphasis on the pioneering work carried out in developing the lignan field, as well as giving a critical analysis of the accumulated knowledge that has made the more recent studies possible.

Podophyllum peltatum



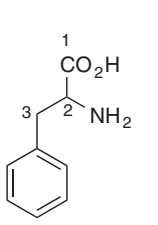
(**1b**) (–)-Podophyllotoxin

Figure 1 *Podophyllum peltatum* and (–)-podophyllotoxin (**1b**). The letters a and b in compound numbers throguhout the manuscript depict (+)- and (–)-enantiomeric forms, respectively. Photograph from Laurence Davin, Washington State University, USA.

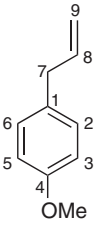
1.23.2 Definition and Nomenclature

1.23.2.1 Allyl-/Propenylphenols

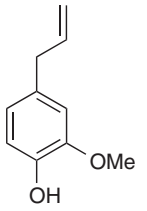
Allylphenols and propenylphenols are metabolites found in many plant species, whose core (C_6C_3) carbon skeleta comprise a C_3 side chain attached to an aromatic ring. They are products of the phenylpropanoid pathway (i.e., the so-called C_6C_3 natural products, which, in turn, are derived from the amino acid Phe (**2**)). They differ from most other phenylpropanoids (e.g., monolignols, hydroxycinnamaldehydes, hydroxycinnamic acids, and their downstream products) by lacking an oxygenated functionality at the terminal carbon (C9) of the side-chain. Additionally, they differ among themselves in the placement and/or absence of the double bond. The allylphenols have a C8–C9 terminal double bond (e.g., methylchavicol (**3**, estragole or 4-allylanisole) and eugenol (**4**)), whereas propenylphenols have it at C7–C8, conjugated with the aromatic ring (e.g., anethole (**5**) and isoeugenol (**6**)). In general, the carbon skeleton is intactly maintained as that originating from Phe (**2**),[2] except for the so-called pseudoisoeugenols (e.g., **7** or **8**). The latter results from a presumed NIH shift of the side-chain,[6] with some pseudoisoeugenols undergoing further functional modification such as epoxide formation.



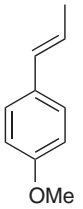
(**2**) Phenylalanine



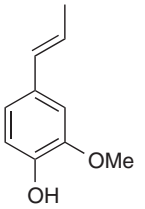
(**3**) Methylchavicol



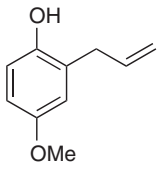
(**4**) Eugenol



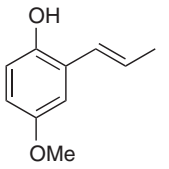
(**5**) Anethole



(**6**) Isoeugenol



(**7**) 2-Allyl-4-methoxyphenol

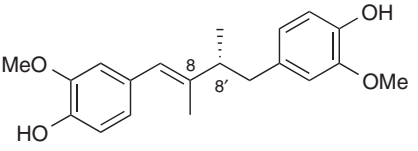


(**8**) Pseudoisoeugenol

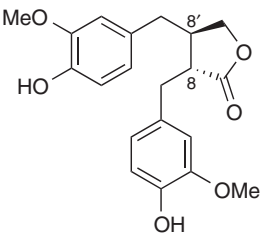
1.23.2.2 Inconsistencies in Current Nomenclature of Lignans and Neolignans

The structurally diverse lignans (and neolignans) encompass an extensive array of distinct carbon skeleta.[1] All are presumed derived from coupling of phenylpropanoid (C_6C_3) units to afford dimers or higher oligomers, linked through C–C and/or C–O interunit bonds. In contrast to the more abundant (by weight) polymeric lignins,[7–9] the lignans (and neolignans) do not have any known essential cell wall structural biopolymeric role(s) in the vascular apparatus.

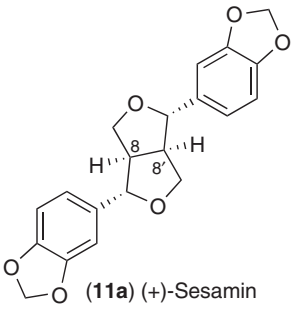
Since the late nineteenth or early twentieth century, numerous dimeric/oligomeric phenylpropanoids have been isolated from various plant species. A number of these were given trivial names indicative of botanical origin, for example, (−)-guaiaretic acid (**9b**) from *Guaiacum officinale* resin,[10–12] (−)-matairesinol (**10b**) from the Matai tree (*Podocarpus spicatus*),[13] (+)-sesamin (**11a**) from *Sesamum indicum*,[14,15] and (−)-olivil (**12b**) from the olive tree (*Olea europaea*).[16–18]



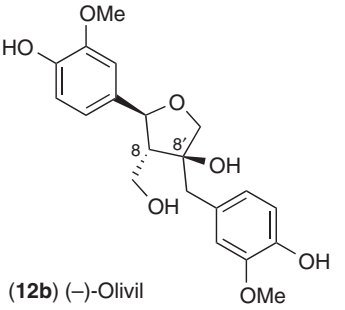
(**9b**) (–)-Guaiaretic acid



(**10b**) (–)-Matairesinol

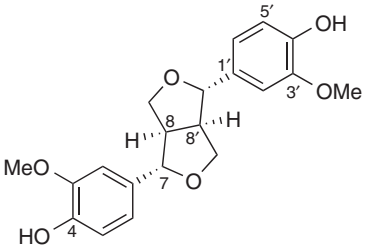


(**11a**) (+)-Sesamin

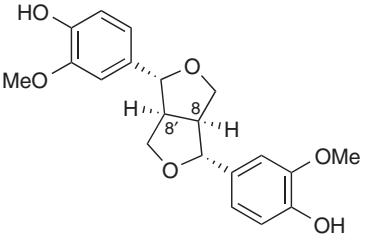


(**12b**) (–)-Olivil

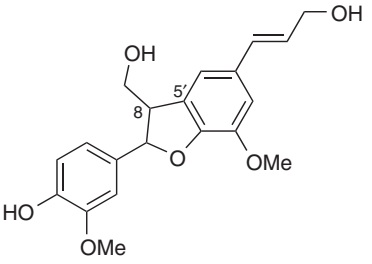
In 1937, Haworth[19] introduced the term lignane, later shortened to lignan, to define various dimeric phenylpropanoids that were considered to result from regiospecific (8–8′) coupling of two C_6C_3 molecules to afford different compounds such as (+)- and (−)-pinoresinols (**13a** and **13b**). A large number of other dimeric phenylpropanoids (and/or higher oligomers thereof) could not be classified according to this definition, and the term neolignan was thus introduced later by Gottlieb[20] to encompass other coupling modes, for example, the 8–5′-linked (±)-dehydrodiconiferyl alcohols (**14a/b**). This neolignan classification was subsequently further modified[21] to next include only phenylpropanoid dimers presumably derived from coupling of allyl-/propenylphenols (i.e., phenylpropanoid monomers lacking C9 oxygenation), such as magnolol (**15**) and (+)-conocarpan (**16a**). Curiously, current IUPAC lignan nomenclature[22] still utilizes the original definitions of both Haworth and Gottlieb in attempts to distinguish between lignans and neolignans, that is, based solely on the differences in interunit linkage type. Using such nomenclature, the lignans would thus be exclusively 8–8′-linked, whereas the neolignans would encompass all other linkage types. To add further to this complexity, the term oxyneolignan was also coined to define ether-linked structures, for example, isomagnolol (**17**).[23]



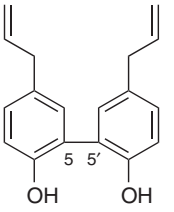
(**13a**) (+)-Pinoresinol



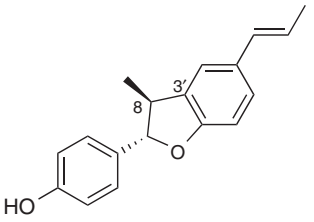
(**13b**) (–)-Pinoresinol



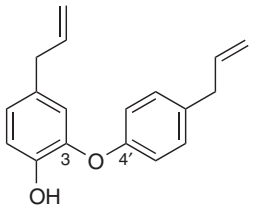
(**14a,b**) (±)-Dehydrodiconiferyl alcohols



(**15**) Magnolol



(**16a**) (+)-Conocarpan

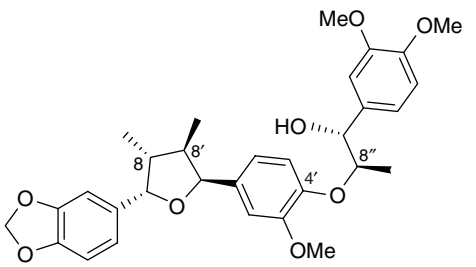


(**17**) Isomagnolol

In our view, however, the continued usage of the terms lignans, neolignans, and oxyneolignans results in serious nomenclature inconsistencies. Two examples serve to illustrate this point (**Figure 2**). The sesquilignan (−)-saucerneol D (**18b**) from *Saururus chinensis*[24] and the dilignan (+)-lappaol F (**19a**) from *Arctium lappa*[25,26] contain both lignan and neolignan interunit linkages according to the definitions of Haworth[19] and Gottlieb,[20] and are thus neither one nor the other.

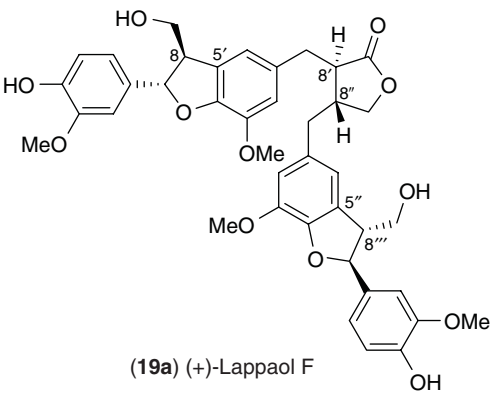
Therefore, it is continued[1] to propose that the original definition of 'lignan' be expanded to encompass all of the lignans, neolignans, and oxyneolignans, regardless of presumed monomeric composition (i.e., irrespective of whether the functionalities present are apparently either monolignol, hydroxycinnamaldehyde, hydroxycinnamic acid, or allyl-/propenylphenol-like). In this way, all of these natural products can be more conveniently characterized as lignans that would differ only through linkage type(s).

Saururus chinensis



(**18b**) (–)-Saucerneol D

Arctium lappa



(**19a**) (+)-Lappaol F

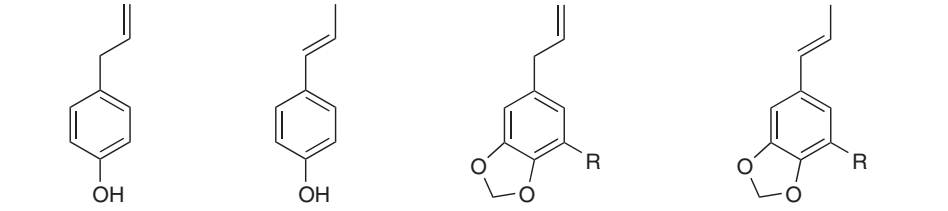
Figure 2 Sesquilignans. (−)-Saucerneol D (**18b**) from *Saururus chinensis* and (+)-lappaol F (**19a**) from *Arctium lappa*. Photograph of *S. chinensis* by Henri Moore, Washington State University USA. Photograph of *A. lappa* by Dr. Toshiaki Umezawa, Kyoto University, Japan.

1.23.3 Chemotaxonomical Diversity: Evolutionary Considerations

To provide an evolutionary and chemotaxonomical perspective of the occurrence of allyl-/propenylphenols and lignans, the three major land plant groups[27,28] are considered individually, as well as algae. The former include the bryophytes[27] (nonvascular land plants, including liverworts, hornworts, and mosses) (**Figure 3**); the pteridophytes[27] (seedless vascular land plants, i.e., lycophytes, horsetails, and all ferns) (**Figure 3**); and the spermatophytes[28] (seed-bearing vascular land plants: gymnosperms (cycads, conifers, ginkgos, and gnetophytes) and angiosperms, **Figure 4**).

1.23.3.1 Allyl-/Propenylphenols and Their Derivatives

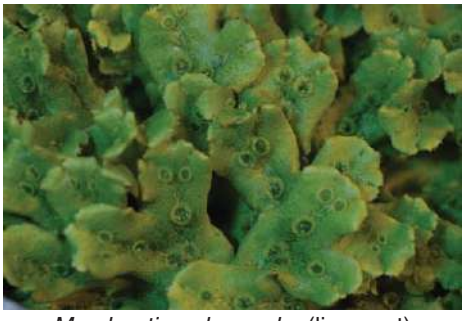
Allyl-/propenylphenols appear to have a fairly broad, albeit not uniform, chemotaxonomic distribution throughout the plant kingdom. In terms of structural diversity, the most commonly reported aromatic ring substitution patterns are those of hydroxylation/methoxylation. To our knowledge, there are no reports of any other heteroatoms (e.g., N, S) being present in allyl-/propenylphenols from plants. The C4 position is typically oxygenated, with further hydroxylation/methoxylation most frequently (but not exclusively) occurring at *ortho* positions (C3/C5) to the C4 oxygenated group. The most common plant allyl-/propenylphenols appear to be eugenol (**4**)/isoeugenol (**6**) and the *O*-methylated derivatives of chavicol (**20**)/*p*-anol (**21**), that is, methylchavicol (**3**)/anethole (**5**), based on current literature searches. Other aromatic ring modifications often reported include either the formation of methylenedioxy bridges, as in safrole (**22**)/isosafrole (**24**) and myristicin (**23**)/isomyristicin (**25**), or less commonly other hydroxylation/methoxylation substitution patterns, such as at C2 and C6 (*ortho* to the C_3 side-chain). These can be exemplified by allyl/propenyl-2,4,6-trimethoxybenzenes (**26**/**27**), 2-allyl-4,5-dimethoxyphenol (**28**), asarones (**29–31**), asaricin (**32**)/carpacin (**33**), croweacin (**34**)/isocroweacin (**35**), apiol (**36**)/isoapiol (**37**), dillapiol (**38**)/isodillapiol (**39**), and nothoapiol (**40**)/isonothoapiol (**41**). As previously mentioned above, the pseudoisoeugenols in *Pimpinella* spp. lack the *p*-oxygenation observed in these compound classes, having instead a 2,5-dioxy substitution as a consequence of side-chain migration,[6] for example, 2-allyl-4-methoxyphenol (**7**) and pseudoisoeugenol (**8**).



(**20**) R = H, Chavicol (**21**) R = H, *p*-Anol (**22**) R = H, Safrole (**23**) R = OMe, Myristicin (**24**) R = H, Isosafrole (**25**) R = OMe, Isomyristicin

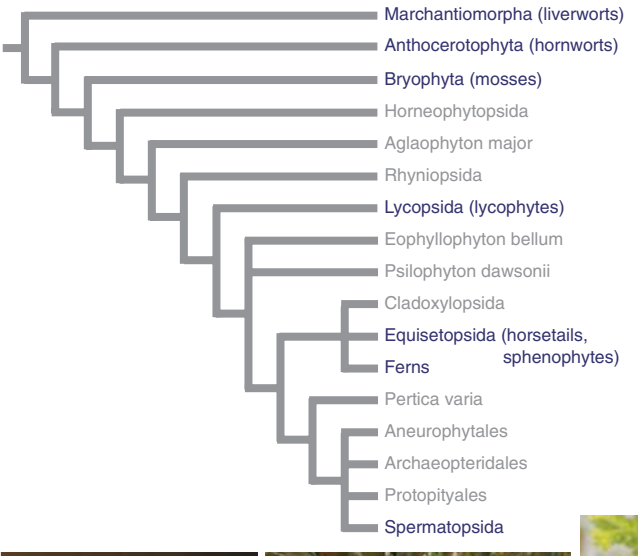
1.23.3.1.1 Algae

Sporadic, albeit somewhat preliminary, reports have been made of allyl-/propenylphenol occurrence in brown (*Spatoglossum variabile*),[29] red (*Jania rubens*[30] and *Hypnea musciformis*[31]), as well as green (*Enteromorpha compressa*,[31] *Caulerpa racemosa*,[32] and *Codium tomentosum*[32]) algae. The reported occurrence of very small amounts of apiol (**36**) and nothoapiol (**40**) in the brown alga (*S. variabile*)[29] growing on shoreline sublittoral rock near Karachi, Pakistan, would benefit from additional biochemical proof, preferably through the analysis of this alga grown under controlled laboratory growth conditions. Although the chemical identification in the study[29] is not in doubt, the possibility that apiol (**36**) and nothoapiol (**40**) may have been absorbed from other detritus at the shoreline or in the waters, or might have resulted from contaminant plant material being collected and analyzed at the same time, needs to be fully eliminated. A similar consideration can be given to preliminary reports of volatile constituents containing traces of eugenol (**4**) and isoeugenol (**6**) in the green (*E. compressa*) and red (*H. musciformis*) algae,[31] as well as that of anethole (**5**) in two green algae (*C. racemosa* and *C. tomentosum*)[32] and a red alga (*J. rubens*).[30]



Marchantia polymorpha (liverwort)

Anthoceros and Phaeoceros (hornworts)



Dicranum (moss)

Pteridium aquilinum (fern)

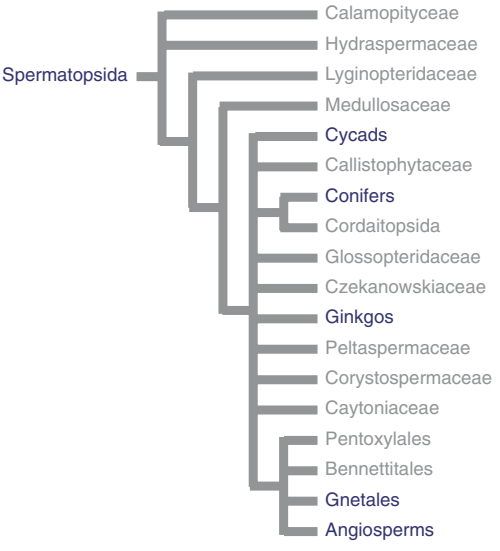
Equisetum hyemale (horsetail)

Selaginella kraussiana (lycophyte)

Figure 3 The tree of life: Embryophytes.[27] Extinct plant lineages are shown in gray, while extant lineages are in blue. Photographs of *Marchantia polymorpha, Dicranum* sp., *Selaginella kraussiana*, *Equisetum hyemale,* and *Pteridium aquilinum* by Henri Moore, Washington State University, USA. Photograph of *Anthoceros* and *Phaeoceros*, copyright Charles F. Delwiche, University of Maryland, USA.

Cycas revoluta (Cycad)

Araucaria araucana (Conifer)

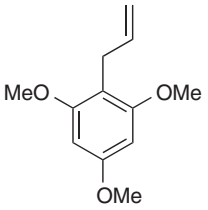


Ginkgo biloba

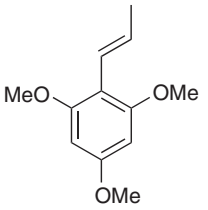
Larrea tridentata (Angiosperm)

Gnetum gnemon (Gnetales)

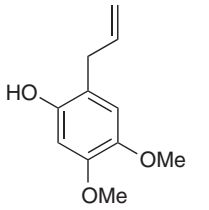
Figure 4 The tree of life: Spermatopsida.[28] Extinct plant lineages are shown in gray, while extant lineages are in blue. Photographs of *Cycas revoluta*, *Gnetum gnemon*, and *Larrea tridentata* by Henri Moore, Washington State University, USA, of *Araucaria araucana* (monkey puzzle tree) by Laurence Davin, Washington State University, USA, and of *Ginkgo biloba* by Daniel Vassãc, Washington State University, USA.



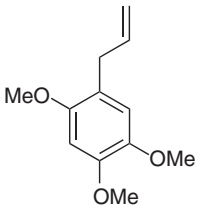
(**26**) Allyl-2,4,6-trimethoxybenzene



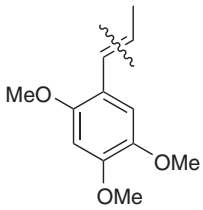
(**27**) Propenyl-2,4,6-trimethoxybenzene



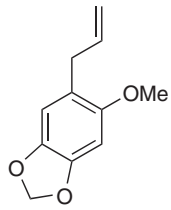
(**28**) 2-Allyl-4,5-dimethoxyphenol



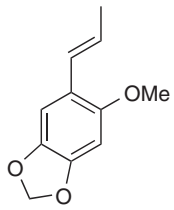
(**29**) γ-Asarone



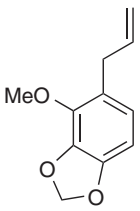
(**30**) α-Asarone (*trans*)
(**31**) β-Asarone (*cis*)



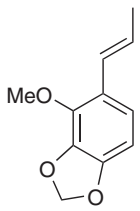
(**32**) Asaricin



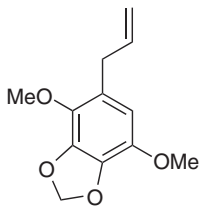
(**33**) Carpacin



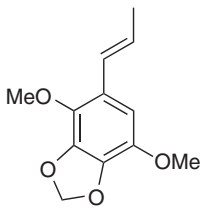
(**34**) Croweacin



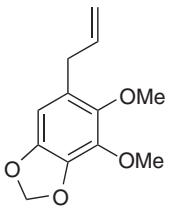
(**35**) Isocroweacin



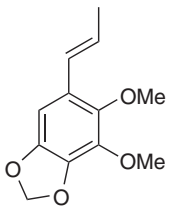
(**36**) Apiol



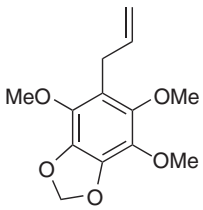
(**37**) Isoapiol



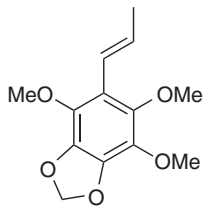
(**38**) Dillapiol



(**39**) Isodillapiol



(**40**) Nothoapiol



(**41**) Isonothoapiol

Their formation by algae is both of interest and of question because, to our knowledge, there is yet no compelling evidence of either algal genes or protein/enzymes encoding the full phenylpropanoid pathway to the monolignols and/or allyl-/propenylphenols as noted earlier (but see Chapter 6.01).[7] Currently, only phenylalanine ammonia lyase activity, the entry point in the phenylpropanoid pathway, has been described in one alga.[33] Furthermore, an *in silico* search for putative phenylpropanoid pathway genes in the recently sequenced green alga, *Chlamydomonas reinhardtii*,[34] was considered unsuccessful, that is, for those genes with homology to the corresponding gymnosperm/angiosperm genes, only fragments of low homology/identity were obtained (S.-J. Kim, unpublished results). This again, suggests, at least in this particular case, that the phenylpropanoid pathway was absent. Thus, if these allyl-/propenylphenol metabolites are indeed formed by such organisms (e.g., *C. racemosa*, *C. tomentosum*, *E. compressa*, *H. musciformis, J. rubens*, and *S. variabile*), this would be of considerable scientific and evolutionary significance. Accordingly, this needs to be confirmed further through more in-depth biochemical studies under controlled laboratory conditions.

1.23.3.1.2 Bryophytes: liverworts, hornworts, and mosses

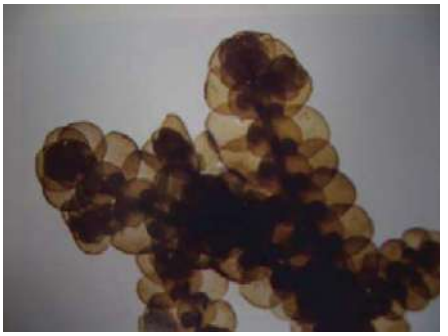
The liverworts (Marchantiophyta) are typically considered to be the most basal of the nonvascular land plants. They have the oldest fossil record among the bryophytes (see **Figure 3**) and apparently date back more than 475 My.[35] They are estimated to contain ~4500–5000 extant species.[36,37] By contrast, hornworts (Anthocerotophyta) contain the fewest species among the bryophytes (~300),[38] and their evolutionary relationship to vascular land plants appears to be uncertain. The hypothesis that hornworts are more basal than liverworts has been supported[39] as well as rejected[40] in recent years; their oldest fossil record is from the late Cretaceous (~70–90 Ma), long after when they are considered to have appeared.[41] The phytochemistry of hornworts is, however, largely unexplored. Mosses, on the contrary, are thought to have emerged up to ~360 Ma (early Carboniferous)[42] and are considered the largest group of bryophytes, totaling ~10 000 species.[43]

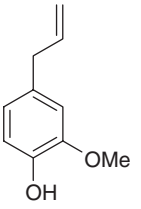
To our knowledge, the only few reports of monomeric allyl-/propenylphenols to date in bryophytes are from liverworts, and mostly come from a series of papers from the Asakawa laboratory from 1979 to 1985.[44–47] These studies described the GC–MS identification of eugenol (**4**) and/or methyleugenol (**42**) in five liverwort species, *Anthelia julacea*, *Conocephalum conicum*, *Frullania davurica*, *Marchesinia brachiata*, and *Marchesinia mackaii* (**Figure 5**). Additionally, γ-asarone (**29**) was reported in an unidentified Jamaican liverwort,[48] and, more recently, α-/β-/γ-asarones (**30/31/29**), (*E*/*Z*)-methylisoeugenols (**43/44**), and methyleugenol (**42**) were described in *M. mackaii* as well.[49] Interestingly, in the latter case, both allylphenols (**29**,**42**) and propenylphenols (**30**,**31**,**43**,**44**) were observed, with propenylphenols apparently being relatively more abundant.

Some of the allylphenols were also reported as being major volatiles in the liverwort species investigated, for example, methyleugenol (**42**) in *M. mackaii*,[45] and were thus proposed to be able to possibly serve as

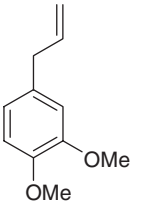
Anthelia julacea

Conocephalum conicum

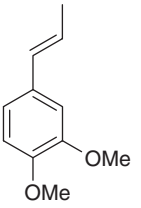

Frullania davurica



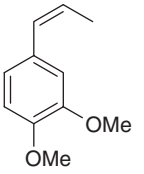
(**4**) Eugenol



(**42**) Methyleugenol



(**43**) (*E*)-Methylisoeugenol



(**44**) (*Z*)-Methylisoeugenol

Figure 5 Allyl-/propenylphenols in liverworts, *Anthelia julacea, Conocephalum conicum,* and *Frullania davurica*. Photographs from Professor Yoshinori Asakawa, Tokushima Bunri University, Japan.

chemotaxonomic markers.[47] These reports are of interest as they suggest a very early (evolutionary) elaboration of both the phenylpropanoid pathway and the allyl-/propenylphenol-forming enzymatic machinery in land plants. Conversely, neither hornworts nor mosses have been reported to accumulate monomeric allyl-/propenylphenols.

1.23.3.1.3 Pteridophytes: lycophytes, horsetails, and ferns

There are very few reports of monomeric allyl-/propenylphenols so far in other (vascular) basal plants. They apparently have not been detected in either lycophytes (the earliest extant vascular plant species, which are thought to have emerged ~420 Ma, in the late Silurian)[50] or ferns (see **Figure 3**). Anethole (**5**) has, however, been reported to occur in trace amounts (as identified by its retention index, mass spectrum, and coinjection with an authentic standard) in horsetail (*Equisetum arvense*, Equisetopsida).[51] As for the algae, this needs to be further biochemically established, that is, as to whether this is a true natural product formed by this species *in vivo* or a minor contaminant.

1.23.3.1.4 Spermatophytes: gymnosperms and angiosperms

1.23.3.1.4(i) Gymnosperms Cycads are often considered to be the most basal extant group of seed plants, with a fossil record dating back to the late Carboniferous (~300 Ma, **Figure 4**); nevertheless, the true evolutionary relationship between gymnosperm groups is at present still not certain.[52–54] There are ~300 species of cycads, distributed mainly through tropical and subtropical areas,[55] but their phytochemistry is not well explored. Nevertheless, the presence of both allyl- and propenylphenol derivatives in the volatiles from male and female cones of one cycad, *Cycas revoluta*,[56] appears to be on a firmer footing relative to the algae and horsetail reports above. Based on GC–MS analysis, together with comparison of the corresponding authentic standards, methylchavicol (**3**) was established as the primary component (~67–93%) of its volatiles, together with smaller amounts of anethole (**5**) and methyleugenol (**42**). To date, however, this is the only cycad known reportedly producing this class of natural products.

Conifers (Pinophyta) are thought to have emerged as early as the late Carboniferous (~310 Ma),[57,58] and now comprise ~630 extant species.[59] An initial review of the literature[8] suggested the absence of allyl-/propenylphenols proper in pine and hemlock (as well as poplar), except for numerous accounts of eugenol (**4**)/isoeugenol (**6**) being detected following pyrolysis GC–MS of presumed lignin-containing conifer tissues, for example, see Faix *et al.*[60] and Camarero *et al.*[61] Such moieties are considered generated by the pyrolysis of monolignol-derived metabolites, including polymeric lignins.

A recent, more in-depth, analysis using the specific structures for searching rather than compound names gave a somewhat different perspective. This resulted in the identification of a relatively small number of papers reporting complex mixtures of various essential oils and oleoresins obtained from either needles, bark, wood, or wounded stem tissue of nine pine (*Pinus*),[62–72] one spruce (*Picea*),[65] one Tasmanian conifer (*Lagarostrobos franklinii*),[73] and three juniper (*Juniperus*)[74] species. Buried within the tables in these studies, in addition to the plethora of mono-, sesqui-, and diterpenes, various allyl-/propenylphenols were detected in low to trace amounts, for example, methylchavicol (**3**) in seven of nine pine species examined,[62–64,66–70,72] with *Pinus taeda*[69] and *Pinus ponderosa*[72] also having methyleugenol (**42**). Three of these pines (*P. brutia*,[71] *P. contorta*,[62] and *P. sylvestris*[65]) as well as spruce (*Picea abies*)[65] reportedly also contained anethole (**5**), with methyleugenol (**42**) and (*E*)-methylisoeugenol (**43**) being detected in *P. brutia*[71] as well. In addition, the oil of *L. franklinii* wood was shown to contain **42**, **43**, and elemicin (**45**).[73] Anethole (**5**) was also detected in *J. brevifolia*, and (*E*)-methylisoeugenol (**43**) in *Juniperus formosana* and *Juniperus rigida*, whereas seven other juniper species apparently contained no allyl-/propenylphenols.[74] These data would thus suggest that both allyl- and propenylphenols can be present in these organisms. However, when the essential oil yields were taken into account, the overall amounts of these allyl-/propenylphenols were very low, that is, ranging from 0.002 to 0.06% of dry weight.

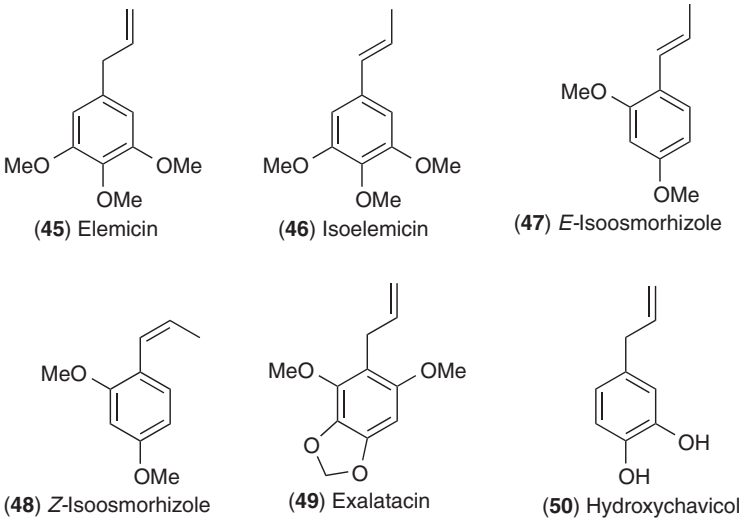
The detection of these substances, at trace to very low levels in the essential oil fractions, raises some interesting questions as to whether they are true gymnospermous natural products or, albeit less likely, result from other effects of an encroaching predator/pathogen, for example, through insect or pathogen attack with conversion of plant-derived monolignol-/lignin-derived components, resulting in release of trace amounts of

such substances. Such reports emphasize the need, however, for a full biochemical clarification to better understand the true chemotaxonomical significance of such observations.

Currently, there are no reports of allyl-/propenylphenols in either the Gnetales (which are thought to have emerged ~270 Ma)[75] or in *Ginkgo biloba* (the single remaining extant species of the Ginkgoaceae family, which is considered to have emerged at least ~170 Ma, see **Figure 4**).[76] Taken together, the evolutionary significance of this apparently scattered chemotaxonomy in the gymnosperms is, at present, not well understood.

1.23.3.1.4(ii) Angiosperms Angiosperms are the most widespread and diverse group of plants, having emerged ~125 Ma[77] (upper Jurassic, **Figure 4**) and are considered to comprise ~260 000 extant species.[78] They are divided into two major groups based on the number of seed embryonic leaves (cotyledons): the monocotyledons (~50–60 000 species, which contain most of the agriculturally cultivated plants) and the more abundant dicotyledons (~200 000 species). With a larger and better-studied number of species, the most frequent report of allyl-/propenylphenols and their derivatives are in the angiosperms. This includes the presence of allylphenols in various monocotyledons, such as eugenol (**4**)/methyleugenol (**42**) in *Lolium perenne* and *Bromus hordeaceus* (Poaceae) essential oils,[79] methyleugenol (**42**), 2-allyl-4,5-dimethoxyphenol (**28**), and γ-asarone (**29**) in the fragrance of the fruit fly orchid *Bulbophyllum cheiri*.[80] Other angiosperm orders where allyl-/propenylphenols have frequently been detected include the Apiales, Asterales, Lamiales, Laurales, Magnoliales, Malpighiales, Myrtales, Piperales, Rosales, and Sapindales. Allyl-/propenylphenols are often substantial contributors to the flavors of culinary spices, adding considerably to their taste/aroma; methylchavicol (**3**) and methyleugenol (**42**) are, for example, present as main components in the essential oil of tarragon (*Artemisia dracunculus*)[81,82] in the Asteraceae, whereas eugenol (**4**) constitutes up to 89% of the oil of cloves (*Syzygium aromaticum*, syn. *Eugenia aromaticum* or *Eugenia caryophyllata*, of the Myrtaceae) (**Figure 6**).[83–86]

The Piperaceae also accumulate a rich variety of specialized metabolites, including both allyl- and propenylphenols in several different chemotypes of *Piper marginatum*. These chemotypes reportedly accumulate as their main allyl-/propenylphenols either safrole (**22**), myristicin (**23**), (*E*)-isoosmorhizole (**47**), (*E*/*Z*)-isoosmorhizole (**47**/**48**), and anethole (**5**), or (α/β)-asarones (**30**/**31**) and exalatacin (**49**),[87] whereas the relatively uncommon hydroxychavicol (**50**) was found in *Piper betel*.[88] Myristicin (**23**) and elemicin (**45**) are also the main phenolic component of the oils of nutmeg (**Figure 6**) and mace (the seed and aril of fruits from *Myristica* spp., respectively), this being accompanied by methyleugenol (**42**), safrole (**22**), and (*E*)-methylisoeugenol (**43**).[89–93] Interestingly, although present in much smaller amounts, isoeugenol (**6**) and a methoxylated eugenol (of undetermined position) were also observed as glycosides.[94]



The allyl-/propenylphenols are perhaps best known, however, to accumulate (to various levels) in the herbs of the Apiaceae. This includes anethole (**5**) and methylchavicol (**3**) in fennel (*Foeniculum vulgare*)[95] and anise (*Pimpinella anisum*),[96] to more highly substituted metabolites such as isoelemicin (**46**) and asaricin (**32**) in

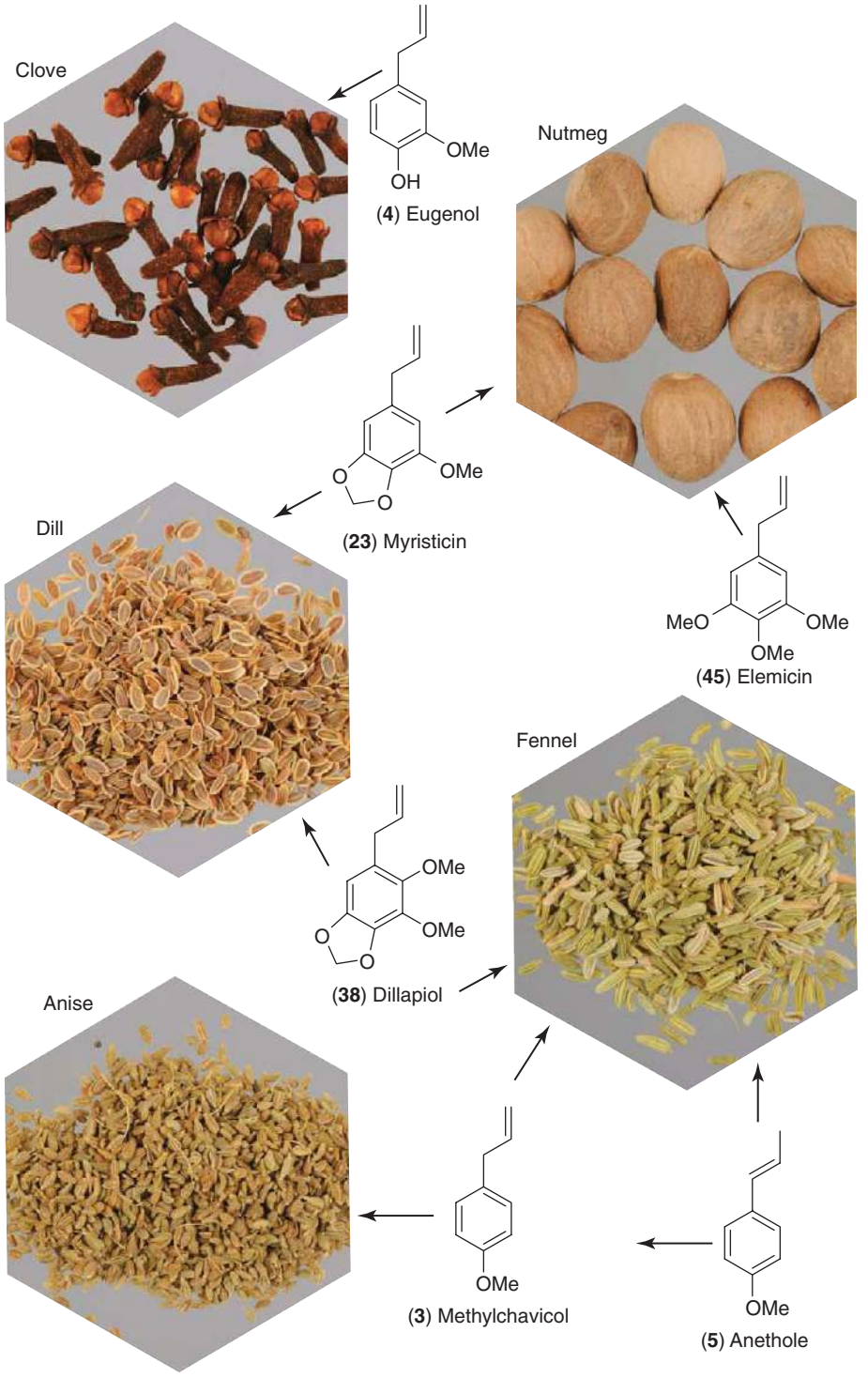


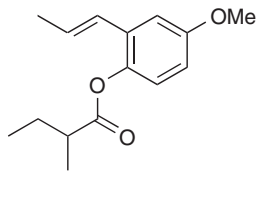
Figure 6 Aroma-contributing molecules in various spices. Photographs by Henri Moore, Washington State University, USA.

Ligusticum mutellina;[97] to myristicin (**23**) in dill (*Anethum graveolens*)[98] and parsley (*Petroselinum crispum*);[99] to apiol (**36**) in parsley[99] and celery (*Apium graveolens*);[100] and to dillapiol (**38**) in dill[98] and fennel[101] (**Figure 6**). The Lamiaceae species produce some of these natural products as well; for example, basil (*Ocimum basilicum*)

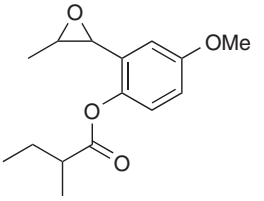
varieties accumulate methylchavicol (**3**), eugenol (**4**), and methyleugenol (**42**),[102,103] and different varieties of *Perilla frutescens* have been found to produce, among other compounds, methyleugenol (**42**), elemicin (**45**), isoelemicin (**46**), myristicin (**23**), dillapiol (**38**), and β-asarone (**31**).[104]

The pseudoisoeugenols, by contrast, appear to have a much more restricted chemotaxonomic distribution. They have been found mainly in the order Apiales, that is, in 22 *Pimpinella* species,[105,106] as well as in *Ferula szowitsiana*,[107] *Prangos pabularia*,[108] *Scaligeria tripartita*,[109] and *Tordylium ketenoglui*,[110] with *P. anisum* being the most extensively studied. Pseudoisoeugenols have also been reported to occur in *Hypericum perforatum*[111] (Clusiaceae) and *Origanum* × *adanense*[112] (Lamiaceae). Of the species examined so far, the most common pseudoisoeugenols reported are pseudoisoeugenyl-2-methylbutyrate (**51**), epoxypseudoisoeugenyl 2-methylbutyrate (**52**, **52a**, EPB), and epoxypseudoisoeugenyl tiglate (**53a/b**), respectively.

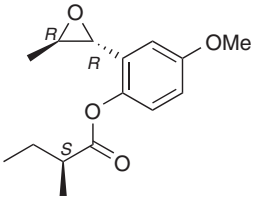
The formation of the epoxide functionality (e.g., in epoxypseudoisoeugenols **52–54**) gives rise to two stereochemical centers, which are, in most reports, not determined in terms of absolute configuration. Nevertheless, based on the relatively small number of structures with assigned configurations, the stereochemistry of the side-chain epoxide formation has been reported as *R,R* (e.g., **52a**), *S,S* (e.g., **53a**), and *S,R* (e.g., **53b**).[113,114] Interestingly, more than one stereoisomer has also been observed in some plant species. (For example, *P. major* reportedly contains both *S,S*- and *S,R*-epoxypseudoisoeugenyl tiglate (**53a** and **53b**), as well as *R,R*-epoxypseudoisoeugenyl 2-methylbutyrate (**52a**) in its roots, whereas *Pimpinella diversifolia* contains both the *R,R* and *S,S* epoxides of diesterified pseudoisoeugenols (**54a** and **54b**)).[113,114]



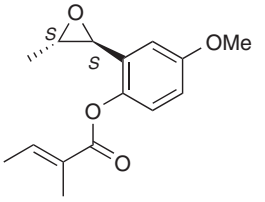
(**51**) Pseudoisoeugenyl 2-methylbutyrate



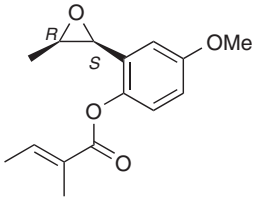
(**52**) Epoxypseudoisoeugenyl 2-methylbutyrate (EPB)



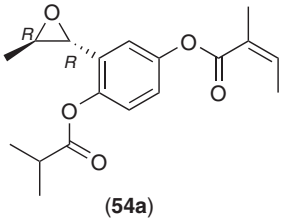
(**52a**) *trans*-Epoxypseudo-isoeugenyl 2-methylbutyrate



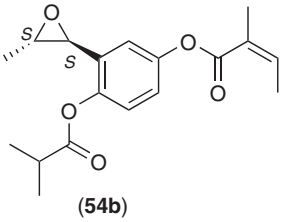
(**53a**) *trans*-Epoxypseudo-isoeugenyl tiglate



(**53b**) *cis*-Epoxypseudo-isoeugenyl tiglate



(**54a**)



(**54b**)

Taken together, the detection of allyl-/propenylphenols in the breadth of plant species examined so far raises interesting questions about how these pathways evolved, and whether these are examples of either convergent or divergent evolution (or both). If the provisional reports of algal allyl-/propenylphenols are correct, then elaboration of their biosynthetic pathways potentially preceded land plant colonization, with this being either sustained or reemerging in the liverworts and the other vascular plant species. However, this needs to be unequivocally established. The apparent lack of uniformity in allyl-/propenylphenol formation (and in known general phenylpropanoid metabolism as well) through extant plant lineages, though, might suggest a

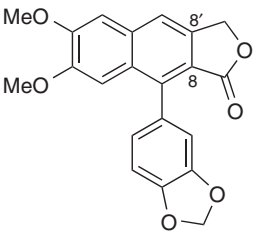
convergent evolutive process as perhaps more likely. That is, the formation of allyl-/propenylphenols might have appeared several times through plant evolution. These processes will only be better understood when the putative allyl-/propenylphenolic biosynthetic processes of algae, as well as those of nonvascular and early vascular plants are better studied, including that of phenylpropanoid pathway gene evolution.

1.23.3.2 Lignans

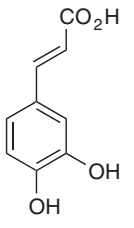
Lignans are known to be present in a large number of plant families, ranging from nonvascular plants such as the liverworts and hornworts, to ferns (in the pteridophytes), to the spermatophytes, but not, to our knowledge, in either the green algae or mosses. As discussed above (Section 1.23.3.1), with the exception of the entry step (phenylalanine ammonia lyase),[33] neither the enzymes nor the encoding genes of the phenylpropanoid pathway have been reported in algae. The sequenced genome of the moss, *Physcomitrella patens*,[115] on the contrary, appears to have genes of relatively high similarity (i.e., from 67 to ~86%) to most (if not all) of the corresponding *Arabidopsis thaliana* genes of the phenylpropanoid pathway proper,[8] from phenylalanine ammonia lyase to cinnamyl alcohol dehydrogenase. However, with the exception of 4-coumarate:CoA ligase,[116] their enzyme activities have not yet been established. To add to the possible complexity of the chemotaxonomy/structural diversity of lignans, a red alga-associated actinomycete bacterial species (*Nocardia* sp.) reportedly produces the 8–8′-linked arylnaphthalene lignan justicidin B (**55**),[117] whereas a biosynthetic pathway to form caffeic acid (**56**) was described in another actinomycete (*Saccharothrix espanaensis*).[118] The former is the first report of any lignan in bacteria and, although unprecedented, demands further biochemical confirmation. If correct, it would represent the most basal elaboration of a lignan-forming pathway from an evolutionary perspective.

1.23.3.2.1 Bryophytes: liverworts, hornworts, and mosses

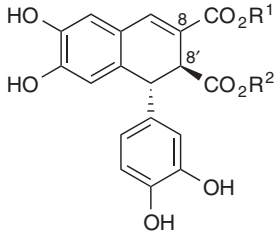
The chemotaxonomical comparison of basal plant lignan occurrence (i.e., in liverworts and hornworts) is of interest, since it possibly provides additional insight regarding the evolution of the phenylpropanoid pathway and of lignan biosynthesis. With one exception, the liverworts analyzed so far have been found to accumulate several 8–8′-linked arylnaphthalene and aryldihydronaphthalene lignans (or derivatives thereof), as well as norlignans. These can be provisionally envisaged as caffeic acid (**56**)-derived, for example, those from *Pellia epiphylla* (**57b–60b**),[119,120] *Lepidozia incurvata* (**57**), *Chiloscyphus polyanthos* (**57**),[121] *Lophocolea heterophylla* (**57b**),[122] *Jamesoniella autumnalis* (**61**, **62a–64a**),[122,123] *Lepicolea ochroleuca* (**65b–67b**),[124] *Bazzania trilobata* (**57b**, **68b**, trilobatins A–K (**69b**, **71b–80b**), **70b**, and the 5′–5″-dimer of **57b** (**81b**)),[125,126] as well as the putative norlignans from *L. incurvata* (**82**, **83**), *C. polyanthos* (**82**, **84**), *Jungermannia exsertifolia* (**82**, **85**),[121] *J. autumnalis* (**86**, **87**),[123] and *B. trilobata* (**82**).[125] Typically, they have aryldihydronaphthalene skeleta (e.g., **57–60**, **65**, **69–80**) with, in a few instances, the pendant aryl group having undergone fission at either the C2–C3 positions resulting in lactone formation as proposed by Tazaki *et al.* [123] (e.g., in **62–64**, **86**, **87**), or at C3–C4 in **68**. Interestingly, all of these metabolites contain (or contained) catechol groups, as well as having dicarboxylic acid (or ester-linked) moieties at carbons C9 and C9′, i.e., there is no apparent methylation of the phenolic groups in these species. A few structures also lack a terminal carboxyl group at C9, thereby providing entry into norlignans (e.g., **82–87**). Relative to our first chemotaxonomic analysis,[1] the lignans (**65–81**) present in *L. ochroleuca*[124] and *B. trilobata*,[125,126] as well as the norlignans **82–85**[121,125] represent new structures.



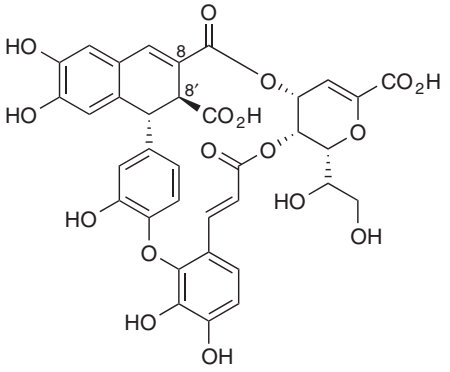
(**55**) Justicidin B



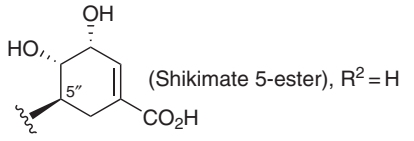
(**56**) Caffeic acid



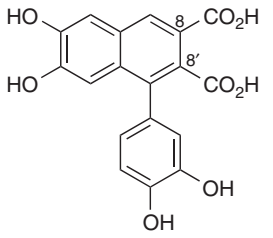
(**57b**) $R^1 = R^2 = H$, (–)-Epiphyllic acid
(**58b**) $R^1 = H$, $R^2 = Me$



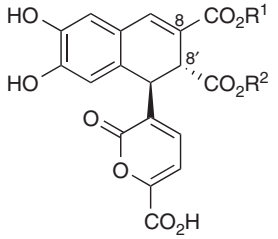
(**60b**) (–)-Pelliatin



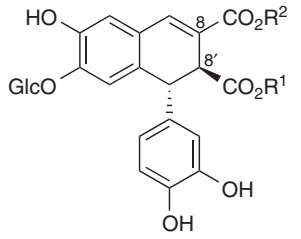
(**59b**) R^1 = (Shikimate 5-ester), R^2 = H



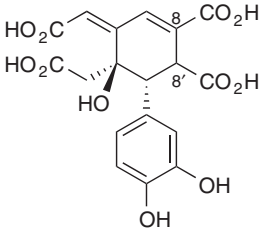
(**61**)



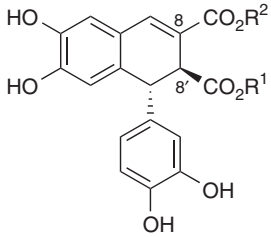
(**62a**) $R^1 = R^2$ = H, (+)-Jamesopyrone
(**63a**) R^1 = Me, R^2 = H
(**64a**) R^1 = H, R^2 = Me

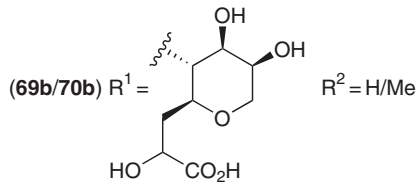


(**65b**) R^1 = Me, R^2 = H
(**66b**) R^1 = Shikimate 5′-ester, R^2 = H
(**67b**) $R^1 = R^2$ = Shikimate 5′-ester

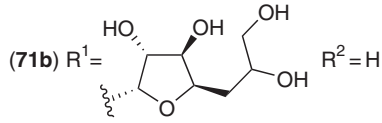


(**68b**) (–)-Bazzania acid

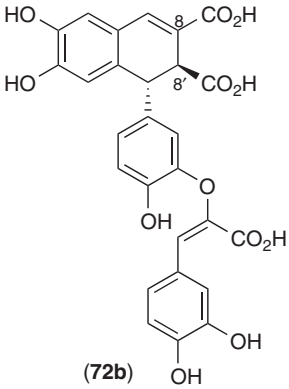




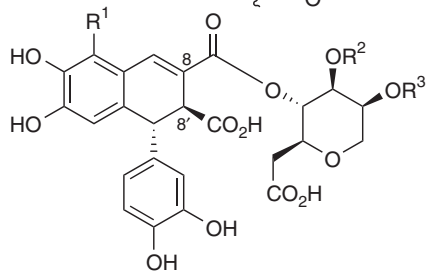
(**69b**/**70b**) R^1 = R^2 = H/Me



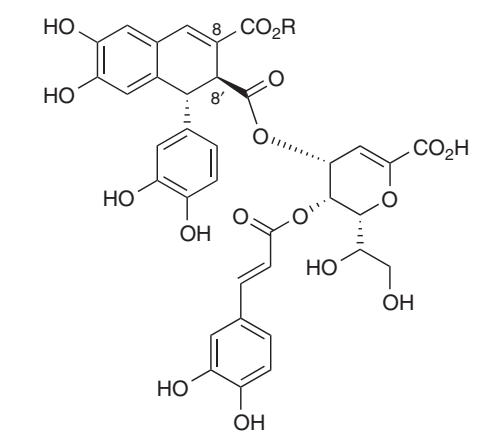
(**71b**) R^1 = R^2 = H

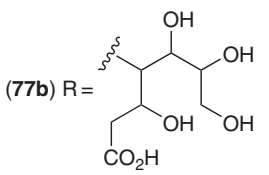


(**72b**)

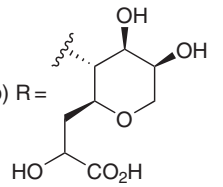


(**73b**) $R^1 = R^2 = R^3$ = H
(**74b**) R^1 = 5-Etherified epiphyllic acid, $R^2 = R^3$ = H
(**75b**) $R^1 = R^2$ = H, R^3 = 9-Esterified jamesopyrone
(**76b**) $R^1 = R^3$ = H, R^2 = 9-Esterified jamesopyrone

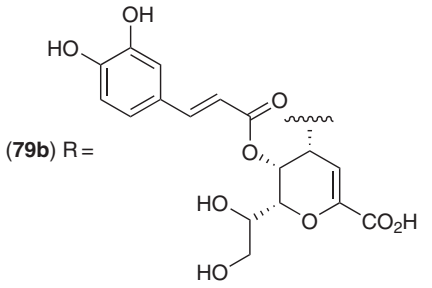




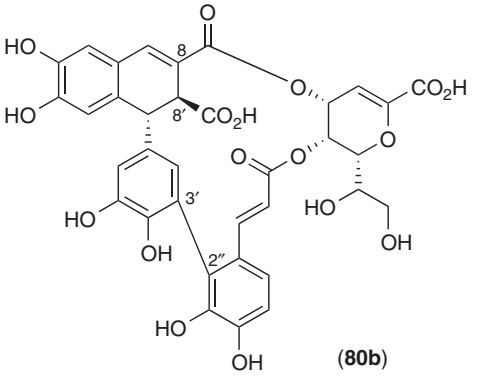
(**77b**) R =



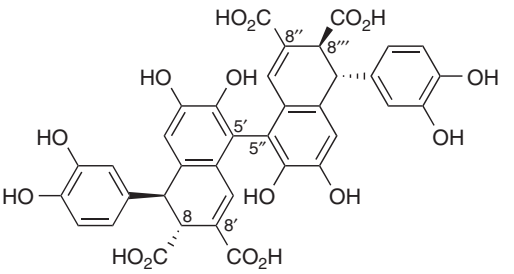
(**78b**) R =



(**79b**) R =



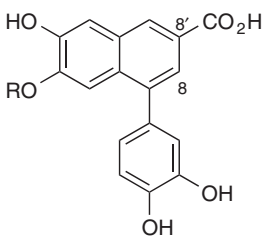
(**80b**)



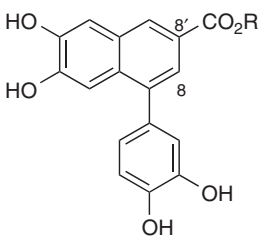
(**81b**)

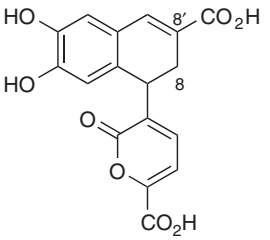
Of the lignans in different liverwort species, epiphyllic acid (**57**) and jamesopyrone (**62**) appear to be two of the most commonly found. Using chiral column chromatography, it was shown that enantiomerically pure (−)-epiphyllic acid (**57b**) accumulates in *Calypogeia azurea*, *L. heterophylla*, and *Aneura pinguis*, whereas the (+)-enantiomer **57a** is found in *Haplomitrium mnioides*, *J. autumnalis*, and *Marsupella emarginata*; (+)-jamesopyrone (**62a**) also accumulated in the latter two species. In other liverwort species, both enantiomers of epiphyllic acid are reported, with one being in enantiomeric excess (e.e.) over the other, that is, the (−)-enantiomer **57b** in >90% e.e. in *Bazzania yoshinagana* and *Heteroscyphus planus*, and the (+)-enantiomer **57a** in >80% e.e. in *Diplophyllum taxifolium*.[127]

Of particular interest also is the report of a presumed allyl-/propenylphenol-derived lignan, (−)-licarin A (**88b**) in *Jackiella javanica*.[128] This differs from the previously described liverwort allylphenols (Section 1.23.3.1.2) in having the pendant double bond conjugated with the aromatic ring. The appearance of a C9/C9′ deoxygenated lignan in this early land plant group is thus again of considerable evolutionary and biosynthetic interest, as it suggests further that pathway elaboration was an early feature in land plant colonization.

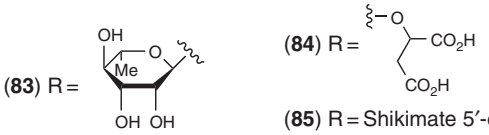


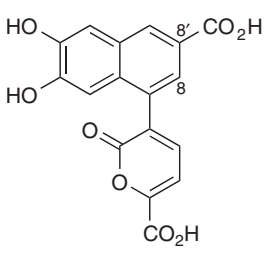
(**82**) R = H



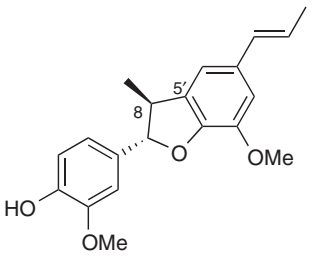


(**86**)



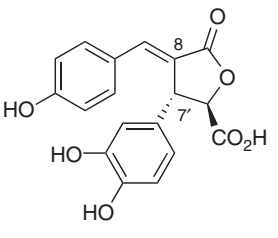


(**87**) Scapaniapyrone

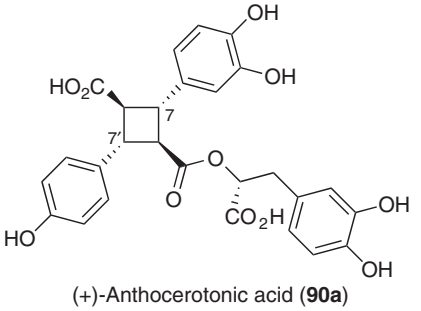


(**88b**) (–)-Licarin A

To our knowledge, the only reports of lignans in hornworts are those of the 8–7′-linked (+)-megacerotonic acid (**89a**) and the 8–7′-linked, 9′-*O*-esterified (+)-anthocerotonic acid (**90a**) present in the hornworts *Megaceros flagellaris*,[129,130] *Dendroceros japonicus*,[130] *Notothylas temperata*,[130] *Phaeoceros laevis*,[130] and *Anthoceros punctatus*,[129,130] respectively. Both are optically active, with the respective (+)-enantiomers being in excess. As before, for the majority of lignans reported in liverworts, these metabolites have highly oxygenated C9 and C9′ end-group moieties and catechol groups (lacking *O*-methylation). Although both lignans are 8–7′-linked, (+)-megacerotonic acid (**89a**) also has a lactone functionality, whereas (+)-anthocerotonic acid (**90a**) has a cyclobutane ring.



(+)-Megacerotonic acid (**89a**)

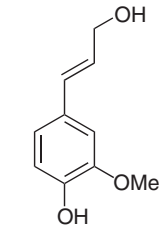


(+)-Anthocerotonic acid (**90a**)

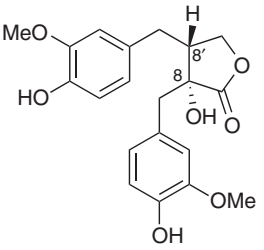
Mosses, on the contrary, have not been reported to accumulate lignans (or, as discussed above, allyl-/propenylphenols), thus differing from the other bryophytes. Their presumed absence in mosses is also of unknown overall significance in evolutionary terms, but underscores the differences in elaboration of various branches of phenylpropanoid metabolism during land plant colonization by early nonvascular plants.

1.23.3.2.2 Pteridophytes: lycophytes, horsetails, and ferns

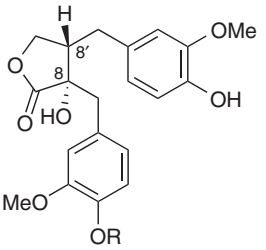
There appears to be very few reports of lignans in the basal vascular plant species so far, although there is considerable structural variety. To date, it is known that the lycophyte *Selaginella doederleinii* accumulates the presumed coniferyl alcohol (**91**)-derived (+)-matairesinol (**10a**), (+)-nortrachelogenin (**92a**, wikstromol) and the glucoside of (−)-nortrachelogenin (**92b**), (−)-nortracheloside (**93b**), as well as the presumed sinapyl alcohol (**94**)-derived 8–8′-linked (−)-*epi*-syringaresinol (**95b**, lirioresinol), and (−)-syringaresinol (**96b**, lirioresinol B).[131] These findings are in agreement with other studies that established the presence of sinapyl alcohol (**94**)-derived lignins in *Selaginella*.[132,133] Additionally, two syringyl-like lignan derivatives (i.e., both *O*-glycosylated and with side-chains containing one and two less carbons, respectively) were also reported to occur in two other *Selaginella* species, namely the 8–5′-linked tamariscinoside C (**97**) from *Selaginella tamariscina*[134] and (−)-moellenoside A (**98b**) from *Selaginella moellendorffii*.[135] In general, however, sinapyl alcohol (**94**) moieties are considered to have evolved mainly in the angiosperms.[136]



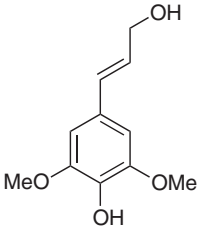
(**91**) Coniferyl alcohol



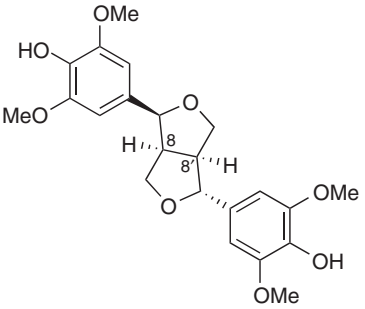
(**92a**) (+)-Nortrachelogenin
[= (+)-Wikstromol]



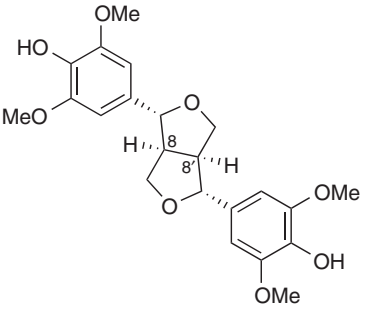
(**92b**) R = H, (–)-Nortrachelogenin
(**93b**) R = Glc, (–)-Nortracheloside



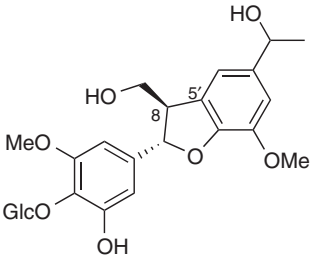
(**94**) Sinapyl alcohol



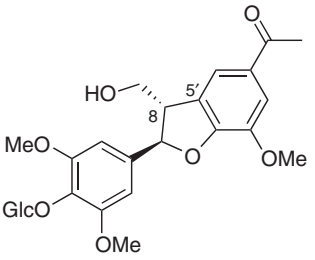
(**95b**) (–)-*epi*-Syringaresinol
[= (–)-Lirioresinol]



(**96b**) (–)-Syringaresinol
[= (–)-Lirioresinol B]



(**97**) Tamariscinoside C

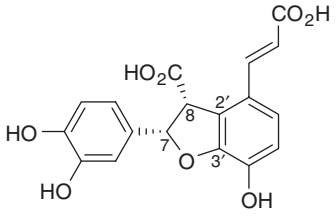


(**98b**) (–)-Moellenoside A

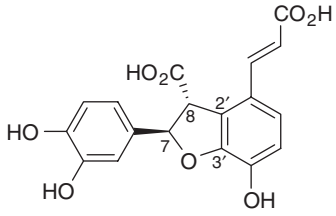
Some examples of reported lignans in fern species are the putatively hydroxycinnamic acid-derived metabolites. This includes the 8–2′-linked (−)-blechnic acid (**99b**) (and its C7 and C8 epimers, **100b** and **101b**, respectively) as well as its shikimate derivative (−)-brainic acid (**102b**), which have been isolated from several Blechnaceae fern species (**Figure 7**).[137,138] These lignans thus share some commonalities with those described aforehand in liverworts and hornworts, that is, with regard to the presence of catechol and carboxylic acid moieties. However, the 8–2′ linkage is quite distinct from the mostly 8–8′-linked aryldihydronaphthalenes in nonvascular lignans.

Other fern lignans can be presumed coniferyl alcohol (**91**)-derived,[1] that is, the two glucosides of (+)-dihydrodehydrodiconiferyl alcohol (**103a**) and (+)-lariciresinol (**105a**), namely, **104b** and **106b** from the fern *Pteris vittata* (Pteridaceae),[139] with these being 8–5′- and 8–8′-linked, respectively. In addition to their different linkage modes, these lignans are noteworthy in having a (saturated) propanol side-chain in **103a** and **104b**, as well as for the *O*-glucoside derivatization found in both **104b** and **106b**.

Didymochlaena truncatula (Dryopteridaceae) contains yet a third structural type, in terms of putative biosynthetic origin. In addition to (−)-nortrachelogenin ((−)-wikstromol (**92b**)) being present, this fern also contains the interesting 5–*O*–4′- and 5–*O*–2′-linked partially dearomatized lignans (+)-didymochlaenone A (**107**) and (−)-didymochlaenone B (**108**).[140] Both lignans are optically active, and both bear an allylphenol-like side-chain functionalization. These natural products are of particular interest since phenoxy radical–radical coupling has (provisionally) occurred at the C5 position, which presumably initially harbored a methoxyl group. Also noteworthy is the presence of methylenedioxy groups, which were previously considered to have evolved later in the gymnosperm lignans,[141] as well as the oxygenation at C2′ to form the conjugated enone moiety of **107** and **108** and the 5–*O*–2′ interunit linkage of **108**.

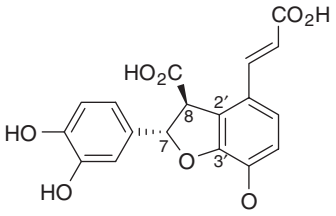


(**99b**) (–)-Blechnic acid

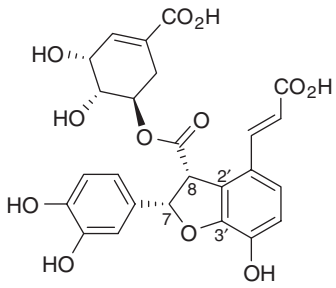


(**100b**) (–)-7-*epi*-Blechnic acid

Blechnum spicant

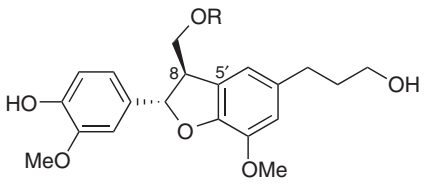


(**101b**) (–)-8-*epi*-Blechnic acid

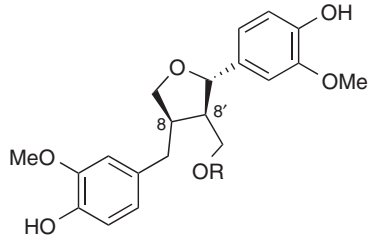


(**102b**) (–)-Brainic acid

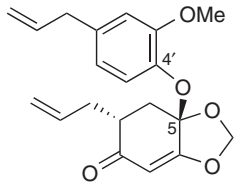
Figure 7 Lignans in *Blechnum spicant* (pteridophyte). Photograph by Henri Moore, Washington State University, USA.



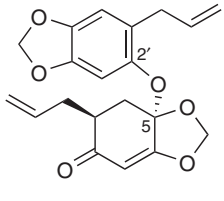
(**103a**) R = H, (+)-*trans*-Dihydrodehydrodiconiferyl alcohol
(**104b**) R = Glc, (–)-*trans*-Dihydrodehydrodiconiferyl alcohol-9-*O*-β-D-glucoside



(**105a**) R = H, (+)-Lariciresinol
(**106b**) R = Glc, (–)-Lariciresinol-9-*O*-β-D-glucoside



(**107**) (+)-Didymochlaenone A



(**108**) (–)-Didymochlaenone B

The fern lignans discussed above can thus be classified in three quite distinct groups based on their side-chain functionalities (i.e., whether containing carboxylic acid, alcohol, or allyl/propenyl moieties). Fern lignans thus appear to have an increased structural diversity relative to those present in earlier groups, including the degree of oxygenation of their side-chain moieties, the linkage modes between monomeric precursors, and glycosylation. No lignans have apparently yet been described in other early vascular plant groups, including the horsetails in which, as discussed before in Section 1.23.3.1.3, only the propenylphenol **5** has been reported to accumulate in trace amounts. Nevertheless, as discussed in the next section, this increased structural diversity trend becomes even more evident in gymnosperm and angiosperm lignans.

1.23.3.2.3 Spermatophytes: gymnosperms and angiosperms

1.23.3.2.3(i) Gymnosperms The distribution of various lignan/norlignan skeleta throughout the gymnosperms is not uniform in terms of structural types, with none so far reported in the cycads. As described earlier (Section 1.23.3.1.4), however, allyl-/propenylphenols have been found in at least one cycad, *C. revoluta*.[56]

The evolution of the different gymnosperm plant groups was, nevertheless, overall accompanied by a vast increase in the structural variety of both lignans and norlignans, with more than a hundred different structures now reported.[142] Although often particularly abundant in conifer heartwood[143] and knots,[144] they have also been isolated from trees at all developmental stages, and from all different tissues, such as bark, roots, needles, cones, and so on. Most lignans are 8–8′-linked,[142] giving rise to carbon skeleta classified as either furofurans (e.g., (−)-sesamin (**11b**) in *Juniperus thurifera*, **Figure 8**);[145] tetrahydrofurans (e.g., shonanin (**109**) in *Calocedrus formosana*[146] and *P. taeda* cell cultures);[147] dibenzylbutyrolactones (e.g., (−)-matairesinol (**10b**) in *P. spicatus*);[13] dibenzylbutanes (e.g., (−)-secoisolariciresinol (**110b**) in *P. spicatus*);[148,149] arylnaphthalenes (e.g., junaphthoic acid (**111**) in *Juniperus sabina*),[150] or aryltetrahydronaphthalenes (e.g., (−)-plicatic acid (**112b**) in *Thuja plicata*[151–153] and (−)-α-conidendrin (**113b**) in *Tsuga heterophylla*).[154,155] As aforementioned, some of these skeleta are also present in more basal plant groups, for example, arylnaphthalenes, which are abundant in liverworts, as well as furofurans and dibenzylbutyrolactones, which have been reported to occur in ferns.

Generally, gymnosperm lignans have either guaiacyl-like aryl groups, that is, coniferyl alcohol (**91**)-like, such as in shonanin (**109**), (−)-secoisolariciresinol (**110b**), and (−)-α-conidendrin (**113b**), or infrequently have 3,4-methylenedioxy bridges, such as with (−)-sesamin (**11b**). Interestingly, some species can accumulate lignans where the C5 carbon is hydroxylated (e.g., (−)-plicatic acid (**112b**) and its derivatives in *T. plicata*), with junaphthoic acid (**111**) having the analogous position additionally *O*-methylated. Although the majority of gymnosperm lignans have hydroxymethyl-like side-chain functionalities (i.e., either as alcohols/ethers or as products presumably derived from further metabolism, such as those affording lactones), there are also a few examples of methyl (i.e., C9- and/or C9′-deoxygenated) end groups in some conifer lignans, for example, junaphthoic acid (**111**) in *J. sabina*. Indeed, although it is also tempting to speculate that these C9- and/or C9′-deoxygenated lignans are allyl-/propenylphenol-derived, this needs to be established biochemically as other possibilities could exist.

Most 8–8′-linked lignans exist as dimers, although higher oligomers have been reported, with the largest documented so far being the (−)-plicatic acid (**112b**)-derived lignans of MW ∼10 000 Da in *T. plicata*.[156] As for the other plant classes described earlier, many of the conifer lignans are either optically pure or highly enriched in one enantiomeric form, for example, 7-hydroxymatairesinol (**114**) is present as ∼85% of the whole lignan mixture in *P. abies* knots, but where the (−)-7*S*- and (−)-7*R*-forms are in a ratio of 32:1.[157]

Lignans with 8–5′ and 8–*O*–4′ linkages are also frequently reported, for example, (±)-dehydrodiconiferyl alcohols (**14a**/**b**) and guaiacylglycerol 8–*O*–4′ coniferyl alcohol ether (**115**) in *P. taeda* cell cultures.[147] Other metabolic modifications often observed include those leading to saturated propanol side-chains, for example, (−)-dihydrodehydrodiconiferyl alcohol (**103b**) in *Juniperus chinensis*,[158]

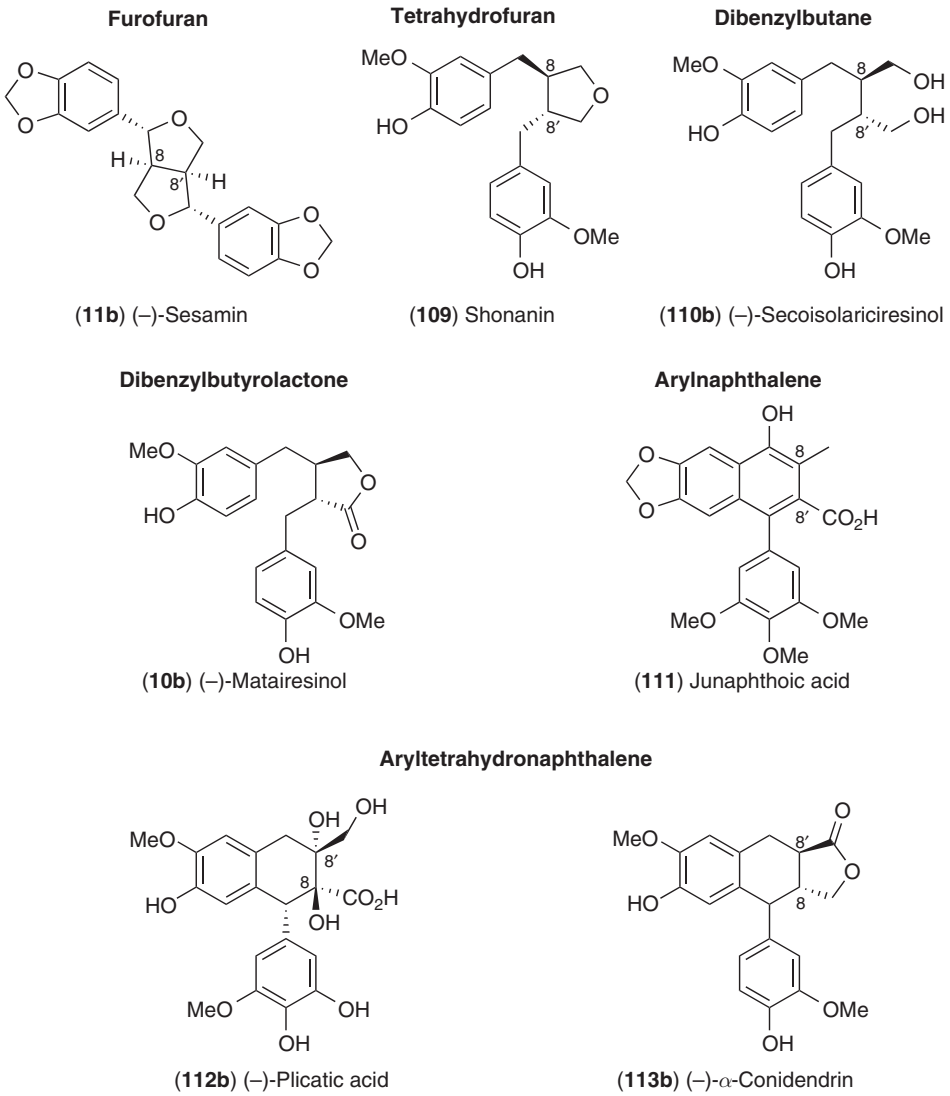
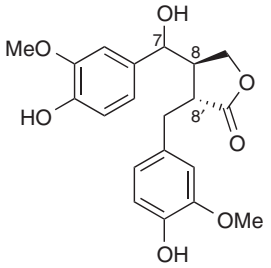
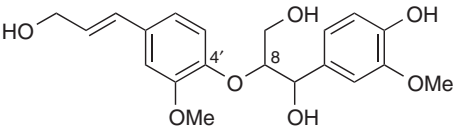


Figure 8 Different classes of 8–8′-linked lignans.

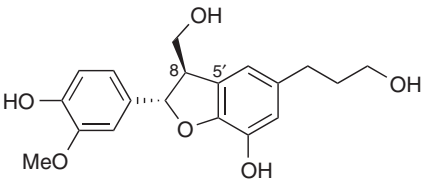
(+)-cedrusin (**116a**) in *Cedrus deodara*[159] and *T. heterophylla*,[155] as well as **117** in *J. chinensis*.[158] The additionally reduced (i.e., cleaved 7–*O*–4′ linkage) peracetylated lignan **118** is also found in *Cryptomeria japonica*.[160] Side-chain reduction apparently occurs on phenylpropanoid monomers as well, with dihydromonolignols such as dihydroconiferyl alcohol (**119**) and its *O*-glucoside (**120**) being present in *Picea glauca* knots.[161] The 8–5′-linked dihydrofuran lignans, (±)-dehydrodiconiferyl alcohols (**14a**/**b**), are thought to be central precursors of (±)-dihydrodehydrodiconiferyl alcohols (**103a**/**b**) and (±)-cedrusins (**116a**/**b**): *Pinus taeda* cell cultures have been shown to regiospecifically *O*-demethylate the C3′ of both dehydrodiconiferyl alcohols (**14a**/**b**) and dihydrodehydrodiconiferyl alcohols (**103a**/**b**).[1] In combination with side-chain reductions, these reactions afforded (±)-cedrusins (**116a**/**b**) as a product in these cell cultures.[1]



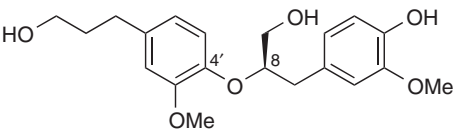
(**114a**) 7*R*, Hydroxymatairesinol
(**114b**) 7*S*, allo-Hydroxymatairesinol



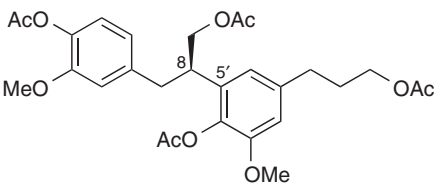
(**115**) Guaiacylglycerol 8-*O*-4′ coniferyl alcohol ether



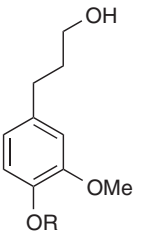
(**116a**) (+)-Cedrusin



117

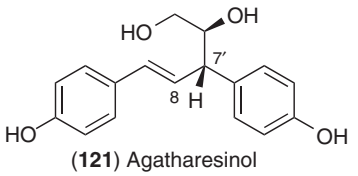


118

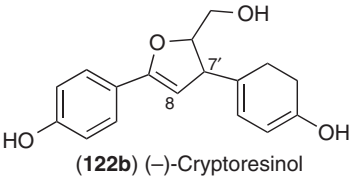


(**119**) R = H, Dihydroconiferyl alcohol
(**120**) R = Glc, Dihydroconiferyl alcohol glucoside

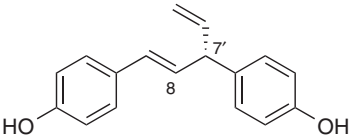
Loss of a terminal carbon at C9 affords norlignans.[142,162] Most of the true gymnosperm norlignans are 8–7′-linked dibenzylpentanes ($C_6C_5C_6$) and are notable components of the Cupressaceae, Araucariaceae, and Taxodiaceae, but are apparently absent from the Pinaceae, Podocarpaceae, and other conifers.[142,162] These are exemplified by agatharesinol (**121**) in *Agathis australis* (Araucariaceae),[163,164] *Sequoia sempervirens*,[165] and *Sequoiadendron gigantea*,[165] as well as (−)-cryptoresinol (**122b**)[166] and (−)-(*E*)-hinokiresinol (**123b**) in *C. japonica*.[167,168] They are structurally quite distinct from the arylnaphthalene norlignans found in the more basal plants mentioned earlier (Section 1.23.3.2.1). Although **121–123** only have monooxygenated phenyl groups, other 8–7′-linked dibenzylpentane norlignans can have catechol, dimethoxyaryl, and trimethoxyaryl moieties, for example, sequirin B (**124**) in *S. sempervirens*, and permethyl-sequirin E (**125**) and permethyl-sequirin G (**126**) in *S. gigantea*.[165] Other 8–7′-linked norlignans include the structurally rare spiro-lignan (−)-athrotaxin (**127b**) from *Athrotaxis selaginoides*[169] and *Metasequoia glyptostroboides*,[170] as well as the presumed product of (−)-hydroxyathrotaxin (**128b**) rearrangement, the cycloheptadiene lignan metasequirin B (**129**), from *M. glyptostroboides*.[171] Additional linkage modes present in conifer norlignans are less common, for example, sequirin D (**130**) in *S. sempervirens*[165] and the 7–7′-linked norlignan **131** in *Araucaria angustifolia*.[172]



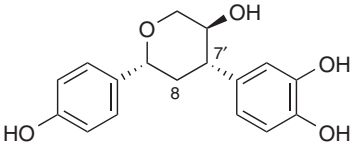
(**121**) Agatharesinol



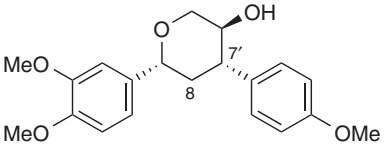
(**122b**) (–)-Cryptoresinol



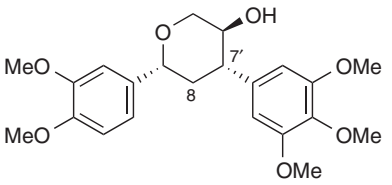
(**123b**) (–)-(*E*)-Hinokiresinol



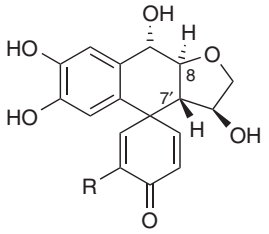
(**124**) Sequirin B



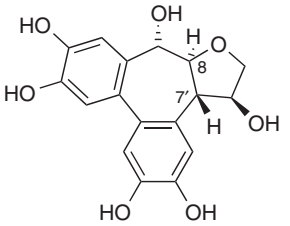
(**125**) Permethyl-sequirin E



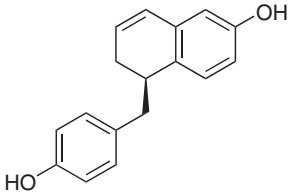
(**126**) Permethyl-sequirin G



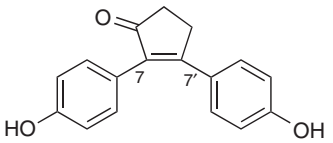
(**127b**) R = H, (–)-Athrotaxin
(**128b**) R = OH, (–)-Hydroxyathrotaxin



(**129**) Metasequirin B

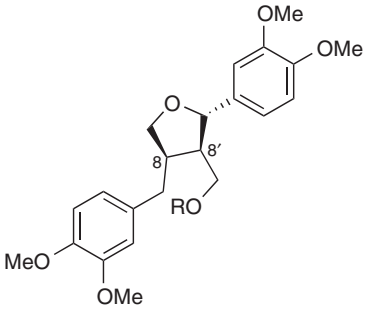


(**130**) Sequirin D

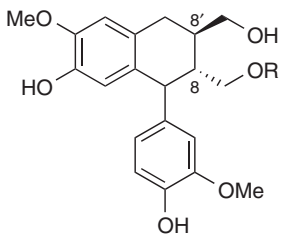


131

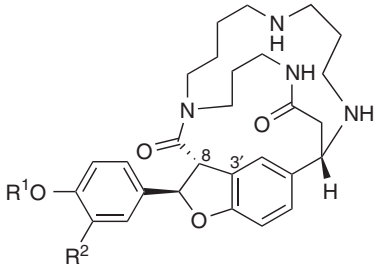
There are only few reports of lignans in Gnetales, most of which are of 8–8′-linked compounds, for example, (+)-syringaresinol (**96a**) in *Gnetum gnemon* roots;[173] (±)-syringaresinols (**96a**/**b**) in *Ephedra alata*;[174] (+)-lariciresinol (**105a**), (+)-isolariciresinol (**133a**), and their 9-acetate derivatives **132a** and **134a** in *Ephedra viridis*;[175] as well as the 8–3′-linked spermine diamides, ephedradines A (**135**), B (**136**), and C (**137**) in mao-kon, the crude drug from *Ephedra* roots.[176–178] The latter are rare examples of N-containing lignan compounds. *Ginkgo biloba*, on the contrary, has been reported to accumulate the 8–8′-linked (+)-sesamin (**11a**) in both its heartwood[179] and leaves.[180] As noted earlier, allyl-/propenylphenols have not been reported in either of the latter two gymnosperm families.



(**105a**) R = H, (+)-Lariciresinol
(**132a**) R = Ac, (+)-Lariciresinol 9-acetate



(**133a**) R = H, (+)-Isolariciresinol
(**134a**) R = Ac, (+)-Isolariciresinol 9-acetate



(**135**) $R^1 = R^2 = H$, Ephedradine A
(**136**) $R^1 = H$, $R^2 = OMe$, Ephedradine B
(**137**) $R^1 = Me$, $R^2 = OMe$, Ephedradine C

Additionally, in many highly valued woody species, nonstructural lignans and less frequently norlignans can be deposited in their heartwoods, for example, (−)-plicatic acid (**112**) and its derivatives, which can constitute up to 20% of its dry weight in *T. plicata*,[156] and which are in part responsible for the wood's durability and texture. Other heartwood-accumulating lignans/norlignans include 7-hydroxymatairesinol (**114**) and (−)-α-conidendrin (**113b**) in *T. heterophylla*,[154] and others such as **116**, **122–124** in *C. japonica*,[166,167] all of which significantly contribute to their heartwood properties. It is also of interest that all three of these species are long-living gymnosperms: *Thuja plicata*, native to the northwestern United States and southwestern Canada, and *C. japonica*, native to Japan and China, can have lifespans in excess of 3000 years, whereas *T. heterophylla*, a native to the west coast of North America, has specimens as old as 1200 years.

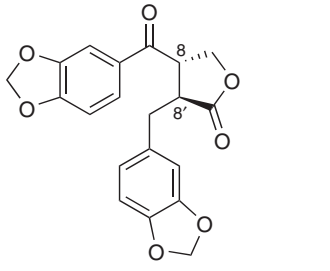
Therefore, the emergence of gymnosperms, in particular the conifers, apparently gave rise to a great increase in the number and variety of lignan and norlignan structures. Although the number of the reports so far may potentially result from a bias of phytochemical studies toward higher plants that are often important wood resources, the gymnosperm lignin structures are ostensibly structurally more varied, that is, in terms of linkage modes, oxygenation patterns of both aromatic groups and side chains, as well as further modifications such as those giving rise to dihydrophenylpropanoids and norlignans. This structural diversity is further expanded within the angiosperms, as described below.

1.23.3.2.3(ii) Angiosperms The emergence of the angiosperms was also accompanied by another massive increase in lignan and (less commonly) norlignan skeletal types, albeit with most of the common structures observed in the gymnosperms being retained in the angiosperms. Of the lignans/norlignans reported to date, however, most are found in the woody and nonwoody dicotyledons, with only a few examples in the monocotyledons. Many contain new skeletal types and many are again optically active.

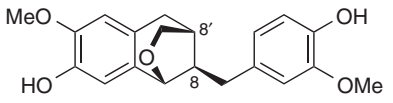
As for the gymnosperms, the terminal groups (C9/C9′) of the angiosperm lignans are frequently modified, that is, with these having lactone/ether (e.g., **138a**, (−)-padocin (**139b**)), hydroxymethyl (e.g., **139b**, (−)-balanophonin (**140b**)), aldehyde (e.g., **140b**), and carboxylic acid (e.g., 5–5′-dehydrodiferulic acid (**141**)) groups, as well as others lacking an oxygenated functionality at C9/C9′ (e.g., **142b–151**).[181,182] Note, however,

that 5–5′-dehydrodiferulic acid (**141**) can be viewed as a lignan artifact of sorts, since it is presumed to be derived from cell wall polymeric carbohydrates containing feruloyl moieties in close proximity (which can undergo radical–radical coupling), and indeed it is only released upon alkali treatment of cell walls.[183,184]

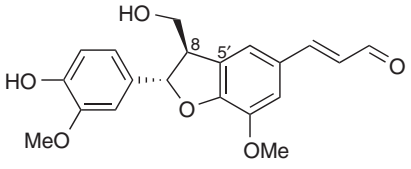
Although aromatic ring substitution patterns in many of the angiosperm lignans are guaiacyl-like (e.g.. **139b–141, 144b**, including those with methylenedioxy groups such as **138a, 146b**), numerous hydroxyphenyl (e.g., **150b**) and trihydroxylated phenyl ring(s) (e.g., **145a, 147b, 152a**) lignans have also been reported. Additional interesting features are those of relatively rare skeletal types, such as the 7–1′-linked (+)-peperomin A (**152a**),[185] the 1–2′-linked (−)-eupomatilone 6 (**153b**),[186] and the 7–1′-linked (+)-iso-ocobullenone (**154a**).[187]



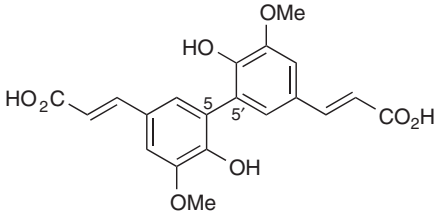
(+)-**138a**



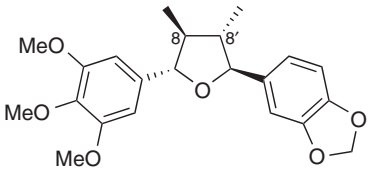
(**139b**) (–)-Padocin



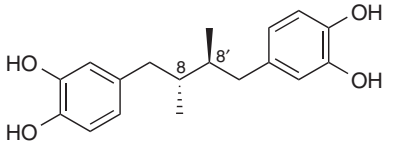
(**140b**) (–)-Balanophonin



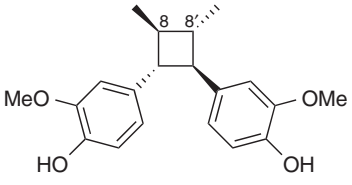
(**141**) 5–5′-Dehydrodiferulic acid



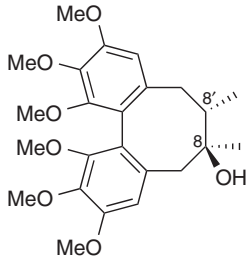
(–)-(**142b**)



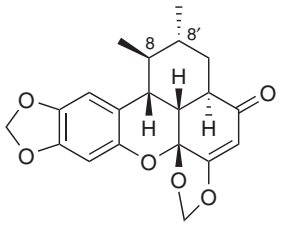
(**143**) Nordihydroguaiaretic acid (NDGA)



(**144b**) (–)-Endiandrin A



(**145a**) (+)-Schizandrin

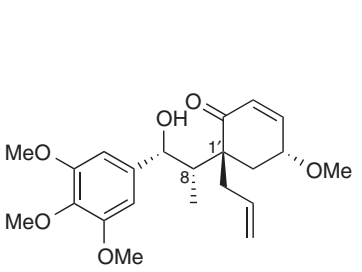


(**146b**) (–)-Sauchinone

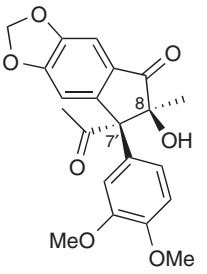
The 8–8′ linkage mode, however, overall remains the most frequent in terms of the number of examples of reported structures. These can be again classified as furofurans (e.g., (+)-sesamin (**11a**) from *S. indicum*);[14,15] tetrahydrofurans (e.g., (−)-olivil (**12b**) from *O. europaea*);[16–18] dibenzylbutanes (e.g., (−)-guaiaretic acid (**9b**) from *G. officinale*);[10–12] dibenzylbutyrolactones (e.g., (−)-matairesinol (**10b**) in *Forsythia intermedia*);[188] aryletrahydronaphthalene derivatives (e.g., (−)-podophyllotoxin (**1b**) from *Podophyllum* spp.);[189] dibenzocyclooctadienes (e.g., (+)-schizandrin (**145a**) from *Schizandra chinensis*);[190] cyclobutanes (e.g., (−)-endiandrin A

(**144b**) in *Endiandra anthropophagorum*);[191] as well as other more uncommon structures, such as the polycyclic (−)-sauchinone (**146b**) from *S. chinensis*[192] and (−)-padocin (**139b**) from *Haplophyllum cappadocicum*.[193]

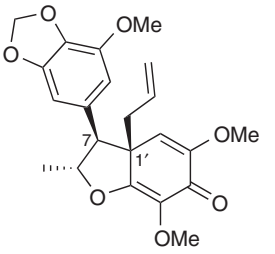
Among other coupling modes reported in angiosperms, these include 8–5′- (e.g., **140b**);[194] 8–1′- (e.g., **147b**);[195] 8–7′- (e.g., **148**);[21] 7–1′- (e.g., **149**);[196] 8–*O*–4′- (e.g., **150b**);[197] and benzodioxocin- (e.g., **151**)[198] linked lignans. Such diverse skeletal types further underscore the varied biochemical processes operative in angiosperms, most of which still remain to be elucidated (see Section 1.23.4). Other structural features of interest are the presumed rearrangement products, such as **152a** and **153b**, formed through putative migration of aryl substituents, or that of **154a** through ring expansion.



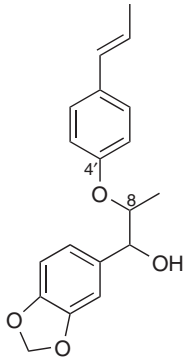
(**147b**) (–)-Megaphone



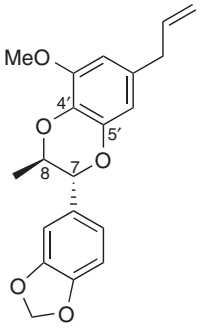
148



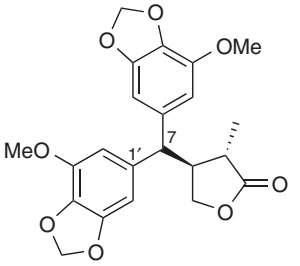
(**149**) Chrysophyllon 1A



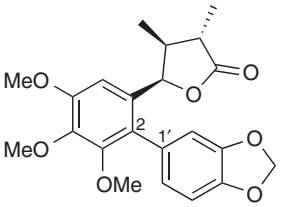
(–)-(**150b**)



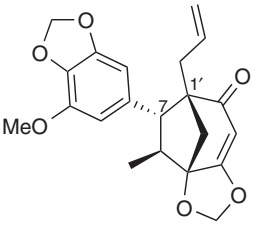
(**151**) Eusiderin B



(**152a**) (+)-Peperomin A

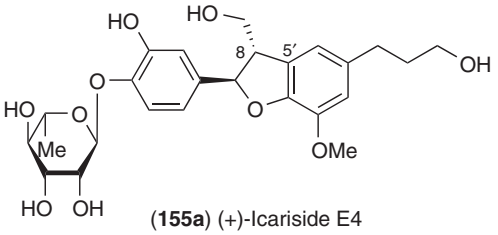


(**153b**) (–)-Eupomatilone 6

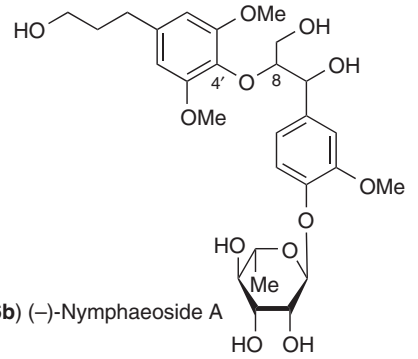


(**154a**) (+)-Iso-ocobullenone

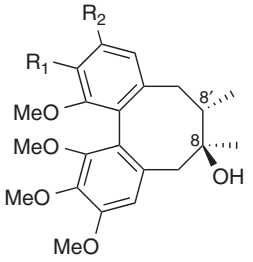
Additionally, the 8–5′-linked (+)-icariside E4 (**155a**) and the 8–*O*–4′-linked (−)-nymphaeoside A (**156b**) found in *Nymphaea odorata* have saturated side-chains.[199] This is a biochemical signature previously observed for some of the lignans in ferns and gymnosperms. The dibenzocyclooctadiene 8–8′-linked lignans present in *Sc. chinensis* (e.g., (+)-schizandrin (**145a**) and gomisins A–C (**157a–159b**))[190,200] are of an apparently unique skeletal type, however, presumably being formed from a linear 8–8′-linked diarylbutane lignan through aryl radical–radical coupling.



(**155a**) (+)-Icariside E4

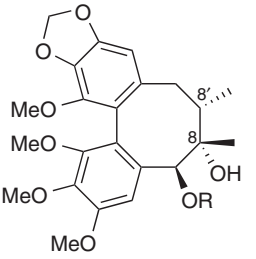


(**156b**) (–)-Nymphaeoside A



(**145a**) $R_1 = R_2 = OMe$, (+)-Schizandrin

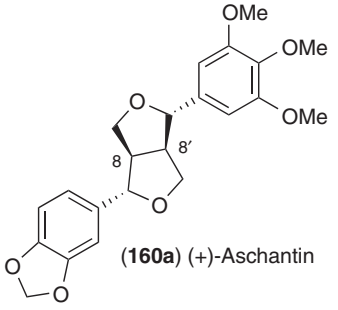
(**157a**) $R_1 + R_2 = OCH_2O$, (+)-Gomisin A



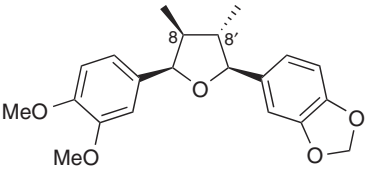
(**158b**) R = (Angelate), (–)-Gomisin B

(**159b**) R = Benzoate, (–)-Gomisin C

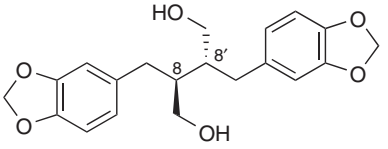
There are also more than 100 distinct lignans[201] isolated and characterized from the genus *Piper* alone, which contains ~700 species including *Piper nigrum*, the source of black and white peppers. Many of these lignans contain common structural motifs, such as the 8–8′-linked furofurans (e.g., (+)-aschantin (**160a**)), the tetrahydrofurans (e.g., (+)-calopiptin (**161a**)), and the diarylbutanes (both linear and cyclic, e.g., (−)-dihydrocubebin (**162b**) and heterotropan (**163**), respectively). However, there are also examples of other linkage types such as 8–3′ (e.g., (+)-conocarpan (**16a**) and schmiditin (**164**)), 8–1′ (e.g., (+)-burchellin (**165a**) and presumed rearrangement products, such as kadsurenin H (**166**)), as well as 8–*O*–4′ (e.g., polysyphorin (**167**)), and 8–2′/3′ (e.g., isofutoquinol A (**168**)).



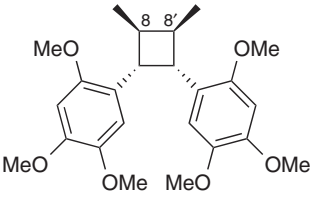
(**160a**) (+)-Aschantin



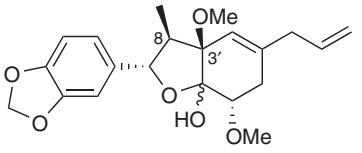
(**161a**) (+)-Calopiptin



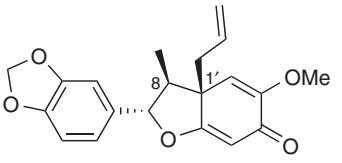
(**162b**) (–)-Dihydrocubebin



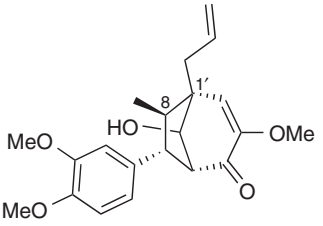
(**163**) Heterotropan



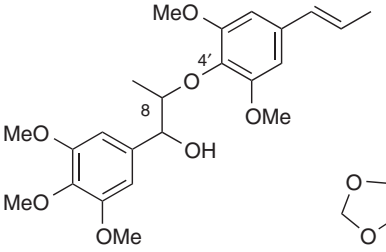
(**164**) Schmiditin



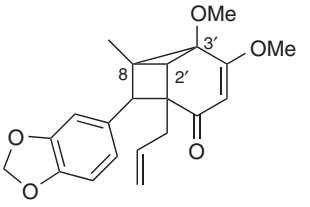
(**165a**) (+)-Burchellin



(**166**) Kadsurenin I

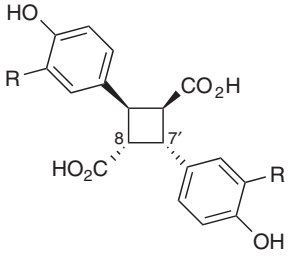


(**167**) Polysyphorin

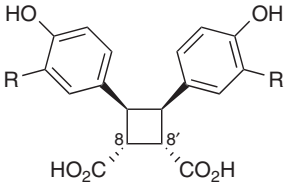


(**168**) Isofutoquinol A

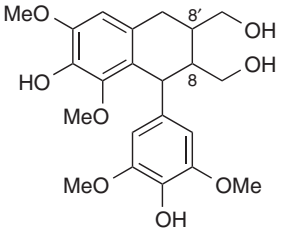
As noted before,[1] only a few monocotyledons are known to accumulate lignans, such as the aryl cyclobutanes, the dihydroxytruxillic/truxinic acids (**169**/**170**) in cereals and grasses.[202,203] Nevertheless, these, as well as **144b** and **163**, may not be necessarily formed through enzymatic catalysis, but perhaps instead as 2,2-photochemical adducts.[202–204] Bamboo (*Phyllostachys edulis*) stems, on the contrary, accumulate, in addition to the 8–8′-linked aryltetrahydronaphthalene lyoniresinol (**171**), the two dilignans phyllostadimers A (**172b**) and B (**173a**).[205] Although all three bamboo lignans contain syringyl-like aromatic moieties, the latter two are 8′–8″-linked dimers but containing the previously unknown 7′–*O*–9/8–*O*–9′ linkages.



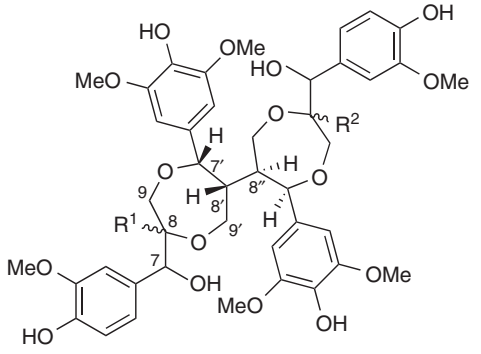
(**169**) R = H or OMe,
Dihydroxytruxillic acids



(**170**) R = H or OMe,
Dihydroxytruxinic acids



(**171**) Lyoniresinol



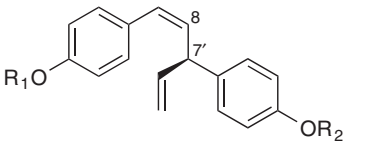
(**172b**) R1 = H, R2 = H, (–)-Phyllostadimer A

(**173a**) R1 = H, R2 = H, (+)-Phyllostadimer B

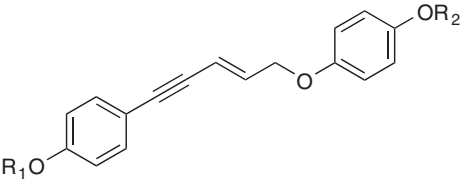
In contrast to the lignans, there are relatively few examples of natural products considered as norlignans in angiosperms. They are, however, most often found in the Asparagales, for example, the 8–7′-linked (+)-(*Z*)-hinokiresinol (**174a**) that occurs in *Asparagus* spp.,[206,207] and in its (−)-form (**174b**) in *Anemarrhena asphodeloides*,[208] as well as the glycosides **175**–**177** in *Hypoxis* spp., which are derived from the (+)-enantiomer

174a.[209,210] The rare acetylenic norlignans compounds, asparenydiol (**178**), asparenyol (**179**), and asparenyn (**180**) also occur in *Asparagus* spp.,[211] as well as rooperol (**181**) glycoside, hypoxoside (**182**) in *Hypoxis* spp.,[212] and nyasicoside (**183**) in *Curculigo capitulata*.[213] Of particular interest is the oxygen insertion forming metabolites **178**–**180**.

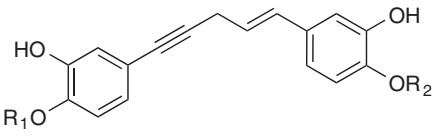
Additional presumed norlignans present in other orders are either 8–8′- or 8–3′-linked. These include the tetrahydrofuran cestrumoside (**184**) in *Cestrum diurnum* (Solanales),[214] the benzofuran machicendiol (**185**) in *Machilus glaucescens* (Laurales),[215] and the arylnaphthalene vitrofolal A (**186**) in *Vitex rotundifolia* (Lamiales).[216] Aromatic substitution patterns in the norlignans are again diverse, with hydroxyphenyl and guaiacyl moieties (and their *O*-methyl and methylenedioxy derivatives) apparently more common. Side-chain functionalities are also varied, being either hydroxymethyl, aldehydic, or deoxygenated.



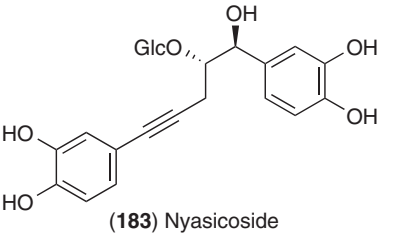
(**174a**) $R_1 = R_2 = H$, (+)-(*Z*)-Hinokiresinol
(**175**) $R_1 = H$, $R_2 = Glc$, (*Z*)-Mononyasine A
(**176**) $R_1 = Glc$, $R_2 = H$, (*Z*)-Mononyasine B
(**177**) $R_1 = R_2 = Glc$,(*Z*)-Nyasoside



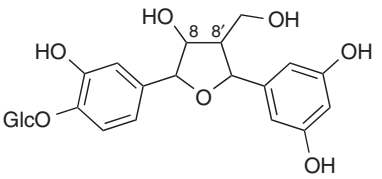
(**178**) $R_1 = R_2 = H$, Asparenydiol
(**179**) $R_1 = H$, $R_2 = CH_3$, Asparenyol
(**180**) $R_1 = R_2 = CH_3$, Asparenyn



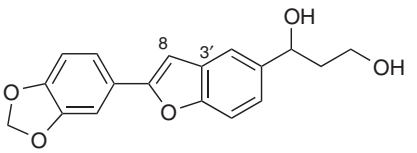
(**181**) $R_1 = R_2 = H$, Rooperol
(**182**) $R_1 = R_2 = Glc$, Hypoxoside



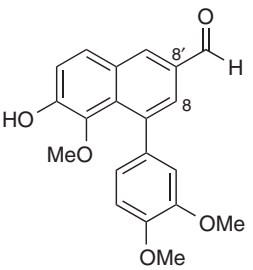
(**183**) Nyasicoside



(**184**) Cestrumoside



(**185**) Machicendiol



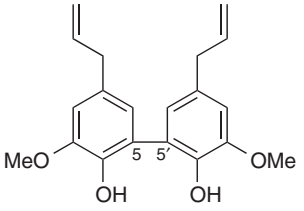
(**186**) Vitrofolal A

1.23.3.3 Evolution of Biochemical Pathways to Allyl-/Propenylphenols and Lignans: Observations on Co-occurrence

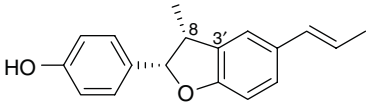
The diversity and distribution of lignans and norlignans from the most basal plants through to the spermatophytes (gymnosperms and angiosperms), allied to their possible absence in mosses, horsetails, and cycads, seem to indicate that their formation was a feature that may also have appeared multiple times during plant

evolution. That is, with their formation being part of a convergent, and not necessarily divergent, biochemical evolutionary process. This is consistent with the phylogenetic/evolutionary 'scattering' of the allyl-/propenylphenols discussed aforehand. Nevertheless, such considerations are at present merely speculative, and underscore the need for a much more thorough (bio)chemical understanding of lignan-forming processes through the diverse extant plant lineages.

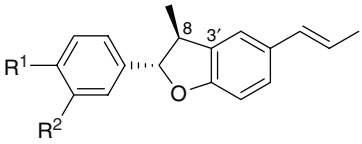
Interestingly, the relative levels of accumulation of lignans, norlignans, and their putative monomeric precursors vary markedly among species. In the case of allyl-/propenylphenols, for example, only a small amount of the lignan, dehydrodieugenol (**187**) (~0.0016% dry wt), accumulates in clove,[217] whereas there is up to ~89% of eugenol (**4**) in the essential oil.[86] *Piper regnellii*, by contrast, accumulates similar amounts (~1.5–1.6% dry wt) of both allyl-/propenylphenols (**23**, **36**, **38**) and the presumed allyl-/propenylphenol-derived lignans (**16a**, **188–193**, **Figure 9**).[218] On the contrary, the creosote bush (*Larrea tridentata*) apparently accumulates no significant amount of monomeric allyl-/propenylphenols, but yet can have ~10% nordihydroguaiaretic acid (NDGA) (**143**) (dry wt) as one of its lignans in the leaves.[219] Biochemical explanations for such variability need to be fully understood in future.



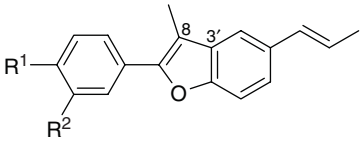
(**187**) Dehydrodieugenol



(**188b**) (–)-*epi*-Conocarpan



(**16a**) $R^1 = OH$, $R^2 = H$, (+)-Conocarpan
(**189a**) $R^1 = OH$, $R^2 = OMe$
(**190a**) $R^1 + R^2 = OCH_2O$, (+)-Regnelline



(**191**) $R^1 = OH$, $R^2 = H$, Eupomatenoid-6
(**192**) $R^1 = OH$, $R^2 = OMe$, Eupomatenoid-5
(**193**) $R^1 + R^2 = OCH_2O$, Eupomatenoid-3

Figure 9 Lignans in *Piper regnellii*. Photographs by Laurence B. Davin, Washington State University, USA.

1.23.4 Lignan Early Biosynthetic Steps: 8–8′ Phenylpropanoid Coupling

The previously discussed chemotaxonomical distribution of allyl-/propenylphenols, lignans, and norlignans resulted in two major observations: The first was that of the still incomplete understanding as to the extent of elaboration of the phenylpropanoid pathway throughout the plant kingdom, with provisional evidence suggesting some algae may have entire biochemical pathways to the allyl-/propenylphenols. The second was that of the enormous structural diversity of the lignans and norlignans, in terms of skeletal types occurring in the major land plant groups. Such observations and reports thus underscored the need to obtain a full biochemical understanding of the various proteins, enzymes, and genes involved in their specific pathways. Progress made so far in this is described below.

1.23.4.1 Discovery of the (+)-Pinoresinol-Forming Dirigent Protein and Encoding Gene

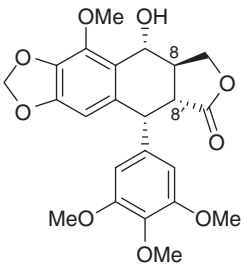
In the course of discovering how lignan formation occurs biochemically in different plant species, our earlier metabolic studies[220–226] established that pinoresinol (**13**) was the central precursor of various 8–8′-linked lignans, that is, leading to the formation of furofuran (e.g., sesamin (**11**)), tetrahydrofuran (e.g., lariciresinol (**105**)), dibenzylbutane (e.g., secoisolariciresinol (**110**)), dibenzylbutyrolactone (e.g., matairesinol (**10**)), and aryltetrahydronaphthalene (e.g., (−)-6-methoxypodophyllotoxin (**194b**, **Figure 10**) and (−)-plicatic acid (**112b**)), depending on the species. Additionally, secoisolariciresinol (**110**) and matairesinol (**10**) were deduced to be intermediates in the biosynthesis of the aryltetrahydronaphthalene (e.g., **194b**)[226] and, by extension, the arylnaphthalene lignans.

Forsythia shrubs (Oleaceae) were key to the discovery of many of these biochemical processes. *Forsythia* species accumulate various 8–8′-linked lignans in differing amounts: Specifically, *F. suspensa* contains (+)-pinoresinol (**13a**), (+)-phillyrin (**195a**), and (+)-phillygenin (**196a**),[227] whereas (−)-matairesinol (**10b**), (−)-arctigenin (**197b**), and (−)-arctiin (**198b**) are present in *Forsythia viridissima*,[227] with *F. intermedia* accumulating all of the above (**Figure 11**).[188]

The genus *Forsythia* was named after the Scottish botanist, William Forsyth (1737–1804). All but one of the known native species originates from northeastern Asia: Japan (*F. japonica*), Korea (*F. ovata*, *F. saxatilis*, *F. koreana,* and *F. densiflora*), and China (*F. giraldina*, *F. likiangensis*, *F. mandschurica*, *F. mira*, *F. suspensa,* and *F. viridissima*). One, *F. europaea*, is, however, endemic to Europe.[228] Other cultivars have been developed from these species, and *F. intermedia* was considered to be a hybrid between *F. suspensa* and *F. viridissima* based on the lignans isolated.[188] A recent phylogenetic study based on chloroplast DNA variation though suggests this not to be the case: *F. intermedia* does not group with either *F. suspensa* or *F. viridissima*, but forms a clade with *F. koreana*, *F. mandschurica*, and *F. saxatilis*.[228]

Initially, presumed *F. suspensa* stem cell wall residue preparations were demonstrated to engender stereoselective coupling of coniferyl alcohol (**91**) to afford (+)-pinoresinol (**13a**).[229] Subsequent solubilization and fractionation of the crude cell wall protein extract ultimately led to the isolation of a (+)-pinoresinol-forming dirigent protein (DP) named from the Latin *dirigere*, to guide or align.[230] This protein influenced the outcome of

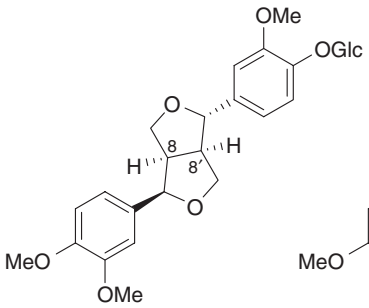
Linum flavum



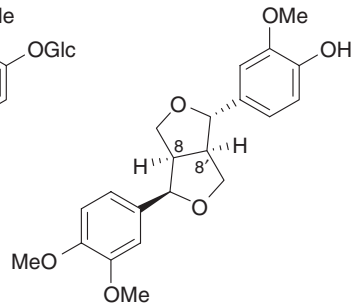
(**194b**) (–)-6-Methoxypodophyllotoxin

Figure 10 (–)-6-Methoxypodophyllotoxin (**194b**) in *Linum flavum*. Photograph by Laurence B. Davin, Washington State University, USA.

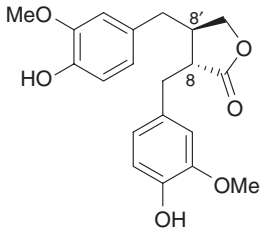
Forsythia intermedia



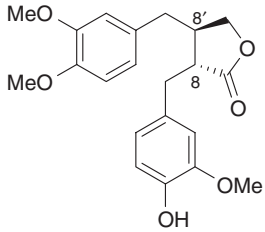
(**195a**) (+)-Phillyrin



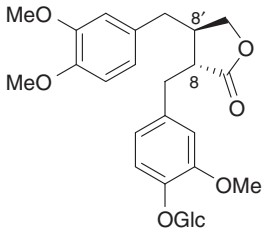
(**196a**) (+)-Phillygenin



(**10b**) (–)-Matairesinol



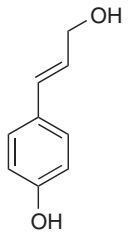
(**197b**) (–)-Arctigenin



(**198b**) (–)-Arctiin

Figure 11 Lignans in *Forsythia intermedia*. Photograph by Laurence B. Davin, Washington State University, USA.

coniferyl alcohol (**91**) radical–radical coupling reactions in the presence of an external one-electron oxidizing agent (such as laccase, ammonium peroxydisulfate, FMN). SDS–PAGE indicated a DP monomeric size of ~26–27 kDa.[230] In the absence of the DP, however, coniferyl alcohol (**91**) gave rise only to racemic mixtures of (±)-pinoresinols (**13a/b**), (±)-dehydrodiconiferyl alcohols (**14a/b**), and (±)-*erythro*/*threo* guaiacylglycerol 8–*O*–4′ coniferyl alcohol ethers (**115a/b**) due to nonregiospecific coupling. The (+)-pinoresinol-forming DP was next established to engender the formation of (+)-pinoresinol (**13a**) in a concentration-dependent manner, that is, higher DP concentrations *in vitro* led to a larger e.e. of (+)-pinoresinol (**13a**), instead of the other possible dimeric products.[231] Interestingly, the monomers used by the (+)-pinoresinol-forming DP did not appear to be seamlessly interchangeable, since neither *p*-coumaryl (**199**) nor sinapyl (**94**) alcohols afforded stereoselectively coupled products in *in vitro* assays.[230]



(**199**) *p*-Coumaryl alcohol

After amino acid sequencing, the corresponding (+)-pinoresinol-forming DP gene (*Fi_DP1*) was isolated from an *F. intermedia* cDNA library of young green stems using a PCR-amplified DNA probe. Analysis of the cloned gene (encoding an ~18.3-kDa predicted protein) suggested that the protein was glycosylated by posttranslational modification and contained a secretory system signal peptide.[232]

Heterologous expression of the corresponding Fi_DP1 protein was then next performed in a eukaryotic system (baculovirus-infected *Drosophila* cells) and, distinct from the native protein, three bands of DP

recombinant protein were visible by SDS–PAGE analysis. This indicated the presence of differentially glycosylated peptides, with these ranging from ~22 to 26 kDa.[232] Nevertheless, the recombinant DP was capable of engendering stereoselective coupling of coniferyl alcohol (**91**), in the presence of an one-electron oxidase/oxidant, to afford (+)-pinoresinol (**13a**).

1.23.4.2 Western Red Cedar Dirigent Proteins

As noted earlier, western red cedar (*T. plicata*) differentially accumulates various 8–8′-linked (−)-plicatic acid (**112b**)-derived (poly) lignans in its tissues, such as needles, stem, and bark, that are considered to be derived from (+)-pinoresinol (**13a**). This differential deposition of lignans in different tissues, which is particularly notable during the transition of its sapwood into heartwood, suggested the existence of distinct metabolic networks involving DPs in various tissues, that is, thereby orchestrating differential expression of the orthologous genes in monolignol radical–radical coupling and downstream metabolism. As a prelude to deciphering the biochemical pathway to (−)-plicatic acid (**112b**) in *T. plicata*, it was thus established that there were nine *DP* genes (*Tp_DP1–9*) having 72–99.5% identity to each other.[233] As for the *Forsythia* DP, they encoded proteins of 180–183 amino acids with each having a predicted molecular mass of ~20 kDa including the signal peptide. Several of these DPs (e.g., Tp_DP5 and Tp_DP8) were demonstrated *in vitro* to engender stereoselective coupling of coniferyl alcohol (**91**) to afford (+)-pinoresinol (**13a**), in the presence of an one-electron oxidase, indicative of the existence of a multigene family.[233]

1.23.4.3 Structural and Mechanistic Studies

The *Forsythia* DP exists apparently in dimeric form, based on the observations made using MALDI-TOF and ESI–MS, analytical ultracentrifugation, sedimentation velocity, and sedimentation equilibrium techniques.[234] It also had a propensity to further aggregate into ~12–18-mers, although this was prevented at increased NaCl concentrations. Additionally, application of circular dichroism (CD) demonstrated that the DP consisted mainly of β-sheet (35–42%) and loop (40–47%) secondary structures (**Figure 12(a)**), this being further supported by *in silico* modeling of the DP secondary structure (e.g., using PSIRED[235] and Folding@home (http://folding.stanford.edu)) (**Figure 12(b)**). Based on subsequent kinetic data and modeling, a steady-state kinetic model for the action of *F. intermedia* DP was proposed (**Figure 12(c)**), whereby the actual binding/coupling substrate was postulated to be a coniferyl alcohol (**91**)-derived free radical in solution ($CA^{\bullet}$).[231] In this proposed model, each DP monomer competes for binding of a $CA^{\bullet}$ with an apparent K_m of about 10 nmol l^{-1} relative to other diffusion-limited reactions (e.g., which would lead instead to the formation of racemic lignans). Upon binding of each $CA^{\bullet}$ to each monomeric DP (and possibly relative stabilization of the radical intermediate), the resulting proteinaceous dimeric complex is orientated in such a way that the two radicals approach each other from their *si–si* faces, thereby accounting for the stereoselectivity observed (**Figure 13(a)**).

1.23.4.4 Discovery of the (−)-Pinoresinol-Forming Dirigent Protein and Encoding Gene

Although *Forsythia* and *Thuja* spp. produce and utilize (+)-pinoresinol (**13a**), other species, for example, *Daphne tangutica*[236] and *A. thaliana*,[237] can accumulate the opposite (−)-enantiomer (**13b**) and/or downstream metabolites thereof. Indeed, *A. thaliana* contains a recently characterized DP homolog that preferentially forms (−)-pinoresinol (**13b**) from coniferyl alcohol (**91**) *in vitro*, again in the presence of a one-electron oxidase/oxidant (K. W. Kim, unpublished results). The corresponding *At_DP* gene encodes for a 21.4-kDa peptide (187 amino acids), and this finding now establishes the existence of distinct (+)- and (−)-pinoresinol (**13a** and **13b**)-forming DPs. Analogous to the *Forsythia* DP, generation of (−)-pinoresinol (**13b**) from coniferyl alcohol (**91**) is DP concentration-dependent. As before, the (−)-pinoresinol (**13b**)-forming DP does not affect substrate oxidation/coupling rates, and the protein lacks a catalytically active redox center. In this case, however, the two $CA^{\bullet}$ approach each other from their *re–re* faces (**Figure 13(b)**), thereby affording (−)-pinoresinol (**13b**) rather than the (+)-enantiomeric form (**13a**). It will next be of interest to

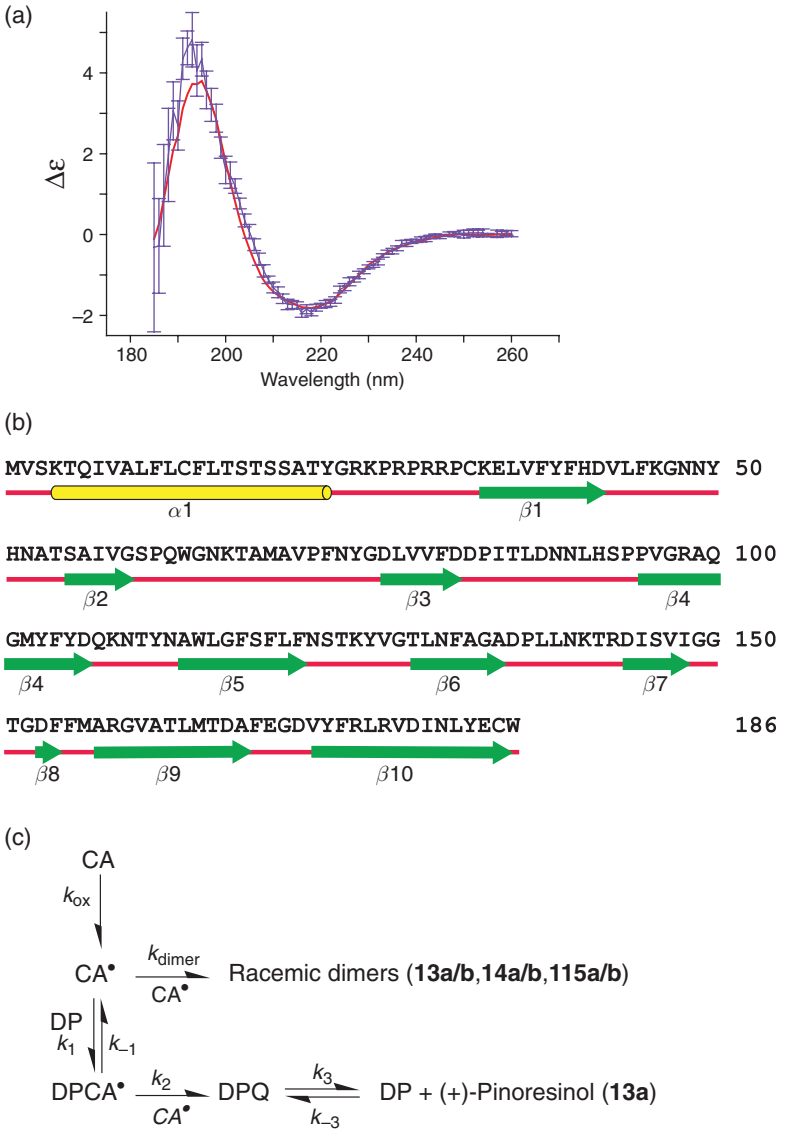


Figure 12 (a) CD spectrum of *F. intermedia* (+)-pinoresinol-forming DP. An expected fit (red) to the observed CD spectrum (purple) is shown for a protein with secondary structural components of 40–47% loop, 35–42% β-sheet, 9–14% turn, and 5–12% α-helix. Adapted with permission from S. C. Halls; N. G. Lewis *Biochemistry* **2002**, *41*, 9455–9461. Copyright 2002 American Chemical Society. (b) Predicted secondary structure of *F. intermedia* (+)-pinoresinol-forming DP using PSIRED server.[235] Yellow cylinder = α-helix, green arrow = β-strand and red bar = coil. (c) Proposed kinetic model for (+)-pinoresinol-forming DP.[231] Abbreviations: CA, coniferyl alcohol (**91**); CA•, coniferyl alcohol radical; DP, dirigent protein; DPCA•, dirigent protein–coniferyl alcohol radical complex; DPQ, dirigent protein quinone–methide intermediate complex; k_{ox}, rate constant of coniferyl alcohol (**91**) oxidation; k_1, rate constant of coniferyl alcohol radical (CA•) binding to DP, k_2, rate constant of second coniferyl alcohol radical (CA•) binding to the DPCA• complex; k_3, rate constant of release of (+)-pinoresinol (**13b**) from DP; k_{-1} and k_{-3} are the corresponding reverse rate constants to k_1 and k_3, respectively. Reproduced with permission from S. C. Halls; L. B. Davin; D. M. Kramer; N. G. Lewis, *Biochemistry* **2004**, *43*, 2587–2595. Copyright 2004 American Chemical Society.

establish fully the structural basis for the differential formation of (+)- and (−)-pinoresinols (**13a** and **13b**). (+) and (−)-Pinoresinol-forming DPs both share high amino acid sequence homology with each other (i.e., ~54% identity, ~70% similarity, not including the signal peptide), perhaps suggesting that only a few amino acids are involved in defining the different stereoselectivities in the active site(s).

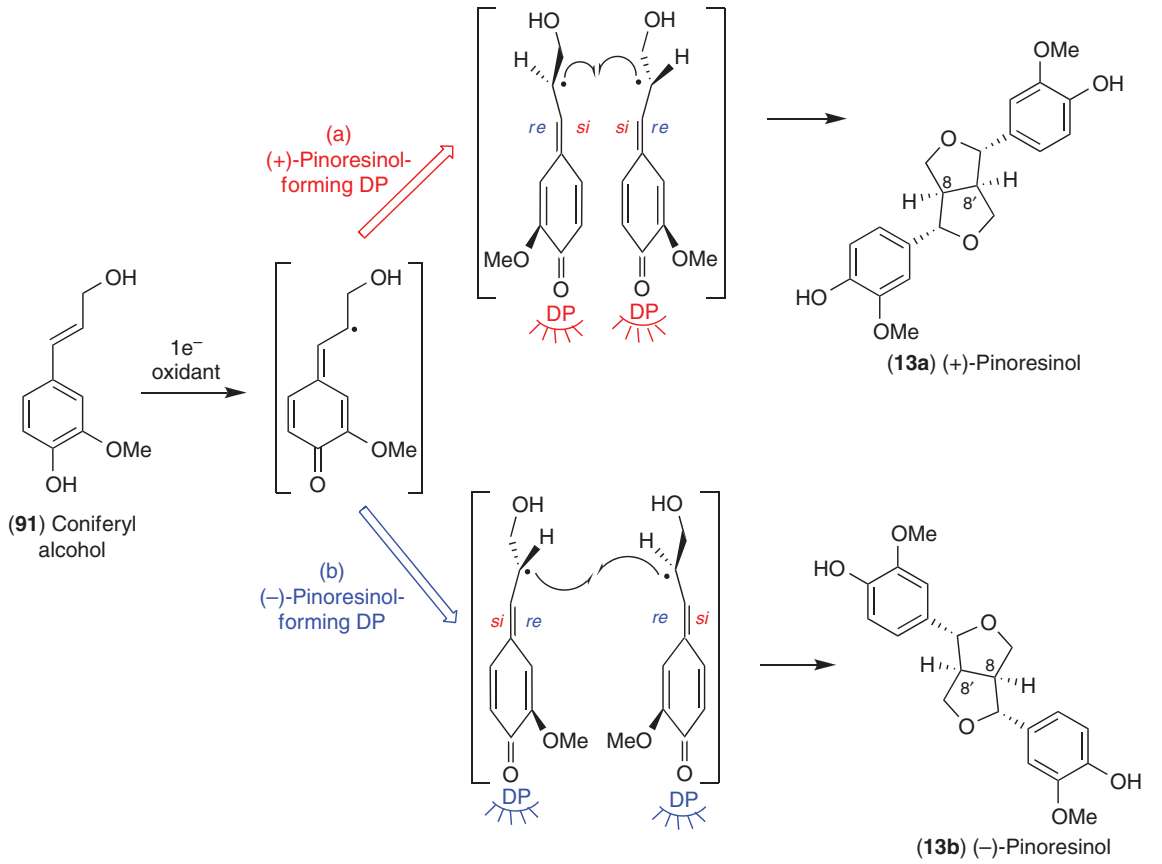


Figure 13 Stereoselective coupling of coniferyl alcohol (**91**) in presence of (a) (+)-pinoresinol- and (b) (–)-pinoresinol-forming dirigent proteins (DPs).

1.23.4.5 Dirigent Protein Tissue Localization and Metabolic Networks

1.23.4.5.1 mRNA tissue localization

Using *in situ* mRNA hybridization techniques, tissue-specific expression of the (+)-pinoresinol-forming DP gene was examined in *F. intermedia*.[238,239] Tissue-printing hybridization of fresh cross-sections of stems, petioles, and roots indicated that DP mRNAs were expressed in the vascular cambium regions of all tissues examined. *In situ* hybridization further confirmed the presence of DP mRNAs in the stem vascular cambium regions (**Figures 14(a)** and **14(c)**), as well as in the ray parenchyma cells adjacent to lignified tracheary elements in the youngest development stage examined (first internode, **Figure 14(b)**), that is, indicating sites of their lignan biosynthetic processes.[239]

Localization of DP gene transcripts was also investigated in western red cedar (**Figure 15**) using a single generic riboprobe to localize mRNA transcripts for all nine DP isovariants[240] (see Section 1.23.4.2). In sapwood, DP transcripts were detected in radial ray parenchyma cells (**Figure 15(c)**) and in vascular cambium; they were also detected in developing cells of cork cambium (**Figure 15(a)**).[240] Of particular interest, no hybridization was observed in the heartwood region under the conditions employed. By contrast, using the same technique, 18S rRNA transcripts (control) were detected in radial parenchyma cells of apparently preformed heartwood, as well as in all the other tissues where the DP was detected (not shown). Taken together, this was a most interesting finding. It demonstrated that the radial ray parenchyma cells were directly involved in the heartwood biosynthetic processes, which ultimately afforded the (–)-plicatic acid (**112b**)-derived lignans, and that other yet unknown biochemical processes were still occurring in the heartwood itself.

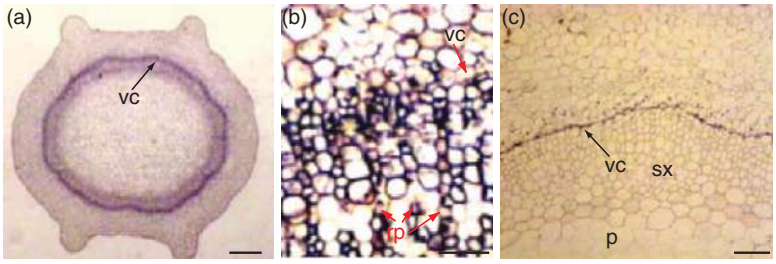


Figure 14 *In situ* hybridization of *Forsythia intermedia* DP mRNA with digoxigenin-labeled riboprobes in the first internode (a and b) and mature 10th internode (c). p, pith; rp, radial parenchyma; sx, secondary xylem; vc, vascular cambium. Bars: 150 μm (a), 25 μm (b), and 50 μm (c). Reproduced from V. Burlat; M. Kwon; L. B. Davin; N. G. Lewis, *Phytochemistry* **2001**, *57*, 883–897. Copyright 2001, with permission from Elsevier.

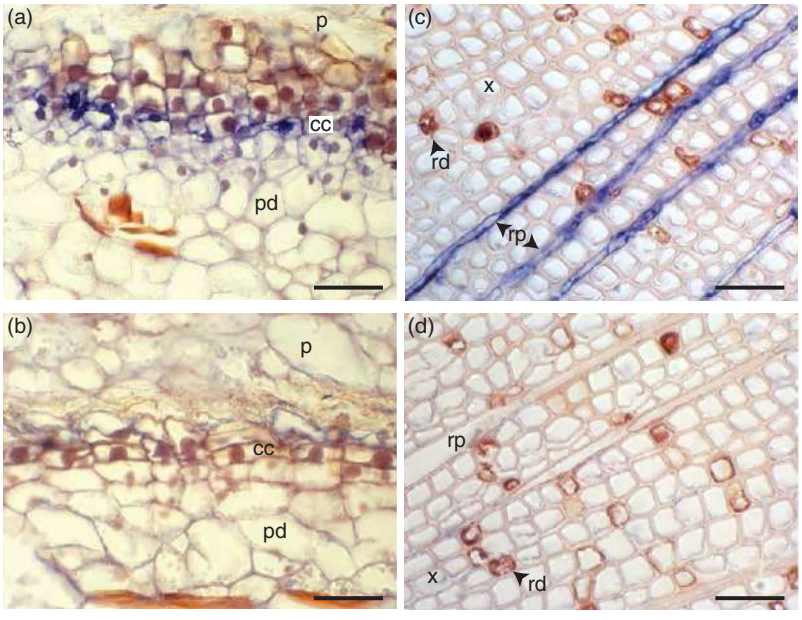


Figure 15 Detection of dirigent protein gene expression in young stem tissues of western red cedar by *in situ* hybridization. Hybridization of antisense probe for dirigent protein was detected in transverse sections by the blue color reaction in cells of cork cambium (a) and radial parenchyma (c) in young stem tissue (sapwood). A RNA probe of the sense strand of the dirigent transcript was used as a negative control (b, d). Abbreviations: cc, cork cambium; p, phellem; pd, phelloderm; rd, resin deposits; rp, radial parenchyma; x, xylem. Bars: 30 μm (a–d). Reproduced from A. M. Patten; L. B. Davin; N. G. Lewis, *Phytochemistry* **2008**, *69*, 3032–3037. Copyright 2008, with permission from Elsevier.

Heartwood tissue properties and the metabolites therein also provide generally a means of readily distinguishing between various woody plants. Thus, this ray parenchyma involvement provides additional insight into how this massive extrusion process occurs, and which is partially responsible for the often metabolite-specific heartwood formation. This may be of particular utility in understanding how the complex biochemical process involved in heartwood generation can be biotechnologically manipulated.

1.23.4.5.2 Dirigent protein tissue localization and proposed proteins harboring arrays of dirigent sites

Dirigent protein and proteins containing presumed arrays of dirigent sites (monolignol radical-binding sites) were also localized at the tissue and subcellular levels in *F. intermedia* using polyclonal antibodies raised against

the DP.[239,241] Overall, the patterns were quite similar to those for DP mRNA localization.[239] In the stems, labeling was localized to the vascular cambium region and young developing xylem, as well as in the cortex outer layers. As stem maturation proceeded, however, the label became restricted to the vascular cambium region (cambium and secondary phloem). Labeling was also mainly restricted to the cambial layers, secondary phloem, and the developing xylem in mature petioles, and to the pericycle layers/vascular tissues of the stele in the roots.

These observations were indicative of the cell/tissue types involved in lignan biosynthesis. Interestingly, high-resolution transmission electron microscopy also showed that in the stems, labeling was associated with the S1 sublayer and compound middle lamella of vessels, ray cells, and fibers, as well as to a lesser extent in their S3 sublayer. This latter set of observations is provisionally considered as indicative of detection of initiation sites for lignification, that is, sites harboring proteins containing arrays of presumed dirigent sites (see Chapters 5.01–5.21).

1.23.4.5.3 Proposed dirigent protein metabolic networks

There is additional evidence for the presence of DP metabolic networks in various plant species, such as western red cedar, spruce (*Picea* sp.), and *Arabidopsis*. This is contemplated even though only a relatively small number of DPs so far in their multigene families appear to be involved in stereoselective coupling to afford either (+)- or (−)-pinoresinols (**13a** or **13b**).

1.23.4.5.3(i) Western red cedar For western red cedar, the overall patterns of temporal and spatial expression of the nine DP isovariants (*Tp_DP1–9*) discussed previously above (Section 1.23.4.2) were analyzed by real-time (RT)-PCR and promoter analysis using the β-glucuronidase (GUS)-reporter gene in *Arabidopsis*.[242] Each DP ortholog was expressed differentially in individual organs, tissues, and cells at all stages of plant growth and development, indicative of the presence of a metabolic gene network. For example, *Tp_DP5* was only associated with the hypocotyl–root transition zone and the developing shoot meristem at 7–12 days old (**Figure 16(a)**), with *Tp_DP8* being strongly expressed throughout the vasculature (**Figure 16(b)**). In contrast, *Tp_DP2* was trichome- and root-specific (**Figures 16(c)** and **16(d)**). Differential expression patterns were also more pronounced in the reproductive tissues. For instance, *Tp_DP1* and *Tp_DP8* had distinct expression profiles in the flowers, that is intense GUS staining in the stamen filament, none in the anther (*Tp_DP1*,

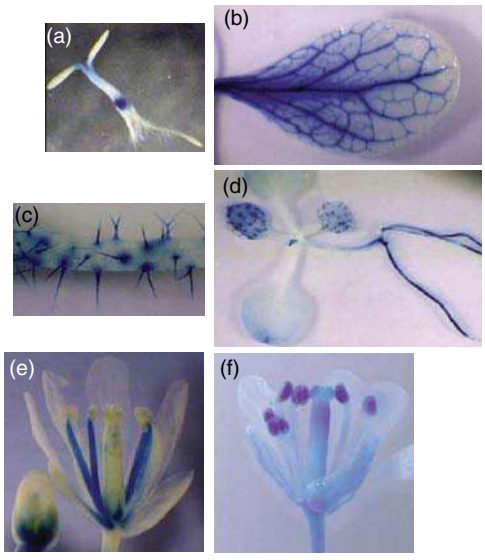


Figure 16 Histochemical localization of GUS activity in selected transgenic *Arabidopsis* plant lines containing various western red cedar *DP* promoter::*GUS* fusions.[242] (a) *TpDP5*gp::*GUS*. (b) *TpDP8*gp::*GUS*. (c and d) *TpDP2*gp::*GUS*. (e) *TpDP1*gp::*GUS*. (f) *TpDP8*gp::*GUS*. Reproduced from M. K. Kim; J.-H. Jeon; L. B. Davin; N. G. Lewis, *Phytochemistry* **2002**, *61*, 311–322. Copyright 2002, with permission from Elsevier.

Figure 16(e)), whereas *Tp_DP8* expression was weak in the stamen filament and strong in the pollen grain (**Figure 16(f)**), as well as in the silique valves (lignifying area, not shown).

As noted earlier, both Tp_DP5 and Tp_DP8 were able to help confer stereoselective coupling of coniferyl alcohol (**91**) to afford (+)-pinoresinol (**13a**), a precursor of (−)-plicatic acid (**112b**). Taken together, these data thus further suggest the existence of distinct metabolic networks involved in the regulation of lignan deposition in this species.

1.23.4.5.3(ii) **Picea *species*** Additional support for the existence of DP multigene families was established from the analysis of both expressed sequence tags (ESTs) and full-length cDNAs from three spruce species (*Picea sitchensis*, *P. glauca*, and *P. glauca* × *engelmannii*), which resulted in the detection of 35 DP and DP-like cDNAs.[243,244] Identities between their predicted amino acid sequences ranged from 99.5 to 17.6%, with predicted molecular masses from ∼17.4 to 21.7 kDa. Phylogenetic analyses showed that they clustered in three subfamilies (DP-a, DP-b, and DP-f), with each gene differentially expressed throughout the tissues analyzed (i.e., shoots, roots, cortex, phloem, cambium, xylem, and embryos). Although the DP-a group also clusters with known (+)-pinoresinol-forming DPs, both these and the others (DP-b and DP-f) currently have no established biochemical functions in all three species.

Nevertheless, a 16.7 k cDNA microarray with 30 ESTs representing at least 22 distinct DP/DP-like genes was used to examine expression profiles under several stress conditions. These included, among others, methyl jasmonate application, wounding, and weevil (*Pissodes strobi*) stem-boring herbivory attacks (alone and in combination), as well as in wood and apex development. Interestingly, upon analysis of the expression profiles, most of the DP/DP-like array elements clustered in accordance with their phylogenetic subfamilies (i.e., DP-a, DP-b, and DP-f). DP-a genes were apparently strongly induced in bark upon either wounding or weevil herbivory (with moderate to weak induction after other treatments). DP-b genes, on the contrary, were induced only weakly (if at all) upon stress treatments. By contrast, DP-f genes gave a more scattered induction/downregulation pattern upon stress treatments, perhaps indicative of more specialized individual functions. The transcript profiles so obtained thus suggest that spruce DP/DP-like genes, especially those from the DP-a subfamily, could play a significant role in constitutive and induced phenolic defense mechanisms against stem-boring insects. Others were speculated to be involved in defense against either pathogens and/or defoliating herbivores, or wounding, or in the formation of compression wood, or perhaps associated with tissue development; however, as indicated above, all of their precise biochemical roles and physiological functions speculated above[243,244] await elucidation.

1.23.4.5.3(iii) **Arabidopsis** There are also 16 isovariants of dirigent proteins harboring consensus regions involved in monolignol radical (and possibly other natural products) binding sites in *A. thaliana*. To date, three are known to be able to engender stereoselective coupling of coniferyl alcohol (**91**) *in vitro* in this species. Nevertheless, all 16 *Arabidopsis* isovariants were cloned (At5g42510, At5g42500, At5g49040, At2g21110, At1g64160, At4g23690, At3g13650, At3g13662, At2g39430, At2g28670, At1g22900, At4g11180, At4g11190, At4g11210, At4g38700, At4g13580), with corresponding promoters isolated and used to obtain transformed *GUS-DIR*/*GFP-DIR Arabidopsis* lines (K. W. Kim, unpublished results). The results obtained are again suggestive of the presence of a comprehensive (cell- and tissue-specific) network with each dirigent gene (homolog) having an unique pattern of expression 'in the vascular apparatus'. This, therefore, provides further potential insight into the presence of (partially overlapping) metabolic networks controlling various aspects of phenoxy radical coupling. For example, eleven of the sixteen genes (see *At_DP5*, *8*, and *10*, **Figures 17(a)**–**17(e)**) were expressed in the lignifying leaf vasculature, two at the base of lignifying leaf trichomes (see *At_DP3*, **Figures 17(f)** and **17(g)**), four in the lignifying leaf hydathodes (vasculature) (see *At_DP10*, **Figures 17(h)** and **17(i)**), twelve in the lignifying abscission zone of the siliques (e.g., *At_DP2*, **Figure 17(j)**), and twelve were differentially expressed in lignifying regions of the stem vasculature (i.e., protoxylem, vascular cambium, xylary and interfascicular fibers, etc (see *At_DP1*, *8*, *12*, and *13*, **Figures 17(k)**–**17(n)**)). These findings were thus in good agreement with those previously obtained using the DP nine-membered multigene found in western red cedar[242] (discussed above).

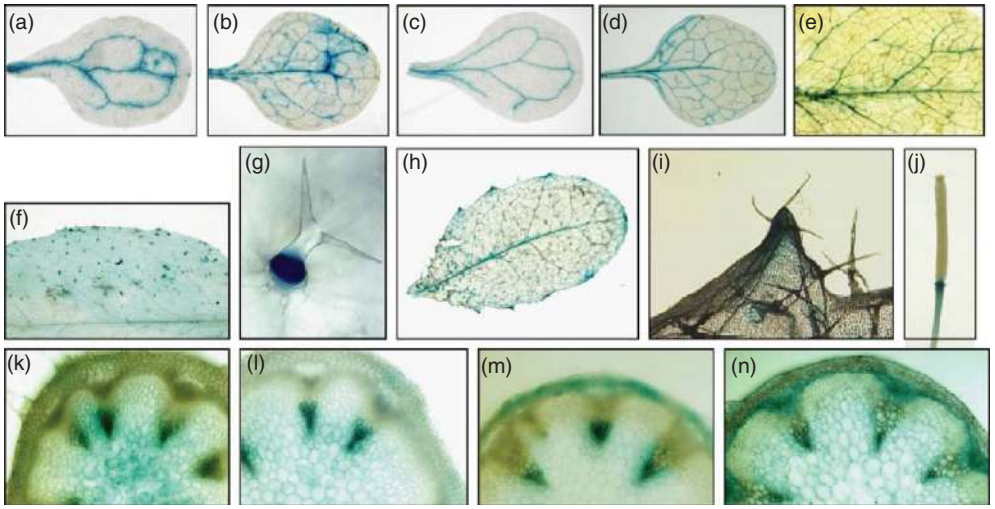


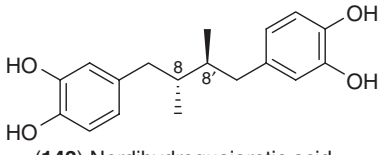
Figure 17 Selected patterns of expression of the 16-membered DP isovariant family using the GUS reporter gene. Cotyledons and leaf tissues at 3 weeks growth for At_DP5 (a, b) and At_DP10 (c, d) and leaf tissues at 4 weeks development for At_DP8 (e). Labeling is seen at the base of the trichomes for At_DP3 (f, g), in the leaf hydathodes for At_DP10 (h, i), in the silique abscission zone for At_DP2 (j) at 5 weeks growth, as well as in the lignifying regions of the stem vasculature for At_DP1 (k), 8 (l), 12 (m), and 13 (n) (Kye-Won Kim, unpublished results).

1.23.4.6 Other Examples of 8–8′ Phenylpropanoid Coupling: Hydroxycinnamic Acid and Allyl-/Propenylphenol-Derived Lignans in Liverworts and the Creosote Bush

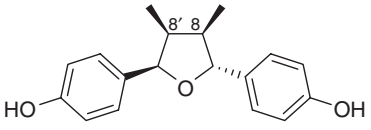
Examples of putative DPs engendering the formation of other types of 8–8′-linked lignans have also been suggested. These include the liverworts, *J. autumnalis* and *L. heterophylla*, whose intact cell cultures were shown to be able to metabolize [8-^{2}H]-caffeic acid (**56**) into either (+)-epiphyllic acid (**57a**) or its (−)-enantiomer (**57b**),[122] respectively, depending on the species. In *J. autumnalis*, the (+)-[8-^{2}H]-epiphyllic acid (**57a**) so formed can then be further metabolized to afford (+)-jamesopyrone (**62a**) and scapaniapyrone (**87**).[122] In either species, the corresponding chiral products that accumulate (i.e., (+)-jamesopyrone (**62a**) in *J. autumnalis* and (−)-epiphyllic acid (**57b**) in *L. heterophylla*) were enantiomerically pure, as determined by chiral HPLC analyses. On the contrary, incubation of caffeic acid (**56**) and H_2O_2 with cell-free extracts of both liverworts *in vitro* led only to the formation of racemic epiphyllic acid (**57a**/**b**). Thus, although the process controlling the proposed stereoselective coupling has not yet been detected; it is provisionally considered to involve DP control.[122] However, this still remains to be established.

As indicated above, 8–8′-regiospecific coupling also exclusively occurs during the formation of the 9–9′-deoxygenated lignans in the creosote bush (*L. tridentata*, Zygophyllaceae, **Figure 18**). This is a desert shrub of 1–3 m height, which is of increasing interest as a medicinal plant (discussed in Section 1.23.10.2). It has also long been used in Native American traditional medicine for treating more than 50 different ailments, including kidney and gall-bladder stones. Its most abundant lignan, NDGA (**143**), accumulates up to ∼5–10% of the leaves' dry weight, and is considered to be the creosote bush's main bioactive principle.[245] Many other 9–9′-deoxygenated lignans have also been detected in this species, the structurally simplest being the tetrahydrofurans (−)-larreatricin (**200b**), (−)-8′-*epi*-larreatricin (**201b**), *meso*-3,3′-didemethoxynectandrin B (**202**), and (±)-3,3′-didemethoxyverrucosins (**203a**/**b**).[246,247] Aryltetrahydronaphthalene lignans, such as norisoguaiacin (**204**), are also present, and apparently have the same C8/C8′ configurations as larreatricin (**200**).

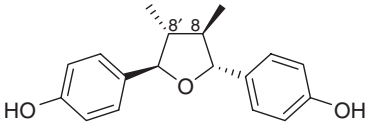
The 8–8′-linked lignans are thought to be derived from allyl-/propenylphenol-coupling products, for example, from *p*-anol (**21**) to afford initially (±)-larreatricins (**200a**/**b**) and its possible tetrahydrofuran ring diastereoisomers (**201**–**203**) with subsequent downstream metabolism then occurring (to generate NDGA (**143**), etc.). However, *p*-anol (**21**), when incubated *in vitro* with oxidases/laccases, generates a range of nonregiospecific coupling products. This, by contrast, does not occur *in vivo* in *Larrea*, suggesting the involvement of DP (or DP-like) proteinaceous control.



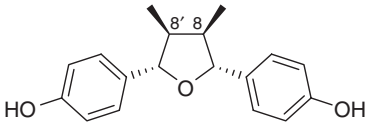
(**143**) Nordihydroguaiaretic acid



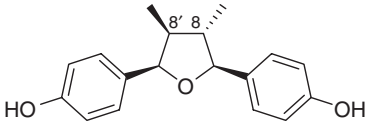
(**200b**) (–)-Larreatricin



(**201b**) (–)-8′-*epi*-Larreatricin

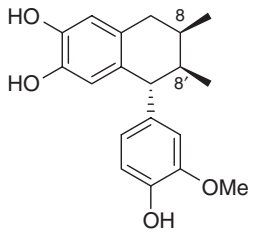


(**202**) *meso*-3,3′-Didemethoxynectandrin B



(**203a/b**) (±)-3,3′-Didemethoxyverrucosins

Larrea tridentata



(**204**) Norisoguaiacin

Figure 18 Various lignans in the creosote bush (*Larrea tridentata*). Photograph by Henri Moore, Washington State University, USA.

1.23.5 Downstream Lignan Metabolism

1.23.5.1 Furofuran Lignans in Sesame

1.23.5.1.1 Methylenedioxy bridge formation

Sesame (*S. indicum*), in addition to coconut, is one of the two oldest oilseed plants used by humanity,[248] with records dating back to about 6000 years.[249] The seed is highly valued as a source of oil, as well as for its antioxidant lignans.[250–252] There are contradictory reports though on the sesame plant origin: According to De Candolle,[253] it originated from the Sunda Isles in the Malay archipelago, following which it was introduced into India and the Euphrates valley to Egypt 2000 or 3000 years ago. Other records also suggest that it originated in the savanna of Central Africa and then spread to Egypt, India, the Middle East, and China.[249]

Sesame seed lignans are all 8–8′-linked and can contain either one or two methylenedioxy bridges (**Figures 19** and **20**). Sesame lignans can also be readily separated into those that are lipid-soluble, including (+)-sesamin (**11a**), (+)-piperitol (**205a**), and (+)-sesamolin (**210a**),[250,251] and those that are water-soluble, for example, (+)-sesaminol 2-*O*-mono- (**207a**), 2-*O*-di- (**208a**), and 2-*O*-tri- (**209a**) glucosides,[254] respectively. Of these, however, **11a** and **209a** are the most abundant lipid- and water-soluble lignans.[255,256] (+)-Sesamolin (**210a**) is of additional interest because of the unusual acetal oxygen insertion between the furanofuran group and the aryl moiety.

Sesame seedpods develop at different stages along the stems, with the oldest being closest to the base (**Figure 20**); the biosynthesis of its lignans is also developmentally regulated with seedpod development.[1,225] For example, at different developmental stages of the pods of 8-week-old plants, two interesting observations

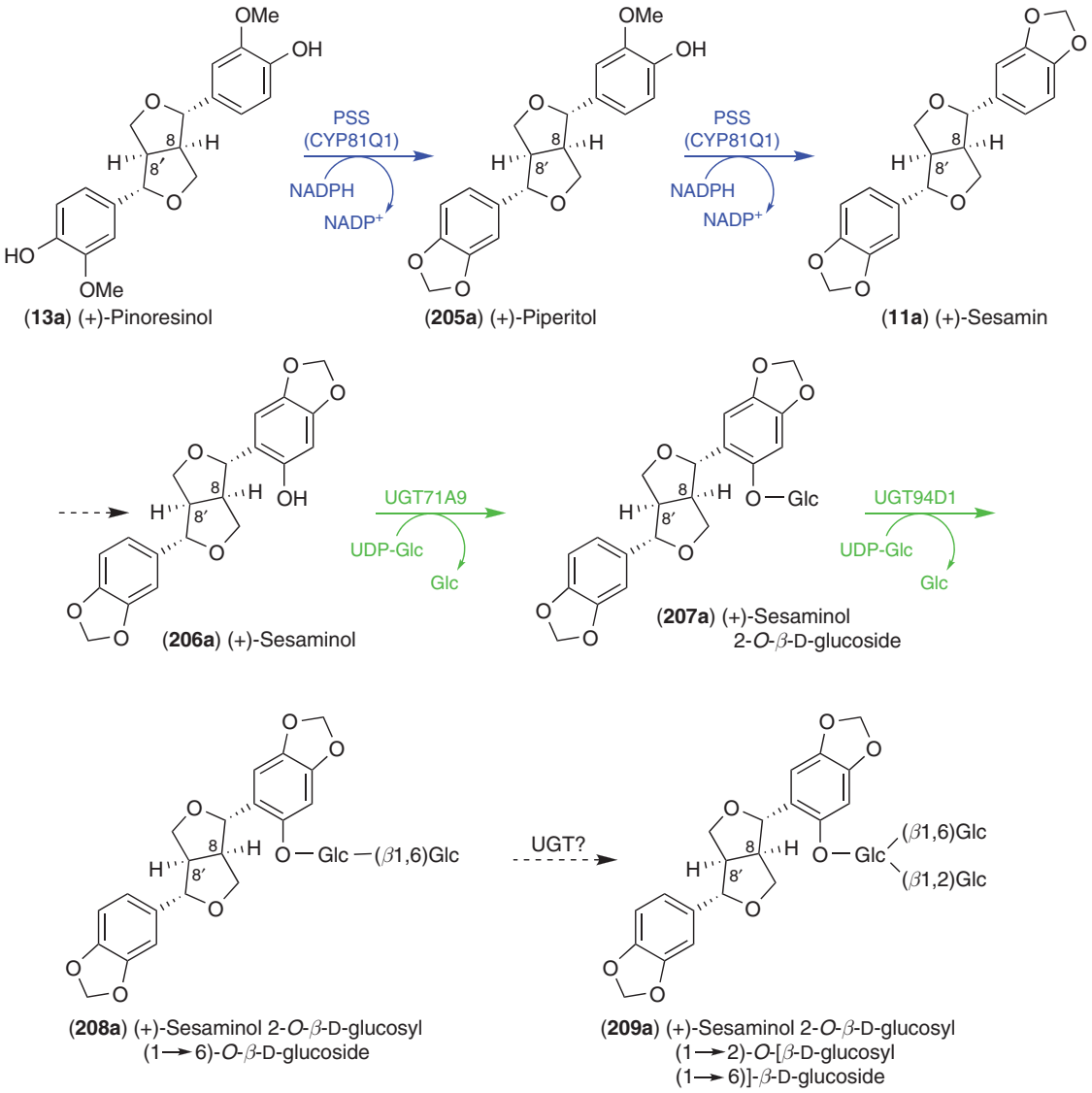
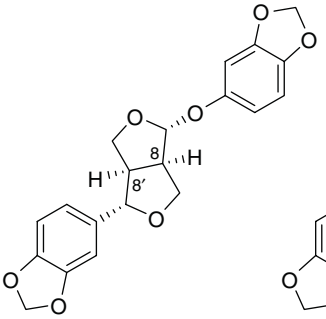


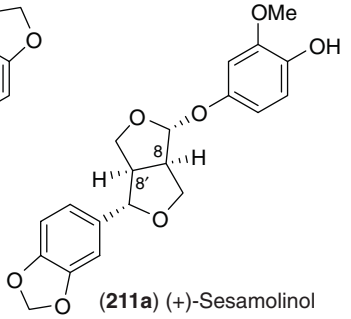
Figure 19 Proposed biosynthetic pathway to sesame (*Sesamum indicum*) lignans.

were previously made.[225] First, when (±)-[3,3′-$O^{14}CH_3$]-pinoresinols (**13a/b**) were administered to intact seeds, only the (+)-antipode (**13a**) was metabolized into the sesame lignans **11a**, **205a**, and **210a**, but not the corresponding enantiomer **13b** (**Figure 19**). Second, the relative efficacy of the incorporation of (+)-[3,3′-$O^{14}CH_3$]-pinoresinol (**13a**) into **11a**, **205a**, and **210a** varied with seed maturation stage.[225]

These findings were extended using microsomal preparations from the first and second stages of seedpod maturation. When incubated with (±)-[3,3′-$O^{14}CH_3$] pinoresinols (**13a/b**), only (+)-[3-$O^{14}CH_2$, 3′-$O^{14}CH_3$]-piperitol (**205a**) was formed when NADPH (1 mmol l^{-1}) was present. The corresponding (−)-enantiomer (**205b**), however, was not biosynthesized. This O_2-requiring, NADPH-dependent, cytochrome P-450 was thus subsequently named (+)-piperitol synthase.[225] Interestingly, although (+)-sesamin (**11a**) was not formed under the assay conditions used, incubation of (+)-piperitol (**205a**) as above resulted in its formation.[1] This, therefore, suggested the involvement of a second cytochrome P-450. In any event, these data established that both methylenedioxy bridges of sesame seed lignans resulted from cytochrome P-450-catalyzed transformations,[225] as had already been noted in alkaloid biosynthesis.[257]



(**210a**) (+)-Sesamolin



(**211a**) (+)-Sesamolinol

Sesamum indicum

Figure 20 (+)-Sesamolin (**210a**), (+)-sesamolinol (**211a**), and sesame (*S. indicum*). Photographs by Laurence B. Davin, Washington State University, USA.

Having thus established the overall enzymology and enantiospecificity of these transformations, subsequent cloning of a gene named *CYP81Q1* from a cDNA library obtained from sesame seed actively synthesizing sesamin (**11a**)[258] confirmed and further extended our original findings.[225] The corresponding protein, CYP81Q1, was heterologously expressed in yeast, with the resulting microsomal preparation individually incubated with (+)-pinoresinol (**13a**) and (+)-piperitol (**205a**) in the presence of NADPH. (Apparently, the corresponding (−)-enantiomers **13b** and **205b** were not, however, tested.) In agreement with our findings, (+)-sesamin (**11a**) was formed upon incubation with (+)-piperitol (**205a**), although it was also formed when (+)-pinoresinol (**13a**) was used as a substrate (**Figure 19**). This suggested that the CYP81Q1 was bi- and not monofunctional, and the name was expanded to (+)-piperitol/(+)-sesamin synthase (PSS) to indicate the bifunctional nature. Again, these data apparently confirmed our findings[225] of the enantiospecificity of methylenedioxy bridge formation and that a cytochrome P-450 was involved.

Identification of the reaction products was confirmed by LC–MS with m/z of 374 ($M + NH_4^+$) and 372 ($M + NH_4^+$) for **205a** and **11a**, respectively. Apparent K_m values were determined, using the yeast microsomal fraction, these being 10.2 and 11.7 μmol l^{-1} for (+)-pinoresinol (**13a**) and (+)-piperitol (**205a**), respectively. No other kinetic data were, however, reported. (+)-Sesamolinol (**211a**) was also not converted into (+)-sesamolin (**210a**) when incubated with this microsomal preparation,[258] perhaps suggesting the involvement of another cytochrome P-450 for the formation of its methylenedioxy bridge.

Two other *CYP81Q1* homologous genes, *CYP81Q2* and *CYP81Q3*, were also isolated from the related species, *Sesamum radiatum* and *Sesamum alatum*, respectively. These were of interest since the seeds of the former accumulate (+)-sesamin (**11a**), whereas those of the latter do not. After individual heterologous expression in yeast, microsomal preparations were again obtained. (+)-Piperitol (**205a**) and (+)-sesamin (**11a**) were formed when (+)-pinoresinol (**13a**) and NADPH were incubated with CYP81Q2, but not when CYP81Q3 was used.[258] This finding provisionally explains the absence of (+)-sesamin (**11a**) in *S. alatum*.

The report that CYP81Q1 apparently catalyzes dual methylenedioxy bridge formation is, however, in contrast to other cytochrome P-450's in alkaloid metabolism. The latter apparently only catalyze a single methylenedioxy bridge-forming reaction,[257,259] for example, the recently characterized CYP719A2 converted

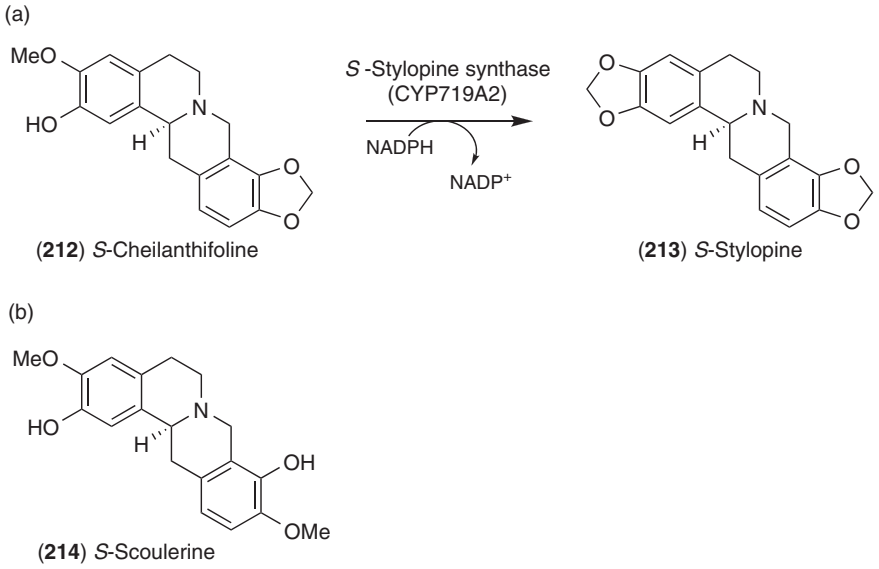


Figure 21 Methylenedioxy bridge formation in isoquinoline alkaloids. (a) Formation of (*S*)-stylopine (**213**) from (*S*) cheilanthifoline (**212**) catalyzed by (*S*)-stylopine synthase, a cytochrome P-450. (b) (*S*)-Scoulerine (**214**) did not serve as a substrate with (*S*)-stylopine synthase in *in vitro* assays.

S-cheilanthifoline (**212**) into *S*-stylopine (**213**) (**Figure 21(a)**), but not *S*-scoulerine (**214**, **Figure 21(b)**) into *S*-cheilanthifoline (**212**)[259] as demonstrated using crude microsomal preparations from *Eschscholzia californica*.[257]

It will be of interest in the future to resolve the ternary structure of this cytochrome P-450. This will hopefully provide further insight into the basis of its catalytic mechanism, and how substrate specificity is controlled, that is, including as to whether (+)-piperitol (**205a**) is first synthesized then released (or not) from the CYP81Q1 active site. If the former occurs (i.e., with product released from the active site), this may explain why the previous studies only detected (+)-piperitol (**205a**) formation when (+)-pinoresinol (**13a**) was incubated with a sesame seed microsomal preparation. That is, the released (+)-piperitol (**205a**) might have been too low in concentration in the assays to compete with the relatively large amounts of (+)-pinoresinol (**13a**) present under saturating conditions.

1.23.5.1.2 Glucosylation

UDP-glucose glucosyltransferases (UGT) presumed to be involved in the formation of (+)-sesaminol 2-*O*-triglucoside (**209a**), the most abundant water-soluble lignan in sesame seeds,[255,256] have also been characterized.[260] They were obtained by screening the sesame seed cDNA library with probes containing a well-conserved UGT sequence. After two rounds of screening, 10 clones were obtained, with each heterologously expressed in *Escherichia coli* as His-tag fusion proteins This resulted in the characterization of UGT71A9 and UGT94D1, which catalyzed the conversion of (+)-sesaminol (**206a**) into (+)-sesaminol 2-*O*-β-D-glucoside (**207a**), and the latter into (+)-sesaminol 2-*O*-β-D-glucosyl (1 → 6)-*O*-β-D-glucoside (**208a**) (**Figure 19**). In both cases, LC–MS confirmed identity of the reaction products: In the reaction catalyzed by UGT71A9, the product had a m/z 555.1450 $[M + Na]^+$ indicating that one glucose had been added to (+)-sesaminol (**206a**), which has a calculated mass of m/z 393.1 $[M + Na]^+$. In the reaction catalyzed by UGT94D1, by contrast, the enzymatically formed product **208a** had a m/z 717.2000 $[M + Na]^+$. The identity of both products was confirmed by ^1H, ^{13}C, and 2D-NMR spectroscopic analyses.

Kinetic parameters were determined for both glucosyltransferases. For UGT71A9, apparent K_m values were 6.32 and 41.0 μmol l^{-1} for (+)-sesaminol (**206a**) and UDP-glucose, respectively, with a k_{cat}/K_m value of 196 000 $mol^{-1}\,l\,s^{-1}$ (for **206a**). For UGT94D1, K_m values were 77.0 and 228.0 μmol l^{-1} for (+)-sesaminol

2-*O*-β-D-glucoside (**207a**) and UDP-glucose, respectively, but with a much lower k_{cat}/K_m of 11 700 mol^{-1} l s^{-1} (for **207a**). The UGT catalyzing the final $\beta 1 \rightarrow 2$ glucosylation to form (+)-sesaminol 2-*O*-triglucoside (**209a**, **Figure 19**) from **208a** was, however, not identified.

As for *CYP81Q1*, two homologs of *UGT71A9*, *UGT71A8*, and *UGT71A10* were cloned by RT-PCR from *S. alatum* and *S. radiatum*, respectively. The encoding proteins showed 98 and 91% similarity to UGT71A9. When the corresponding recombinant proteins were heterologously expressed in *E. coli*, both were able to enzymatically convert (+)-sesaminol (**206b**) into (+)-sesaminol 2-*O*-β-D-glucoside (**207a**), suggesting that this glucosyltransferase is conserved in the *Sesame* genome.

1.23.5.1.3 Oxygen insertion

The insertion of oxygen to form the acetal bridge between the furanofuran group and the aryl moiety during (+)-sesamolin (**210a**) formation still remains to be clarified,[261] as the precise sequence of oxygenation and methylenedioxy bridge formation is unknown. Interestingly, (+)-pinoresinol (**13a**), (+)-piperitol (**205a**), or, less likely, (+)-sesamin (**11a**) could all potentially serve as substrates for the oxygen insertion step. However, radiolabeled precursor administration experiments[225,262] did not resolve among these possibilities, and the enzyme participating in acetal bridge formation has not yet been identified.

(+)-Sesamolin (**210a**) is known to be rearranged into (+)-sesaminol (**206a**), however, during sesame oil processing.[252] This reaction is acid-catalyzed, with the acetal bridge being cleaved to afford the presumed intermediate oxonium ion and sesamol (**215**), which can then undergo nucleophilic attack (**Figure 22(a)**). In the presence of even a trace amount of water, however, both samin (**216**) and sesamol (**215**) are instead formed (**Figure 22(b)**).[252] Although the enzymatic formation of (+)-sesamolin (**210a**) has not yet been characterized, a possible mechanism for acetal bridge formation might involve rearrangement of (+)-sesaminol (**206a**) to afford (+)-sesamolin (**210a**) (**Figure 23**).

1.23.5.2 Pinoresinol/Lariciresinol Reductases and Pinoresinol Reductase

Discovery of genes encoding either (+)- or (−)-pinoresinol-forming DPs began to bring the much needed clarification to hitherto enigmatic differences in optical activities of various lignans from different plant species. This section thus focuses next on the discovery of pinoresinol/lariciresinol reductases (PLRs),[263,264] which catalyze conversions of pinoresinol (**13**) into either lariciresinol (**105**), secoisolariciresinol (**110**), or both. This includes discussion of their quite distinct enantiospecificities in various plant species, as well as progress made toward establishment of the operative biochemical mechanisms and the presumed involvement of enzyme-bound intermediary quinone methides. These studies also led to the discovery of related reductases, the 'provisionally' annotated phenylcoumaran benzylic ether reductases[265] (PCBERs, see Section 1.23.7.2), as well as the biochemical mechanisms/3D structures of PCBER and the related isoflavone reductases (IFRs).[266]

Partly due to serendipity, other PLR-related genes were also isolated in our investigations in 1999, but which instead encoded enzymes affording entry into the allyl-/propenylphenol pathways, for example, chavicol/eugenol synthases (CES).[2–4,267,268] Deduction of the allyl-/propenylphenol biochemical pathway came, however, from both delineation of PLR biochemical mechanisms and mechanistic considerations of substrate-to-product relationships in norlignan (*E*/*Z*-hinokiresinol (**123**/**174**)) biosynthesis (see Section 1.23.8). As far as PLRs proper are concerned though, most work has been carried out with *Forsythia*, western red cedar, *Linum* sp., and *A. thaliana* as summarized below.

1.23.5.2.1 Forsythia PLR: discovery of (+)-pinoresinol/lariciresinol reductase

The first-known PLR, discovered in *F. intermedia*,[263,269] catalyzes the sequential NADPH-dependent enantioselective conversion of (+)-pinoresinol (**13a**) into (+)-lariciresinol (**105a**), and then the latter into (−)-secoisolariciresinol (**110b**), respectively (**Figure 24**).[269] It was purified to apparent homogeneity and its measured kinetic parameters demonstrated that it efficiently reduced both substrates with apparent K_m, V_{max}, and k_{cat}/K_m values for **13a** and **105a** of 27/121 µmol l^{-1}, 4.5/7.0 pkat μg^{-1} protein, and 5800/

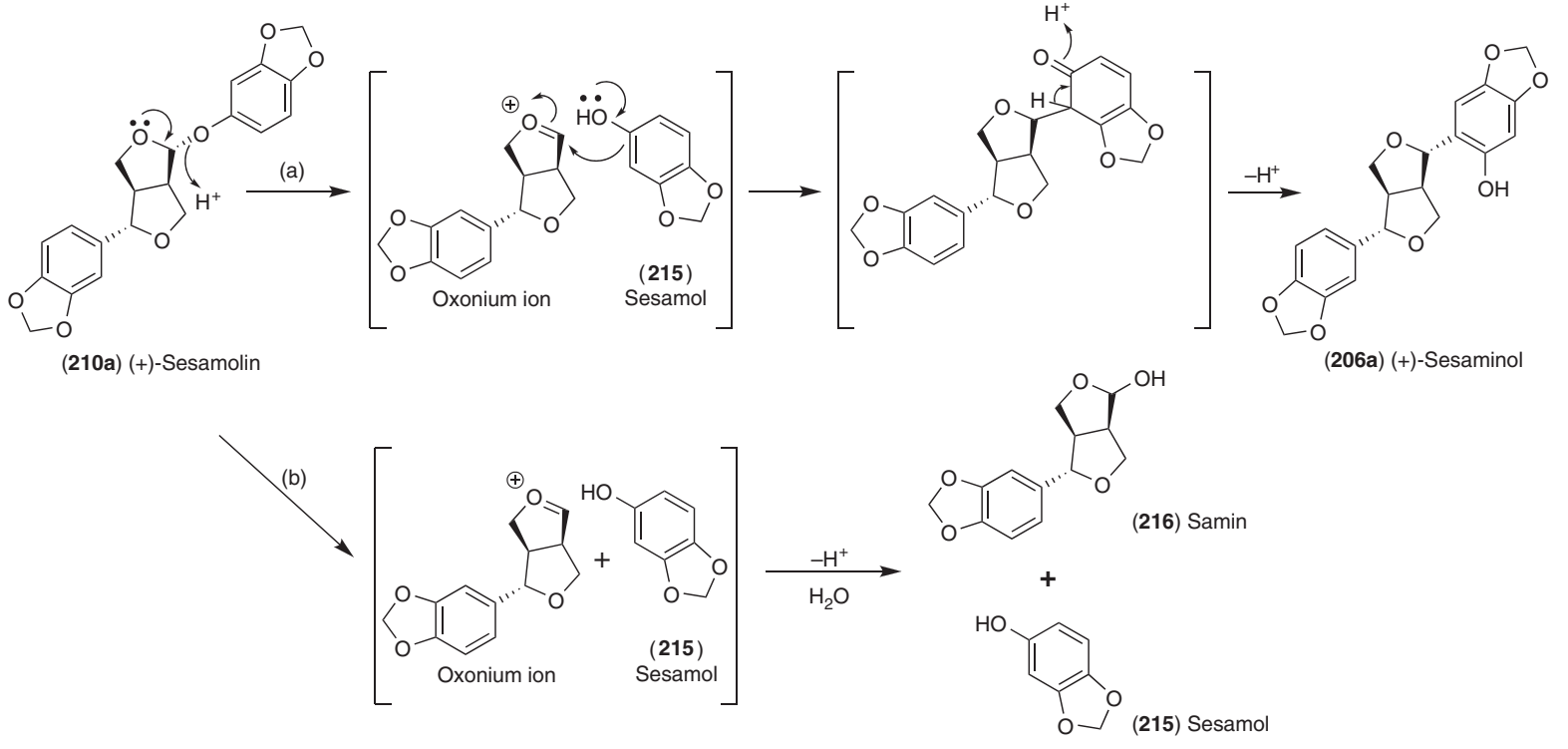


Figure 22 Proposed mechanism for the formation of (a) (+)-sesaminol (**206a**) via (+)-sesamolin (**210a**) rearrangement during sesame oil processing and (b) sesamol (**215**) and samin (**216**) from (+)-sesamolin (**210a**) in presence of H_2O.

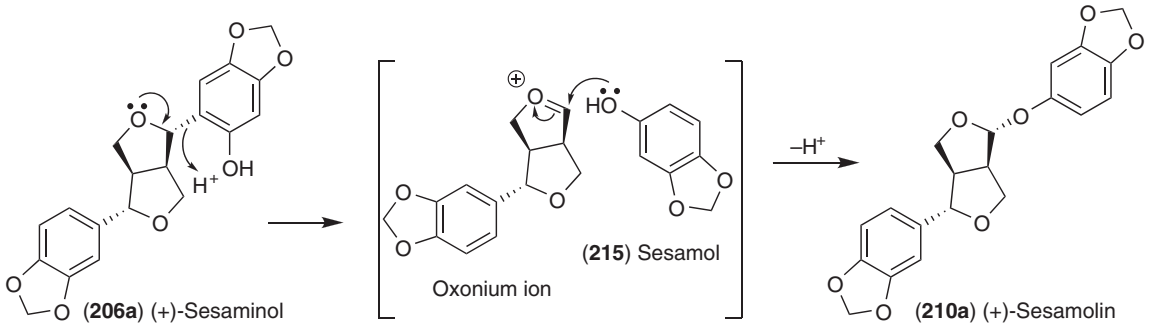


Figure 23 Putative rearrangement resulting in oxygen insertion to afford (+)-sesamolin (**210a**) from (+)-sesaminol (**206a**).

2000 mol^{-1} l s^{-1},[263] respectively. Kinetic data were thus reasonably consistent with enzyme turnover data for others in the general phenylpropanoid pathway to the monolignols **91**, **94**, and **199**.[7] The corresponding cDNA PLR_Fi1 was subsequently cloned,[263] with this encoding a polypeptide of 312 amino acids having a calculated molecular mass of 34.9 kDa.

1.23.5.2.2 Gymnosperm PLR/PLR homologs: discovery of PLR/PLR homologs of differing PLR and pinoresinol reductase enantiospecificities

PLR and/or PLR homolog cDNAs were also isolated from both western red cedar and western hemlock, that is, *PLR_Tp1–PLR_Tp4* and *PLR_Th1/PLR_Th2*, respectively, with the corresponding proteins having ~52–61% identity and ~66–79% similarity to PLR_Fi1.[264]

When heterologously expressed in *E. coli*, the western red cedar *PLR_Tp2* as well as the *PLR_Tp1* homolog (sharing >70% identity) gave recombinant proteins of very different enantiospecificities. As for the *Forsythia* PLR_Fi1, PLR_Tp2 efficiently converted (+)-pinoresinol (**13a**) into (−)-secoisolariciresinol (**110b**) using (+)-lariciresinol (**105a**). However, interestingly, PLR_Tp2 was also able to slowly reduce (−)-pinoresinol (**13b**) into (−)-lariciresinol (**105b**), albeit with the latter not further metabolized (**Figure 24**). This was indicative of distinct PLR and pinoresinol reductase (PR) activities, depending on the enantiomer present *in vitro*.

In contrast, the homolog PLR_Tp1 less efficiently catalyzed the NADPH-dependent conversion of (−)-pinoresinol (**13b**) into (−)-lariciresinol (**105b**), as well as (−)-lariciresinol (**105b**) into (+)-secoisolariciresinol (**110a**). (+)-Pinoresinol (**13a**) was also slowly converted into (+)-**105a**, but this was also not metabolized further (**Figure 24**). Moreover, kinetic analyses carried out using **13a**, **13b**, **105a**, and **105b** individually indicated that PLR_Tp2 was overall ~150-fold more catalytically efficient relative to the PLR_Tp1 homolog.[270] This was consistent with the proposed role of PLR_Tp2 in (−)-plicatic acid (**112b**) biosynthesis.

1.23.5.2.3 Linum species PLR: additional discovery of genes encoding (−)-PLR activity

The genus *Linum* in the family Linaceae consists of about 230 species,[271] which can be divided into six sections based on morphological characters:[272] *Linum*, *Dasylinum*, *Linastrum*, *Cathartolinum*, *Syllinum*, and *Cliococca*. Of the *Linum* species, flax (*L. usitatissimum*, 'the most useful' in Latin[273]) is the most common, which is grown for both its fibers (linen) and its seeds (oil and lignans). Indeed, it was apparently one of the first domesticated plants where the cultivation most likely began in the Fertile Crescent within the valleys of the Tigris and Euphrates some 8000 years ago.[274]

Flaxseed contains various ester-linked oligomers of the 8–8′-linked secoisolariciresinol diglucoside (SDG, **217**, **Figure 25**), which are covalently attached to hydroxymethyl glutaryl moieties (HMG, **218**),[275,276] such as the lignan **219**. SDG (**217**)/HMG (**218**)-containing flaxseed lignans have become somewhat better understood structurally in recent years, with several new features described. Although the macromolecular backbone is composed mainly of SDG (**217**) moieties ester-linked to HMG (**218**) molecules as indicated above,[275,276] the flavonoid herbacetin diglucoside (HDG, **220**) has been reported as a backbone component (estimated to be

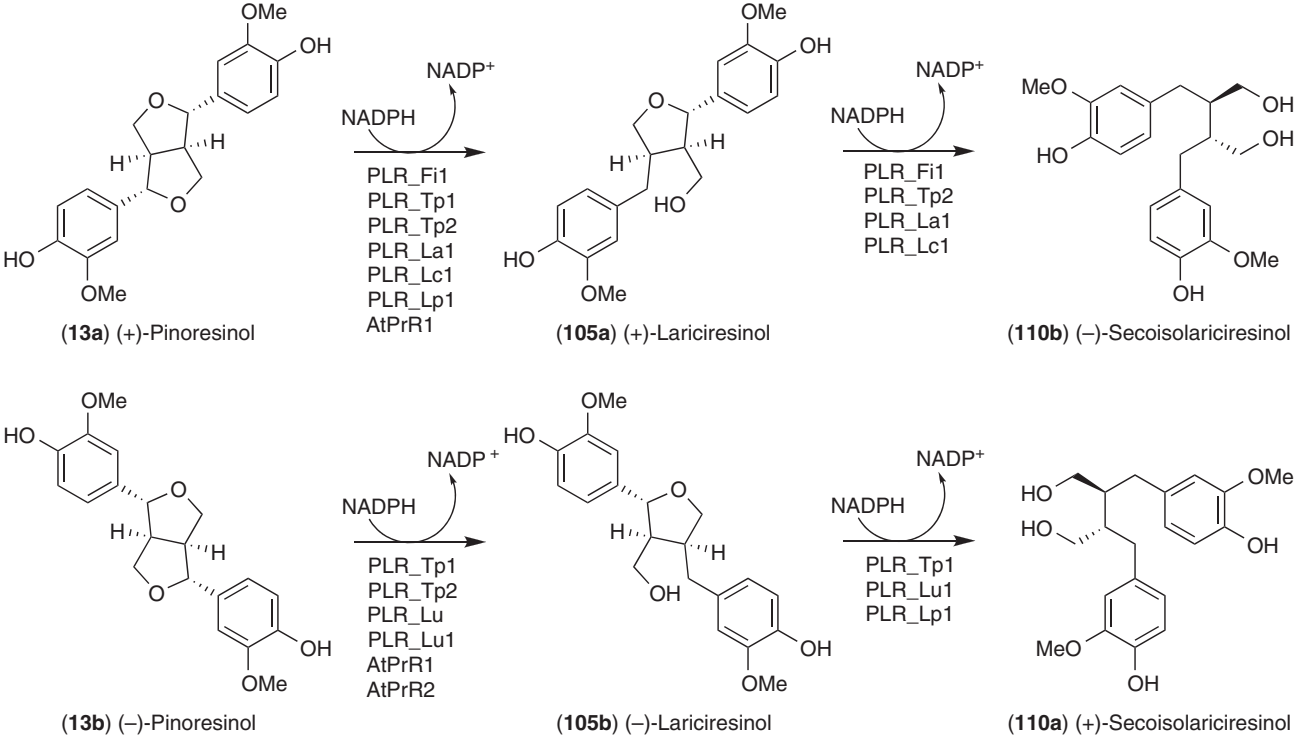
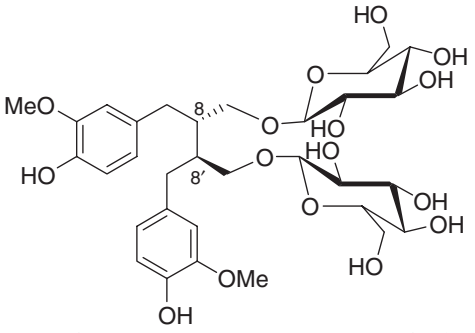
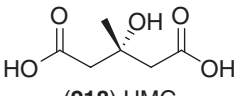


Figure 24 Reactions catalyzed by pinoresinol/lariciresinol reductases (PLRs), pinoresinol reductases (PRs), and homologs thereof. At, Arabidopsis thaliana; Fi, Forsythia intermedia; La, Linum album; Lc, L. corymbulosum; Lp, L. perenne; Lu, L. usitatissimum; Tp, Thuja plicata.

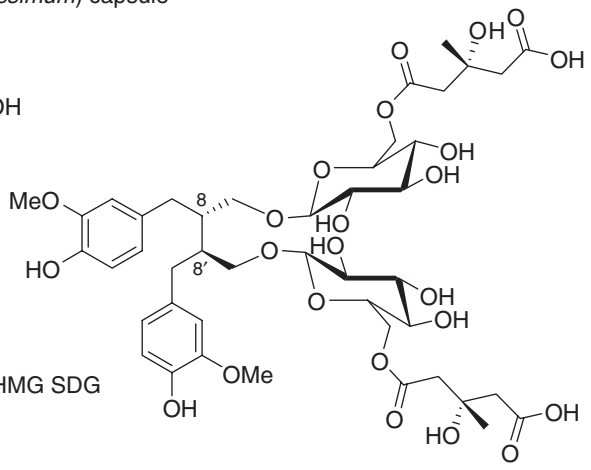
Flax (*Linum usitatissimum*) capsule



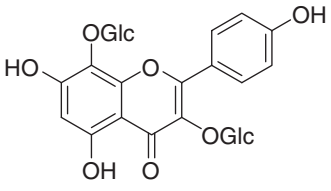
(**217**) Secoisolariciresinol diglucoside (SDG)



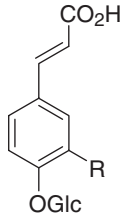
(**218**) HMG



(**219**) 6a,6a′ di-HMG SDG



(**220**) Herbacetin diglucoside (HDG)

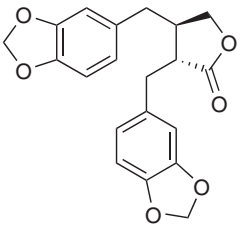


(**221**) R = H, *p*-Coumaric acid glucoside
(**222**) R = OMe, Ferulic acid glucoside

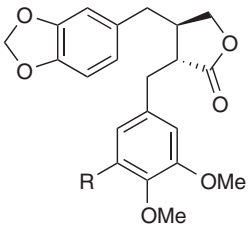
Figure 25 Various phytochemical constituents in flax (*Linum usitatissimum*). Photograph by Laurence B. Davin, Washington State University, USA.

~10 × lower in amount than that of SDG (**217**)).[277] The average flaxseed lignan 'macromolecule' was also recently provisionally estimated to contain ~3–5[276,278] (but varying between 1 and 7)[278] SDG (**217**) and/or HDG (**220**) backbone units. Both *p*-coumaric and ferulic acid glucosides (**221, 222**) are considered possibly to be linked as well,[279] but apparently as terminal (i.e., end) groups, with a negative correlation being noted between their amounts and the average length of the macromolecular chain.[278] The SDG (**217**) released from these conjugates upon alkali treatment mainly contains the (+)-secoisolariciresinol (**110a**) enantiomer, with the ratio of (+) (**110a**) to (−) (**110b**) being ~99:1.[275]

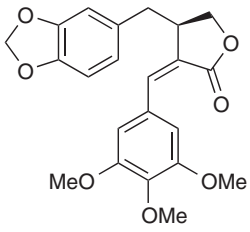
Interestingly, at full flower development, flax floral tissues reportedly also accumulate other 8–8′ dibenzylbutyrolactone lignans, such as (−)-hinokinin (**223b**), (−)-bursehernin (**224b**), (−)-yatein (**225b**), (−)-*E*-anhydropodorhizol (**226b**), (−)-matairesinol dimethyl ether (**227b**), and (−)-thujaplicatin trimethyl ether (**228b**).[280,281]



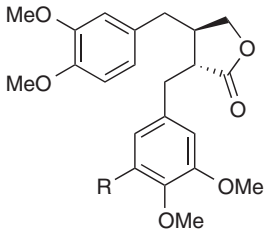
(**223b**) (–)-Hinokinin



(**224b**) R = H, (–)-Bursehernin
(**225b**) R = OMe, (–)-Yatein



(**226b**) (–)-*E*-Anhydropodorhizol



(**227b**) R = H, (–)-Matairesinol dimethyl ether
(**228b**) R = OMe, (–)-Thujaplicatin trimethyl ether

In terms of enzymology, PLRs of differing enantiospecificities have been very preliminarily characterized in four *Linum* species, that is, *L. usitatissimum* (PLR_Lu[282–284] and PLR_Lu1[285]), *L. album* (PLR_La1),[285] *L. corymbulosum* (PLR_Lc1),[286] and *L. perenne* (PLR_Lp1);[287] these are of ~60–74% identity and ~78–87% similarity to PLR_Fi1, respectively. Of these, PLR_Lu[282–284] and PLR_Lu1[285] were individually expressed heterologously in *E. coli.* and both catalyzed the conversion of (−)-pinoresinol (**13b**) into (−)-lariciresinol (**105b**) and, in one instance, of the latter into (+)-secoisolariciresinol (**110a**) as well. These data thus begin to provide a biochemical explanation for the predominance of the (+)-secoisolariciresinol (**110a**) enantiomer in flaxseed.

The other recombinant PLRs examined from *L. album* (PLR_La1)[285] and *L. corymbulosum* (PLR_Lc1)[286] apparently, also converted (+)-pinoresinol (**13a**) into (−)-secoisolariciresinol (**110b**), as described above for the *Forsythia* (PLR_Fi1) and western red cedar (PLR_Tp2) PLRs (**Figure 24**). These *Linum* PLRs, however, have not been subjected to detailed kinetic parameter characterization, and thus their relative enzymatic efficacies are as yet unknown. Nevertheless, the reported enantiospecific properties of the PLRs are consistent with the known optical activities of the isolated lignans: *L. album* accumulates (−)-podophyllotoxin (**1b**), (−)-6-methoxypodophyllotoxin (**194b**), and their derivatives thereof,[288] whereas (−)-hinokinin (**223b**)[289] is found in *L. corymbulosum.*

On the contrary, a report of a PLR from *L. perenne*, which accumulates justicidin B (**55**),[290] gave contradictory findings.[287] In assays carried out with (±)-pinoresinols (**13a/b**) in the presence of NADPH and increasing amounts of the recombinant PLR_Lp1, the enzyme reportedly preferentially utilized (+)-pinoresinol (**13a**) for the first reduction and (−)-lariciresinol (**105b**) for the next[287] (**Figure 24**). Such preliminary studies need, however, to be followed up with more conclusive and comprehensive kinetic parameter determinations. Accordingly, to establish the actual substrate specificities, pure enantiomeric substrates need to be used, that is, **13a**, **13b**, **105a**, and **105b**, rather than using racemic mixtures.

1.23.5.2.4 Arabidopsis *PLR homologs: Pinoresinol reductases*

Following β-glucosidase treatment of wild-type *A. thaliana* roots, small amounts of (−)-lariciresinol (**105b**) in ~88% e.e. were released.[237] To begin to provide clarification to this enantiomeric preponderance, two *Arabidopsis* PLR cDNAs (*AtPrR1* and *AtPrR2*) were obtained, whose corresponding proteins had ~59% identity and 75–78% similarity to PLR_Fi1. When incubated with (±)-pinoresinols (**13a/b**) *in vitro*, the recombinant AtPrR2 generated (−)-lariciresinol (**105b**) in 96% e.e., whereas AtPrR1 afforded both enantiomers **105a/b** with ~6% e.e. of the (+)-antipode **105a** (**Figure 24**). Interestingly, AtPrR1 was more catalytically efficient in *in vitro* assays, with a k_{cat}/K_m ~11–16 times higher for (±)-pinoresinols (**13a/b**) (12 800 and 19 000 $mol^{-1}\,l\,s^{-1}$ for **13a** and **13b**, respectively) than that of AtPrR2 for (−)-pinoresinol (**13b**) (1170 $mol^{-1}\,l\,s^{-1}$). On the contrary, the

predominance of (−)-lariciresinol (**105b**) *in vivo* was more in apparent agreement with the catalytic properties of AtPrR2.

To better understand the role of each gene in the formation of (−)-lariciresinol (**105b**), T-DNA insertion mutants were obtained, with homozygous lines further selected (i.e., *atprr1-1* and *atprr1-2* for *AtPrR1*, and *atprr2* for *AtPrR2*). A double mutant, *atprr1-1 atprr2*, was also generated. There was a significant increase (1.6–1.8×) in lariciresinol (**105**) content in the roots of the single mutants as compared to that of the wild-type line, with (−)-lariciresinol (**105b**) from *atprr1* and *atprr2* mutants having an e.e. of 94–96 and 82%, respectively. By contrast, lariciresinol (**105**) was not detectable in the double mutant *atprr1-1 atprr2*. Instead, it accumulated a relatively large amount of (−)-pinoresinol (**13b**) in 74% e.e. (~11 times WT levels of (±)-lariciresinols (**105a/b**)), this being in agreement with the preponderance of the upstream DP-mediated (−)-pinoresinol (**13b**) formation as previously discussed (Section 1.23.4.4).

1.23.5.2.5 Tissue localization of PLRs and PRs

Various studies have begun to establish tissue-specific spatial and temporal patterns of gene expression of both PLRs and PRs. These findings have also provided the required comparative insight with those of pinoresinol-forming DPs as well.

1.23.5.2.5(i)* In situ *hybridization of* Forsythia *PLR: Comparison with DP gene expression PLR_Fi1 localization in *F. intermedia* tissues was carried out using *in situ* hybridization.[291] As for the (+)-pinoresinol-forming DP, the strongest signal intensity was observed in the stems, as compared to leaves, petioles, and roots. For the stem tissues, the early development stage (first internode) gave labeling in the vascular cambium and primary phloem, respectively (**Figure 26(a)**), then later (i.e., from second to twentieth internode) in both ray parenchyma cells (**Figure 26(b)**) and vessels adjacent to ray parenchyma (twentieth internode, **Figure 26(c)**), in addition to the vascular cambium. For the roots, petioles, and leaves, PLR gene expression was only associated with meristematic tissues.[291]

When compared with (+)-pinoresinol-forming DP gene expression in *F. intermedia*, both PLR and DP mRNA expression were overall co-localized in the cambial regions (**Figures 14** and **26**). The DP mRNAs were also observed in ray parenchyma cells of the first stem internode stage (**Figure 14(b)**), whereas, by contrast, PLR mRNAs, were noted in these same cell types at later stages (**Figures 26(b)** and **26(c)**).[238,239] In western red cedar, however, the (+)-pinoresinol-forming DP gene expression was evident in both cambial regions, as well as ray parenchyma cells (**Figures 15(a)** and **15(c)**). Taken together, both the PLR and the DP thus appear to be restricted to the same cell types, although whether there are more subtle temporal differences in their individual gene expression patterns cannot be gauged at this time.

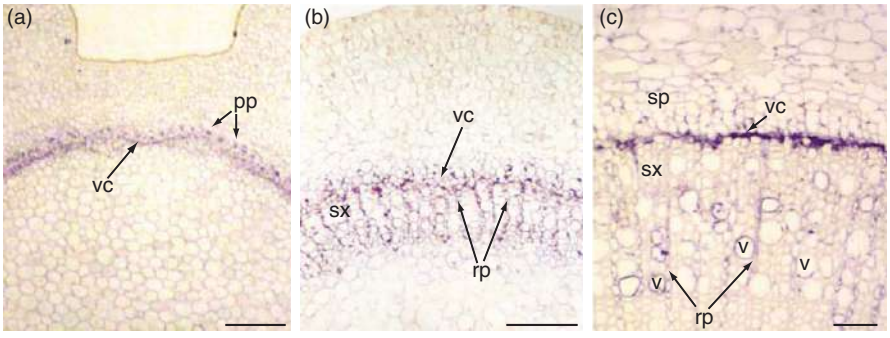


Figure 26 PLR_Fi1 mRNA accumulation in *F. intermedia* stems at different developmental stages as shown by *in situ* hybridization. (a) First stem internode. (b) Second stem internode. (c) Twentieth stem internode. PLR_Fi1 mRNA strongly accumulated in the vascular cambium regions (vc) (a–c), the ray parenchyma cells (rp) (b and c) and in the vessels (v) adjacent to the ray parenchyma cells (c). Abbreviations: pp, primary phloem; rp, radial parenchyma; sp, secondary phloem; sx, secondary xylem; v, vessel; vc, vascular cambium. Bars: 100 μm (a), 60 μm (b), and 300 μm (c). Reproduced from M. Kwon; L. B. Davin; N. G. Lewis, *Phytochemistry* **2001**, *57*, 899–914. Copyright 2001, with permission from Elsevier.

1.23.5.2.5(ii) Flax PLR gene expression Spatiotemporal expression of PLR in flax was also investigated using both semiquantitative RT-PCR and GUS reporter gene (using the *PLR_Lu* promoter) strategies.[292] The RT-PCR analyses indicated that PLR_Lu expression occurred in the seed coats at all five developmental stages (10, 16, 20, 24, and 35 days after flowering, respectively), with the highest level being observed at stage 3 of seed maturation. GUS staining was carried out on both vegetative and reproductive organs. As expected, strong tissue-specific expression occurred in developing seeds, but not in stems, leaves, and roots (which do not accumulate SDG-derived lignans to any considerable extent). Gene expression was mainly localized to the seed coat (**Figure 27**, seeds shown at stage 3 of maturation), but was not detectable in the embryo (**Figures 27(b)** and **27(c)**). The seed coat is the known site of accumulation of SDG-HMG lignans, such as **219**.[293]

1.23.5.2.5(iii) AtPrR localization in Arabidopsis Quantitative RT-PCR analyses have also been used to study expression levels of both *AtPrR1* and *2* genes in *Arabidopsis*, with both reportedly expressed at similar levels in the root tissues, and to a lesser extent, in stems.[237]

1.23.5.2.6 Structural biology studies: PLR and PLR homolog

Comprehensive structural biology studies of both PLR and a PLR homolog have provided the much needed insight into their overall catalytic mechanisms, as well as that of the PLR-related proteins (PCBERs, IFRs, and CESs, discussed later). The PLR and PLR homolog studies include establishing stereospecificity of hydride transfer resulting in an inversion of product configuration; evidence for presumed involvement of enzyme-bound quinone methide intermediates; and progress made toward an understanding of the biochemical basis for distinct enantiospecificities.

1.23.5.2.6(i) Stereospecificity of hydride transfer with resulting inversion of product configuration *Forsythia* PLR is a type A reductase, as established when using [4*R*-^{3}H]- and [4*S*-^{3}H]-NADPH as cofactor; only the *pro-R* hydrogen on the nicotinamide ring of NADPH was abstracted and transferred to form both (+)-lariciresinol (**105a**) and (−)-secoisolariciresinol (**110b**), respectively (**Figure 28**). Moreover, the incoming hydride also took up the *pro-R* position in the corresponding products. For example, when (±)-pinoresinols (**13a/b**) and (±)-lariciresinols (**105a/b**) were incubated with the *Forsythia* PLR in the presence of [4*R*-^{2}H]-NADPH, the enzymatic products were established to be (+)-[7′*R*-^{2}H]-lariciresinol (**105a**) and (−)-[7*R*,7′*R*-^{2}H]-secoisolariciresinol (**110b**), respectively. Furthermore, using (±)-[7,7′-^{2}H$_2$]-pinoresinols (**13a/b**) and [7,7′-^{2}H$_3$]-lariciresinols (**105a/b**) as substrates and unlabeled NADPH, NMR spectroscopic analyses of the resulting enzymatic products **105a** and **110b** also demonstrated that the incoming hydride took up the *pro-R* position in the corresponding substrate (**Figures 29(b)** and **29(d)**; the corresponding ^{1}H-NMR spectra of unlabelled **105** and **110** are shown for comparison in **Figures 29(a)** and **29(c)**). From these observations, it was demonstrated that an 'inversion' of configuration had occurred at C7 for (+)-lariciresinol (**105a**) and at C7/C7′ for (−)-secoisolariciresinol (**110b**), with the presumed intermediate(s) being the enzyme-bound quinone methide(s) (**Figure 28**).[269]

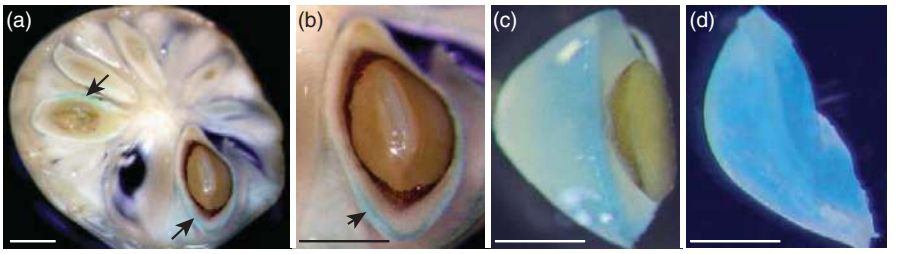


Figure 27 Histochemical localization of GUS in flax (*L. usitatissimum*) containing the *PLR_Lu* promoter. Cross-sections of: (a) Flaxseed capsule. (b and c) Seed showing seed-coat stained and unstained embryo. (d) Seed-coat. Bars represent 1 mm. Reproduced with permission from C. Hano; I. Martin; O. Fliniaux; B. Legrand; L. Gutierrez; R. R. J. Arroo; F. Mesnard; F. Lamblin; E. Lainé, *Planta* **2006**, *224*, 1291–1301. **Figure 4n–4q**. Copyright 2006.

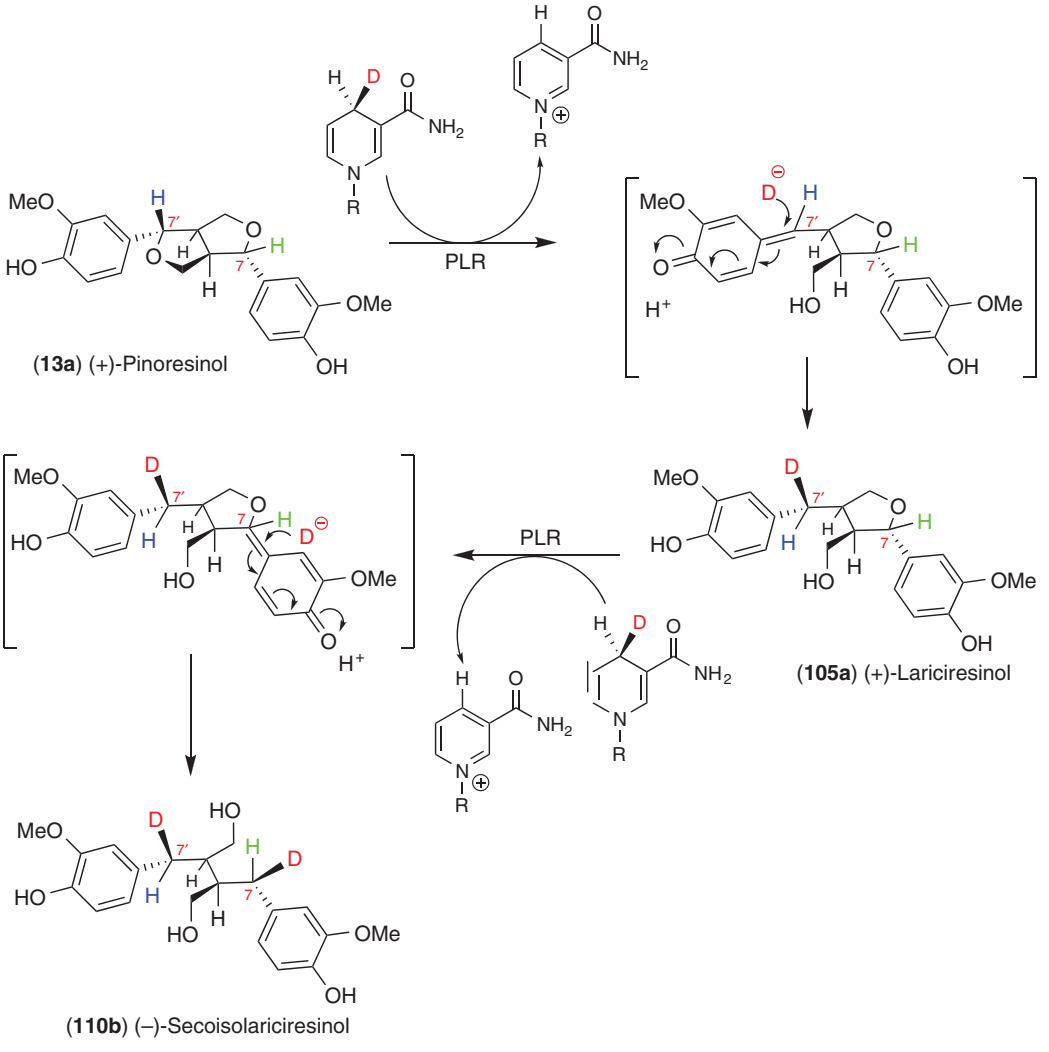


Figure 28 Stereospecificity of hydride transfer during *F. intermedia* PLR catalysis resulting in proposed formation of quinone methide intermediate from (+)-pinoresinol (**13a**) and inversion of product configuration for (+)-lariciresinol (**105a**) and (–)-secoisolariciresinol (**110b**). Adapted from A. Chu; A. Dinkova; L. B. Davin; D. L. Bedgar; N. G. Lewis, *J. Biol. Chem.* **1993**, *268*, 27026–27033.

1.23.5.2.6(ii) Structural biology/substrate versatility studies Based on sequence similarities (and, as discussed later, structural features as well), PLRs are classified within the somewhat diverse short-chain dehydrogenase/reductase (SDR) superfamily.[266] In general, most SDRs are about 250 amino acid residues in length, and share many structural similarities, for example, a conserved N-terminal nucleotide cofactor-binding sequence motif (GXXGXG, the so-called Rossman fold), and a conserved catalytic Lys residue. SDRs, in general, are able to process a wide range of substrates and may have as little as only 15–30% sequence identity. However, typically, their tertiary structures/folding patterns are quite similar.[294]

Of those involved in phenylpropanoid metabolism, the *T. plicata* PLR (PLR_Tp1) homolog (apo-form), as well as the modeled PLR_Tp2, were the first to have their X-ray crystal structures determined (at 2.5 Å resolution for PLR_Tp1, **Figure 30(a)**).[266] Both are dimers as demonstrated by sedimentation equilibrium experiments (with PLR_Tp1), which gave an apparent molecular mass of 69.9 ± 0.4 kDa, whereas MALDI-TOF analysis gave a distinct monomer at 35 096 *m*/*z*, as well as a broad less intense dimer centered at ~70 099 *m*/*z*.[266] PLR_Tp2 and the PLR_Tp1 homologs have two functional domains, a conserved nucleotide cofactor-binding N-terminal domain and a presumed substrate-binding C-terminal domain, with a deep cavity between

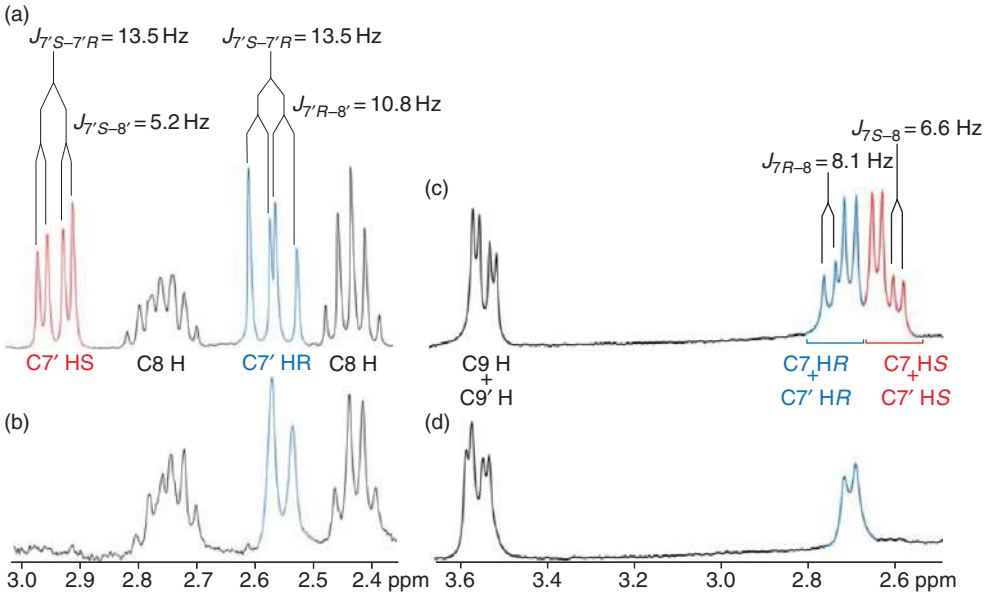
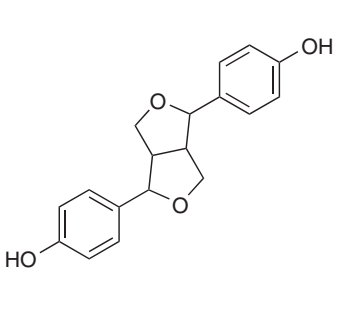


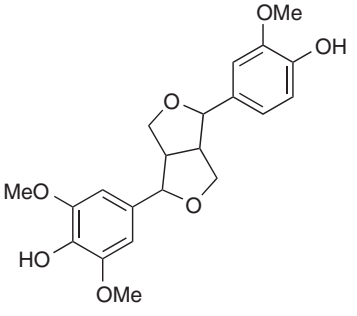
Figure 29 Partial ^{1}H NMR spectra of lariciresinol (**105**) and secoisolariciresinol (**110**) showing regions for C7′/C8/C8 and C7/C7′/C9 proton resonances, respectively. (a) Synthetic (±)-lariciresinols (**105a/b**). (b) Enzymatically synthesized (+)-[7,7′ *S*-$^{2}H_2$]lariciresinol (**105a**) obtained following incubation of PLR_Fi1 with (±)-[7,7′ $^{2}H_2$]pinoresinols (**13a/b**). (c) Synthetic (±)-secoisolariciresinols (**110a/b**). (d) Enzymatically synthesized (−)-[7,7′ *S*-$^{2}H_3$]secoisolariciresinol (**110a**) obtained following incubation of PLR_Fi1 with (±)-[7,7′ *S*-$^{2}H_3$]lariciresinols (**105a/b**). Redrawn from A. Chu; A. Dinkova; L. B. Davin; D. L. Bedgar; N. G. Lewis, *J. Biol. Chem.* **1993**, *268*, 27026–27033.

them. The highly conserved catalytic lysine residue (Lys138) was present in the active site and presumed to be involved in catalysis, this being further established by site-directed mutagenesis,[266] where, for example, the K138A mutant of PLR_Tp1 had its PLR activity abolished.

Substrate/cofactor docking *in silico* suggested that Lys138 was in close proximity to the phenolic group of the substrate. Lys138 was thus considered to act as a general base to initiate catalysis, abstracting the phenolic proton of pinoresinol (**13**) to facilitate the formation of the putative quinone methide intermediate[269] (**Figure 31(a)**). Hydride addition (from NADPH) at C7 then occurs to afford the first product, lariciresinol (**105**), which is envisaged to leave the active site, and then bind again through the phenolic moiety (attached to the remaining furan group), adjacent to Lys138. A second round of catalysis can then proceed, with the presumed intermediacy of a second quinone methide as before, and hydride addition at C7′ affording secoisolariciresinol (**110**). Quinone methide intermediacy was further supported by substrate versatility studies, where ligballinol (**238**), medioresinol (**239**), and syringaresinol (**96**) were processed as substrates, whereas sesamin (**11**) was not; that is, there was an apparent requirement for a free phenolic group in the substrate thereby enabling quinone methide formation as had been proposed earlier in 1993.[269]



(**238**) Ligballinol



(**239**) Medioresinol

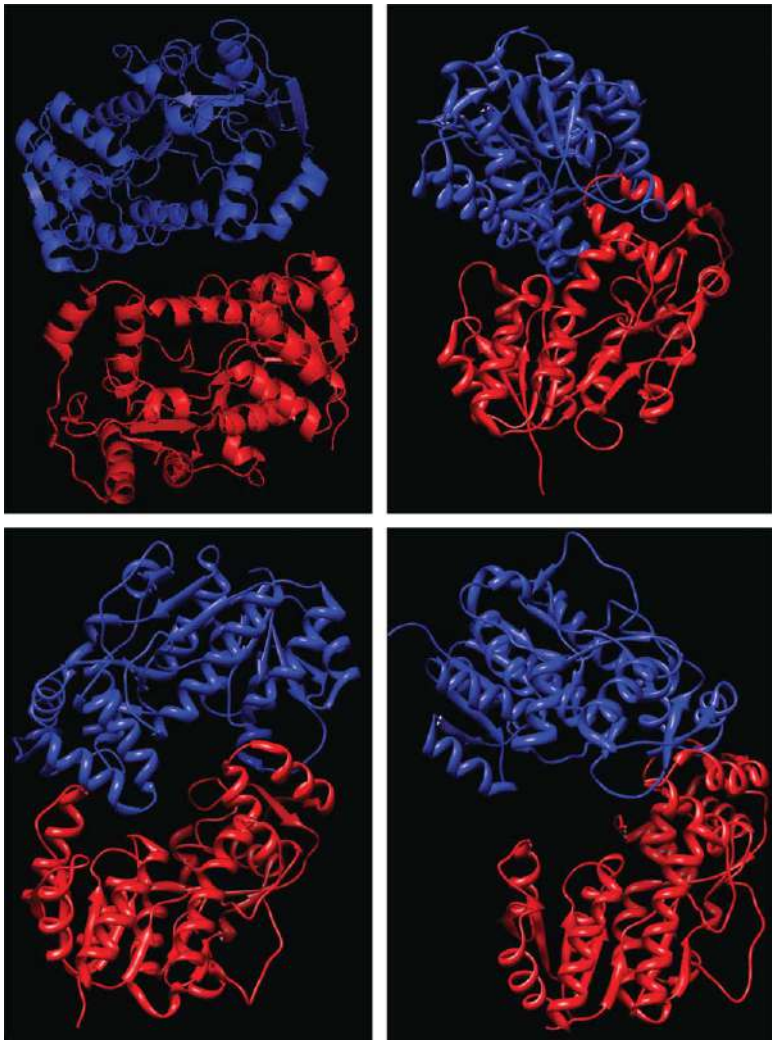


Figure 30 Schematic outlines of the dimeric forms of (a) *Thuja plicata* PLR_Tp1, (b) *Pinus taeda* PCBER_Pt1, (c) *Medicago sativa* IFR_Ms1, and (d) *Ocimum basilicum* EGS depicted in ribbon form by alignment of their twofold axes.

1.23.5.2.6(iii) PLR and PLR homolog enantiospecificity In general, PLR_Tp2 and PLR_Tp1 display opposite preferences with regard to substrate enantiomeric forms. From comparison of both structures, various potential (candidate) residues present in the corresponding active sites were considered as possibly involved in stipulation of the distinct enantiospecificities observed, that is, Phe164, Ser167, Val268, and Leu272 in PLR_Tp1, and Leu164, Gly167, Gly267, and Phe271 in PLR_Tp2. To investigate the question of differential enantiospecificity, site-directed mutagenesis of some of these residues (in both enzymes) was performed, with mutants obtained indeed displaying altered overall enantiospecificities.[270] Interestingly, the G267V PLR_Tp2 mutant protein was almost inactive (<5% of wild-type PLR_Tp2 activity) with regard to its activity toward (+)- and (−)-pinoresinols (**13a** and **13b**), whereas the V268G PLR_Tp1 mutant protein had its enantiospecificity almost completely reversed. However, this was achieved mainly by abolishing the ability to process (−)-pinoresinol (**13b**). Thus, the overall underlying reasons for distinct enantiospecificities are beginning to be understood.

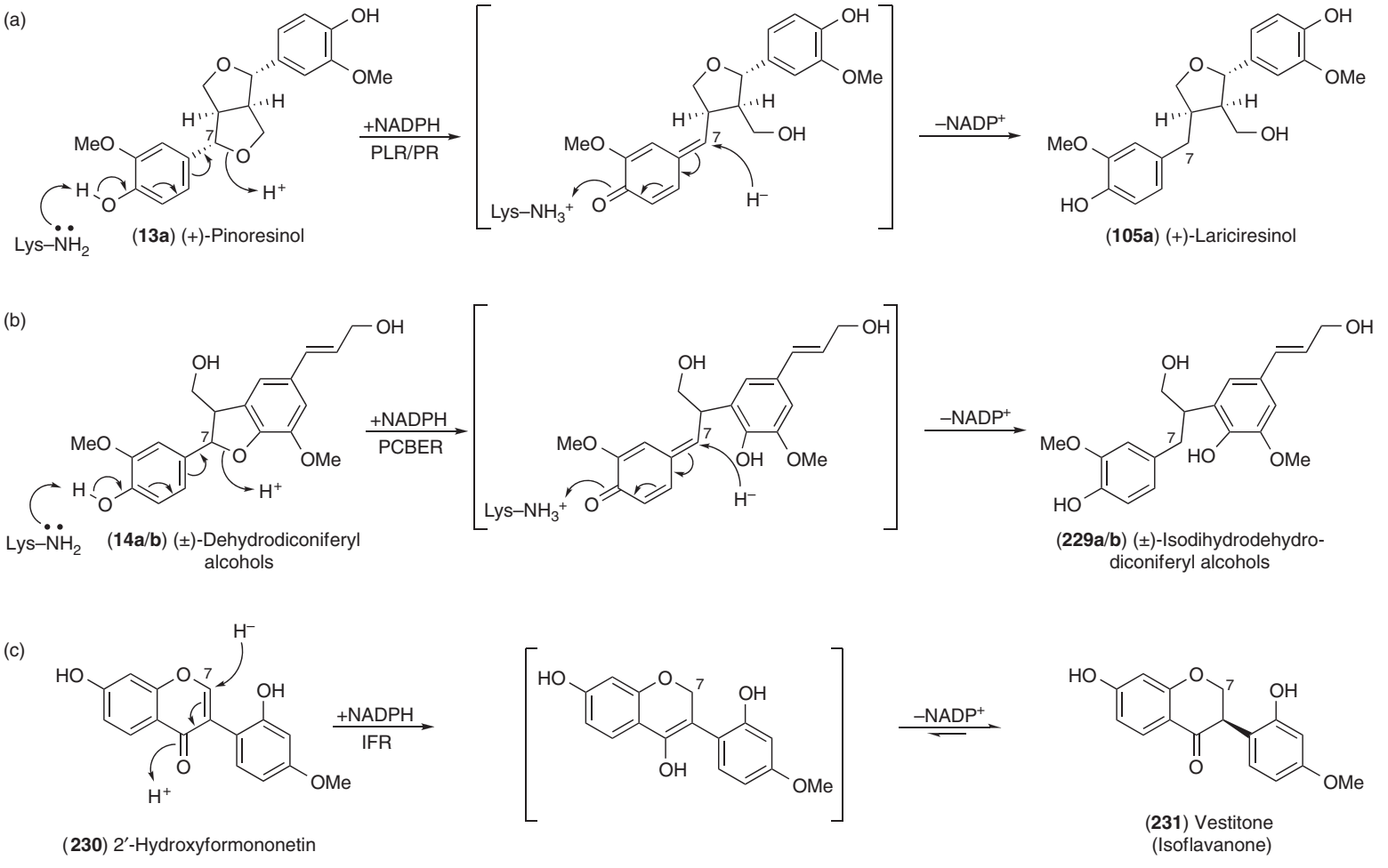


Figure 31 (Continued)

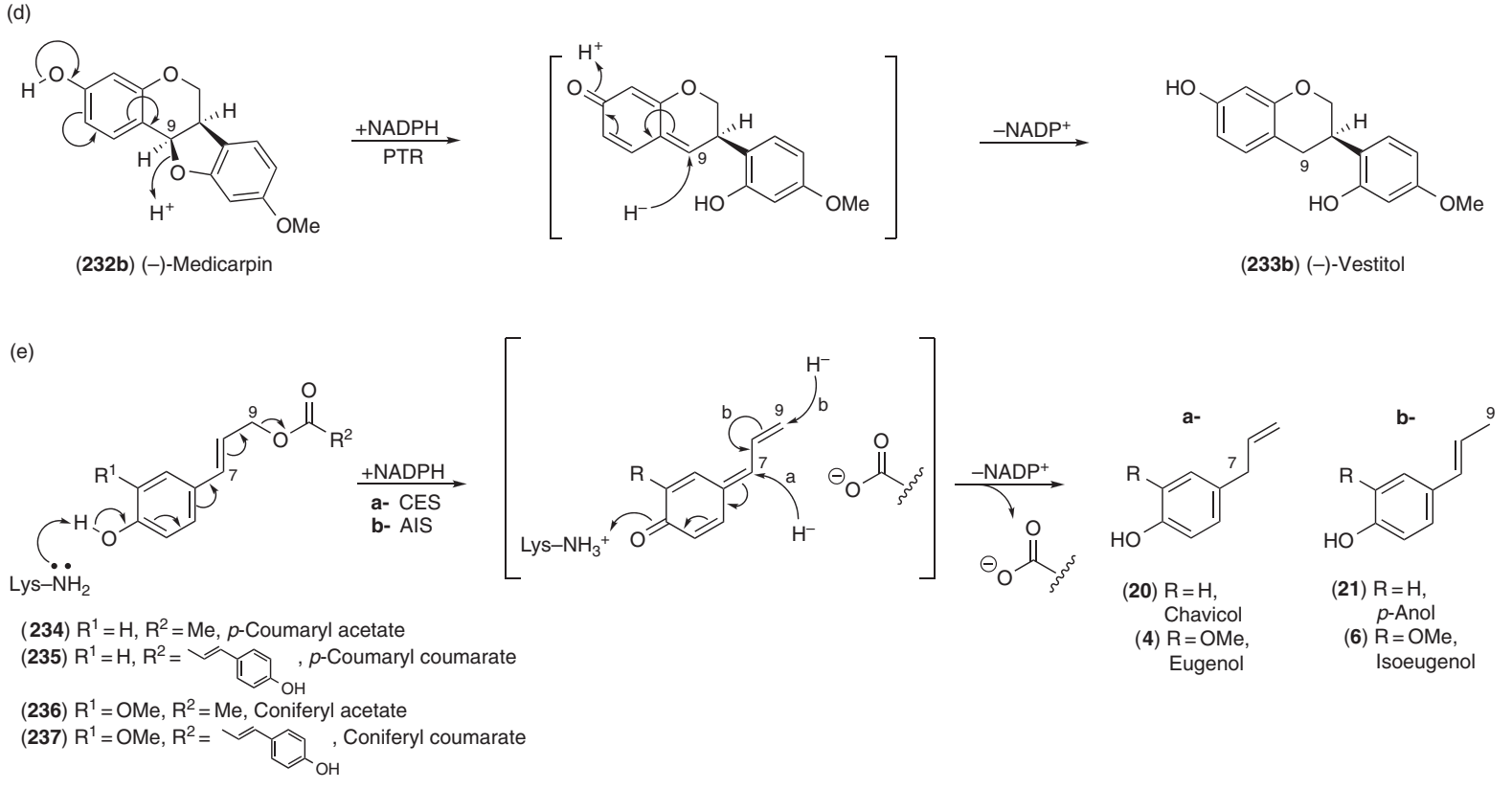


Figure 31 Proposed reaction mechanisms for (a) pinoresinol–lariciresinol reductases/pinoresinol reductases, PLR/PR, (b) phenylcoumaran benzylic ether reductase, PCBER, (c) isoflavone reductase, IFR, (d) pterocarpan reductase, PTR, and (e) chavicol eugenol synthase/anol isoeugenol synthase, CES/AIS.

1.23.5.3 Secoisolariciresinol Dehydrogenase

In many plant species, secoisolariciresinol dehydrogenase (SDH) catalyzes the conversion of secoisolariciresinol (**110**) to matairesinol (**10**). The latter is, in turn, a central intermediate to numerous 8–8′-linked lignans, such as the aforementioned (−)-podophyllotoxin (**1b**) and (−)-plicatic acid (**112b**).

1.23.5.3.1 Discovery of SDH and encoding gene

Using *F. intermedia* cell-free extracts, radiotracer/stable isotope studies initially established that (−)-secoisolariciresinol (**110b**) was enantiospecifically converted into (−)-matairesinol (**10b**) in the presence of NAD^+;[220,222] the (+)-enantiomer (**110a**) was not utilized. This was unambiguously established by the incubation of both (±)-[9,9′-^{3}H]- and (±)-[Ar-^{2}H]-secoisolariciresinols (**110a/b**) to afford (−)-[9-^{3}H]- and (−)-[Ar-^{2}H]-matairesinol (**10b**), respectively. The resulting secoisolariciresinol dehydrogenase (SDH) was also purified (>6000-fold) from *F. intermedia* stems, the encoding gene (*SDH_Fi321*) cloned, and fully functional recombinant protein expressed in *E. coli*;[295] the corresponding 831-bp cDNA encoded a 277-amino acid protein. Using recombinant SDH_Fi321, catalysis was observed to occur via intermediacy of (−)-lactol (**240b**) (**Figure 32**), indicative of the bifunctional nature of the SDH-catalyzed transformation. Interestingly, however, (−)-lactol (**240b**) intermediacy was not detected though when assaying the native *F. intermedia* SDH.

Additionally, an SDH gene (*SDH_Pp7*) was isolated from the (−)-podophyllotoxin (**1b**)-containing species, *Podophyllum peltatum*. Following the heterologous recombinant protein expression of the presumed native SDH in *E. coli*, and purification to apparent homogeneity, the kinetic parameters (K_m and V_{max} values) were determined using (−)-lactol (**240b**) as the substrate. The apparent K_m, V_{max}, and k_{cat}/K_m values of 160.2, 118 pkat μg^{-1} protein, and 23 600 mol^{-1} l s^{-1} obtained were also consistent with kinetic data for the general phenylpropanoid pathway enzymes to the monolignols.[7]

NMR spectroscopic analyses were also employed to establish the stereospecificity of hydride transfer from (−)-secoisolariciresinol (**110b**) to [4-^{2}H]-NAD^+, and whether this occurred at C-4 of either the *pro-R* position, the *pro-S* position, or both.[296] To investigate this further, enzymatic syntheses of both [4*R*-^{2}H]- and [4*S*-^{2}H]-NADH were carried out, as well as that of [4-^{2}H]-NAD^+.[296] Analysis of [4*R*-^{2}H]- and [4*S*-^{2}H]-NADH by ^{1}H (**Figures 33(b)** and **33(c)**) and heteronuclear multiple-quantum correlation (HMQC)-NMR spectroscopy, and comparison with natural-abundance NADH (**Figure 33(a)**), enabled the 4*S*-proton to be assigned to δ 2.65 ppm and the 4*R*-proton to δ 2.77 ppm.

Further NMR studies of the [4-^{2}H]-NADH generated during dehydrogenation of (−)-secoisolariciresinol (**110b**) by SDH established the stereochemistry of hydride transfer. Following incubation of (±)-secoisolariciresinols (**110a/b**) with partially purified SDH in the presence of [4-^{2}H]-NAD^+, the enzymatically generated (deuterated) NADH was purified and analyzed by ^{1}H- (**Figure 33(d)**) and HMQC- (not shown) NMR spectroscopy. The broad doublet signal observed at δ 2.63 ppm ($\mathcal{J}_{4S-5}$ 3.0 Hz, **Figure 33(d)**), in addition to lack of geminal coupling, established the deuterium to be in the 4*R* position. Thus, the hydride abstracted from (−)-secoisolariciresinol (**110b**) and/or (−)-lactol (**240b**) was added to the *pro-S* position of [4-^{2}H]-NAD^+, affording [4*R*-^{2}H]-NADH, thereby establishing SDH as a type B dehydrogenase.

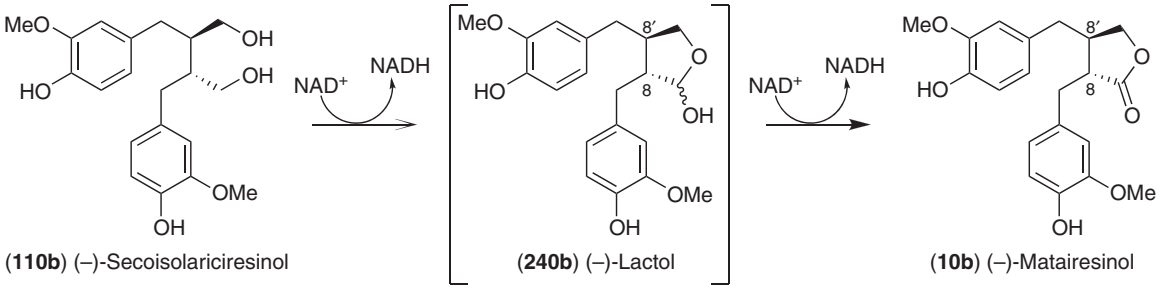


Figure 32 Reaction catalyzed by secoisolariciresinol dehydrogenase (SDH).

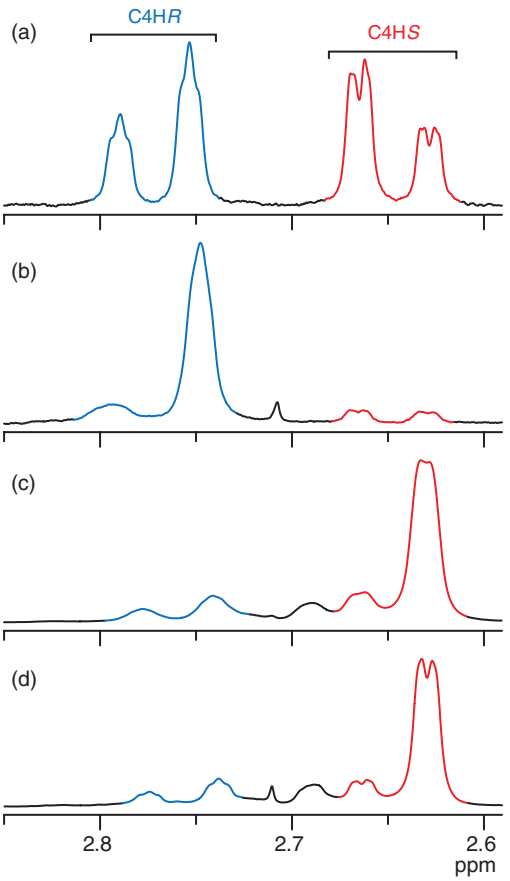


Figure 33 Partial ^{1}H NMR spectra of NADH showing spectroscopic regions for 4*S* and 4*R* protons at C4 of the dihydropyridine ring. (a) Natural-abundance NADH. (b) [4*S*-^{2}H]-NADH. (c) [4*R*-^{2}H]-NADH. (d) Enzymatically synthesized [4*R*-^{2}H]-NADH with SDH_Pp7 following incubation with (±)-secoisolariciresinols (**110a/b**) and [4-^{2}H]-NAD^{+}. Reproduced from S. G. A. Moinuddin; B. Youn; D. L. Bedgar; M. A. Costa; G. L. Helms; C. Kang; L. B. Davin; N. G. Lewis, *Org. Biomol. Chem.* **2006**, *4*, 808–816. Copyright 2006, with permission from the Royal Society of Chemistry.

1.23.5.3.2 Structural biology studies

SDH belongs to the short-chain dehydrogenase (SDR) superfamily. *P. peltatum* SDH (SDH_Pp7) crystal structures were obtained in apo-form, as well as their binary and ternary complexes using NAD^{+} and NAD^{+}/(−)-matairesinol (**10b**), at 1.6, 2.8, and 2.0 Å resolution, respectively.[297] In solution, SDH exists as a homotetramer (as determined by size-exclusion chromatography and multiangle laser light scattering), with each subunit having a single α,β domain structure. In its active site, the SDR-conserved Ser153–Tyr167–Lys171 (the so-called catalytic triad) can apparently function, in combination with the ribose 2′-OH of the cofactor, as a putative proton-relay system (**Figures 34(a)–34(c)**). These residues are considered responsible for cofactor binding, with Tyr167 also placed in close proximity to the alcohol functionality of (−)-secoisolariciresinol (**110b**) upon binding. Participation of the proposed catalytic triad was further supported by site-directed mutagenesis of residues Tyr167 and Lys171 to Ala, which abolished catalytic activity, and Ser153 to Ala, which displayed modest activity.[296]

The entropically favored NAD^{+} binding is provisionally considered to be further stabilized by hydrogen bonding between the cofactor ribose 2′- and 3′-OH groups and Tyr167 and Lys171, respectively (**Figure 34(b)**). This, in turn, would favor Tyr167 deprotonation to afford the corresponding phenolate (which is further stabilized by the neighboring Ser153 hydroxyl group). Tyr167 is thus considered to function

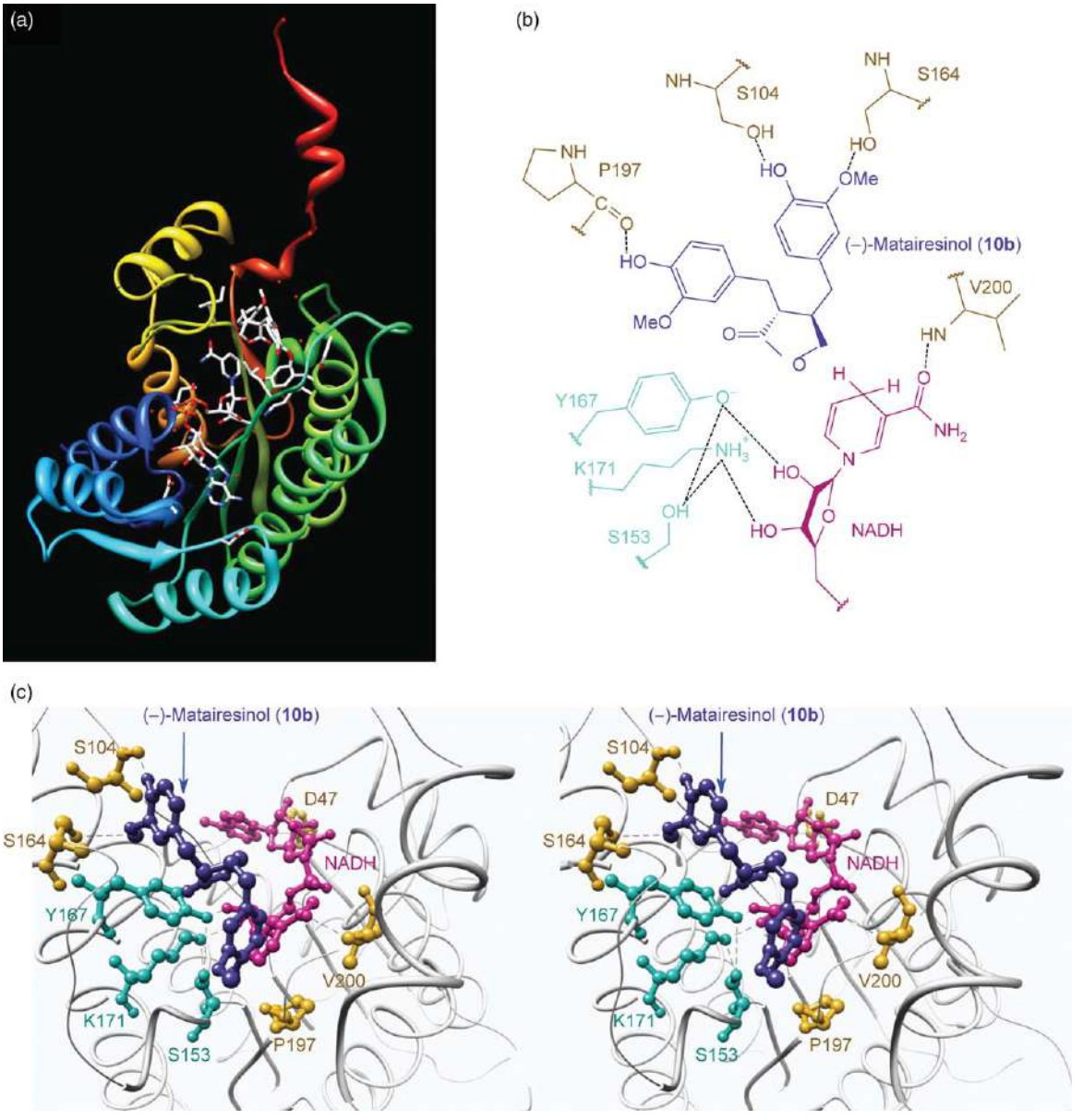


Figure 34 (a) Ternary complex of SDH_Pp7 with NAD^+ and (−)-matairesinol (**10b**). (b) The Tyr167 phenolic and Lys171-protonated amino groups are hydrogen-bonded to 2′- and 3′-OH of NADH, respectively. (c) Structures of the substrate-binding pocket of SDH_Pp7 in ternary complex as a stereoview. The catalyt are shown in blue triad residues, Ser153, Tyr167, and Lys171.

as a general catalytic base, abstracting (or perhaps more accurately influencing the local electronic environment of) the alcohol proton of (−)-secoisolariciresinol (**110b**), thereby facilitating intramolecular cyclization to form (−)-lactol (**240b**), via concerted transfer of the C9-hydride to the *pro-S* position of NAD^+. The NADH thus formed is released from the active site, with a second round of catalysis involving a similar process with (−)-lactol (**240b**), resulting in formation of the lactone product, (−)-matairesinol (**10b**), with concomitant hydride transfer to NAD^+ as before.

1.23.5.4 Creosote Bush Lignan Metabolism: Enantiospecific Polyphenol Oxidase

As summarized earlier, the creosote bush (*L. tridentata*) accumulates only 8–8′-linked lignans, such as NDGA (**143**). The latter is presumed to form via regiospecific 8–8′ coupling of the monomeric precursors, affording larreatricin (**200**) and/or other coupled forms (**201–203**) as possible biosynthetic pathway intermediates. In this context, (+)-larreatricin (**200a**) (but not its (−)-antipode (**200b**)) was enantiospecifically hydroxylated *in vitro* by an *L. tridentata* polyphenol oxidase (PPO, **Figure 35**),[298] this being in agreement with the observed 92% e.e. of (−)-larreatricin (**200b**) *in planta*.[247] Additional reductive steps can thus be envisaged to afford NDGA (**143**).

The native protein catalyzing C3′/C3 hydroxylation was purified (~1700-fold) to apparent homogeneity to afford, via SDS–PAGE analysis, an ~43-kDa protein band. Using the purified PPO from *L. tridentata* twigs, hydroxylation at C3′ was favored, affording (+)-3′-hydroxylarreatricin (**241a**) in an ~7:1 ratio relative to the C3-hydroxylated (+)-3-hydroxylarreatricin (**242a**).[298] Peptide fragments from trypsin digestion of the isolated protein were analyzed by microcapillary reversed-phase-HPLC nanoelectrospray tandem MS (μLC-MS/MS), and gave sequences homologous to conserved PPO motifs. The corresponding gene was obtained from a cDNA library, which encoded an ~66-kDa protein. The higher predicted molecular weight (66 kDa vs ~43 kDa) was indicative of posttranslational processing, a feature common to PPOs, this being further confirmed by comparison of deduced and observed amino acid sequences. Of particular note was the presence of five histidine residues, His188, 197, 319, 323, and 353, which are considered involved in the binding of the copper atoms. Larreatricin-3′-hydroxylase is the first characterized example of an enantiospecific PPO.

1.23.5.5 Additional (Preliminary) Studies Toward Justicidin B, Hinokinin, and Podophyllotoxin/6-Methoxypodophyllotoxin Biosynthesis

1.23.5.5.1 Justicidin B

The achiral justicidin B (**55**) described earlier was first reported in *Justicia procumbens*,[299] and is also found in a number of *Linum* species.[290,300–306] To begin to establish whether *L. perenne* PLR (PLR_Lp1, see Section 1.23.5.2.3) was involved in justicidin B (**55**) biosynthesis, shoot cultures were transformed with *Agrobacterium rhizogenes* containing a vector with an RNAi construct for the silencing of the *PLR_Lp1* gene.[287] From the hairy roots later obtained (six independent lines and control lines with and without empty constructs), justicidin B (**55**) levels were reduced to ~24% of control levels, this presumably resulting from the *PLR_Lp1* mRNA levels being partly suppressed. These data are thus provisionally consistent with our earlier projections[1] that pinoresinol (**13**) serves as a common obligatory intermediate to arylnaphthalenes and so on, via the action of PLR and other downstream enzymes.

1.23.5.5.2 Hinokinin

(−)-Hinokinin (**223b**), which can accumulate in *L. corymbulosum* suspension cultures, differs from (−)-matairesinol (**10b**) via two methylenedioxy bridge additions. Bayindir *et al.*[286] proposed two possible pathways for its

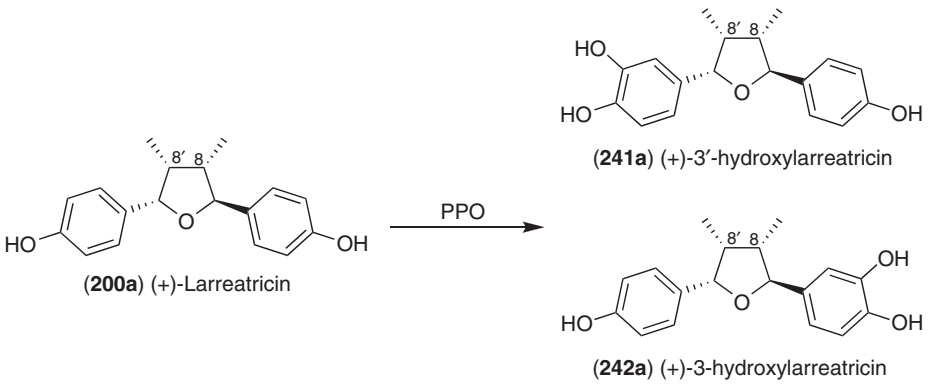


Figure 35 Reaction catalyzed by enantiospecific PPO, larreatricin-3′-hydroxylase, from the creosote bush.

formation: One from secoisolariciresinol (**110**) with further modification, and a second (more unlikely one) from sesamin (**11**) involving a reductase catalyzing an analogous conversion to that of PLR. *L. corymbulosum* shoots were transformed with *A. rhizogenes* harboring the construct (*plr-Lc1*-ihpRNAi) for silencing of *PLR_Lc1* (see Section 1.23.5.2.3) to generate hairy root cultures as above. The resulting transformed cultures were unable to produce (−)-hinokinin (**223b**). No metabolic profiling was, however, reported to indicate whether any upstream metabolites accumulated, such as (+)-pinoresinol (**13a**). Nevertheless, these data were provisionally consistent with the common intermediacy of PLR and pinoresinol (**13**) for the biosynthesis of dibenzylbutyrolactones, such as **223b**, as previously proposed.[1]

1.23.5.5.3 Podophyllotoxin/6-Methoxypodophyllotoxin

As indicated in the previous sections, (−)-podophyllotoxin (**1b**) is found in *Podophyllum* species, and with its derivative, (−)-6-methoxypodophyllotoxin (**194b**), is also present in some *Linum* species. Using *Linum flavum* roots that accumulate **194b**, the pathway from (+)-pinoresinol (**13a**) to (−)-matairesinol (**10b**) was shown to be operative.[226] It was also established, using mass spectral analyses, that following the uptake of matairesinol (**10**) by *L. flavum* root tissue, and its further metabolism for 6 h, conversion to 7-hydroxymatairesinol (**114**) occurred.[226] Moreover, when [7-^{3}H]-7-hydroxymatairesinol (**114**) was also administered to *L. flavum* roots, and allowed to be metabolized for 6 h, the radiolabel was incorporated into (−)-6-methoxypodophyllotoxin (**194b**), suggesting **114** as a pathway intermediate.

In an apparent contrast to the above results,[226] it has been suggested that (−)-deoxypodophyllotoxin (**243b**) might serve instead as a common precursor to both (−)-podophyllotoxin (**1b**) and (−)-6-methoxypodophyllotoxin (**194b**), respectively.[307] A presumed deoxypodophyllotoxin 6-hydroxylase was preliminarily characterized from a crude microsomal preparation obtained from *L. flavum* cell suspension cultures: It was able to catalyze the formation of (−)-β-peltatin A (**244b**) from (−)-deoxypodophyllotoxin (**243b**),[308] with apparent K_m values of 20 and 36 µmol l^{-1} for **243b** and NADPH, respectively (as determined from crude microsomal preparations). In addition, another crude microsomal preparation from *L. album* cell suspension culture was able to convert (−)-deoxypodophyllotoxin (**243b**) into (−)-β-peltatin A (**244b**) in the presence of

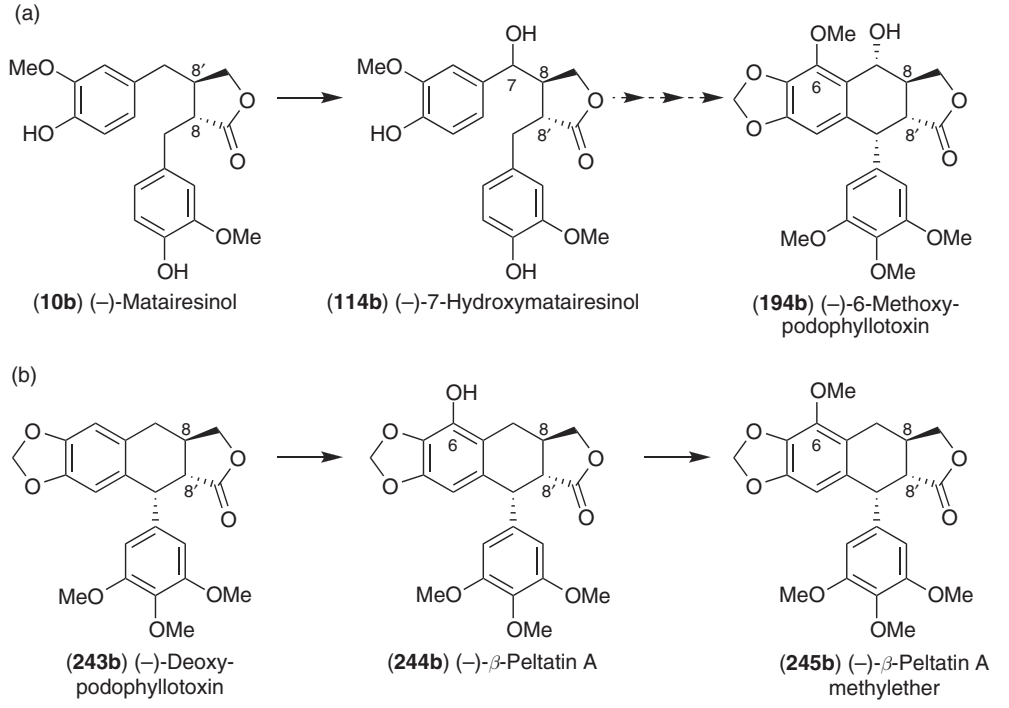


Figure 36 Proposed pathways to (−)-6-methoxypodophyllotoxin (**194b**) from (−)-matairesinol (**10b**) via (a) (−)-7-hydroxymatairesinol (**114b**)[226] or (b) (−)-deoxypodophyllotoxin (**243b**).[288,308]

NADPH (with apparent K_m values of 3 and 41 μmol l^{-1} for **243b** and NADPH, respectively).[288] No other kinetic data were, however, obtained.

In any event, the complete pathway to **1b**/**194b** still remains to be elucidated, that is, in terms of the various hydroxylations, *O*-methylation, and methylenedioxy bridge-forming steps, as well as for aryltetrahydronaphthalene ring formation.

Additionally, a presumed *β*-peltatin 6-*O*-methyltransferase was preliminarily detected in cell suspension cultures of *Linum nodiflorum*, which converted *β*-peltatin A (**244b**) into *β*-peltatin A methylether (**245b**). Apparent K_m values were estimated (using proteins recovered after $(NH_4)_2SO_4$ precipitation) as 40 and 15 μmol l^{-1} for **244b** and *S*-adenosyl methionine, respectively. Neither matairesinol (**10**), nor pinoresinol (**13**), nor podophyllotoxin (**1**), apparently served as substrates.[307] To date, however, none of these enzymes has been purified and/or the encoding genes cloned.

1.23.6 Other Phenylpropanoid Coupling Modes: 8–2′, 8–3′ (8–5′), and 8–*O*–4′-Linked Lignans

Although many additional coupling modes in lignans are evident (see Section 1.23.3.2) from the different skeletal types and optical activities reported, most studies have thus far focused mainly on the formation of the 8–8′-linked lignans. Nevertheless, the generation of other optically active skeletal types can be envisaged to occur through DP-assisted coupling of two radical intermediates, although whether they are derived from monolignols, allylphenols, or hydroxycinnamic acids, and/or combinations thereof to generate different linkage modes is currently unknown.

Interestingly, DP participation has been reported as involved in the formation of the toxic chiral (atropisomeric) (+)-gossypol (**247a**),[309] a terpenoid derived from the coupling of the achiral precursor hemigossypol (**246**) in moco cotton (*Gossypium hirsutum* L. var. *marie galante*) flower petals. The optical activity in this case results from restricted rotation around the C–C biphenyl linkage (**Figure 37**). When laccase/O_2 were added to partially purified and presumed DP-enriched fractions from flower petals, the e.e. of product formation increased to ~65% (from being 59% without laccase/O_2), with the overall amounts of (+)-gossypol (**247a**) also increased. This contrasts with one-electron oxidation of hemigossypol (**246**) *in vitro*, which produces only the racemic products (±)-**247a/b**. These data, therefore, suggest involvement of DPs beyond the phenylpropanoid pathway, albeit still utilizing phenolic substrates.

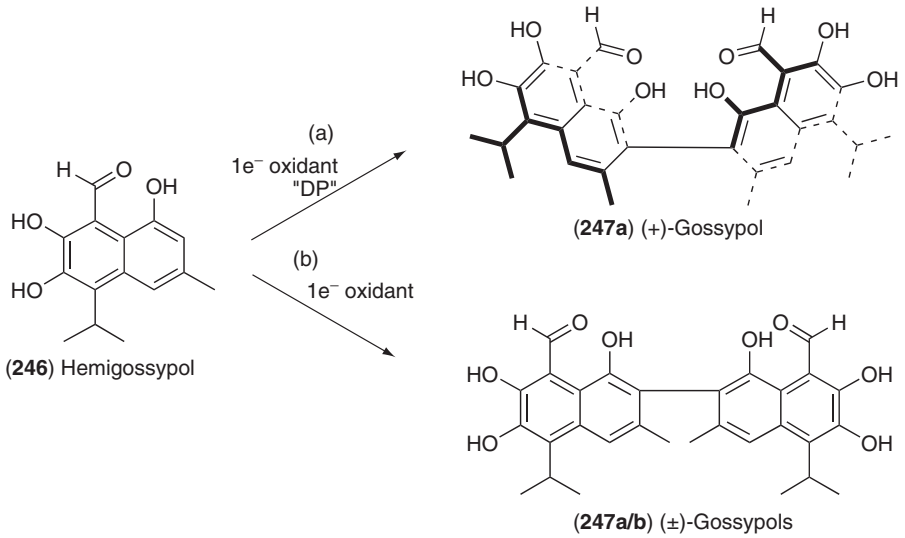


Figure 37 Stereoselective and non-stereoselective formation of gossypol (**247**) from hemigossypol (**246**). (a) Stereoselective: presence of an $1e^-$ oxidant and a putative dirigent protein (DP) in cotton flower petals to afford (+)-gossypol (**247a**).[309] (b) Non-stereoselective: presence of $1e^-$ oxidant affording (±)-gossypols (**247a/b**).

Some of the other coupling modes described earlier in the chemotaxonomical analysis of lignans have been preliminarily investigated from a biochemical perspective, as summarized below.

1.23.6.1 8–2′ Coupling

The optically active 8–2′-linked lignans, (−)-(*Z*/*E*)-blechnic acids (**248b**/**99b**), and its derivative (−)-*E*-brainic acid (**102b**), in the fern *Blechnum spicant* were demonstrated to be derived from labeled [U-^{14}C]-, [1-^{13}C]-, [2-^{13}C]-, and [3-^{13}C]-Phe (**2**), as well as [9-^{14}C]-cinnamic (**249**), [8-^{14}C]- and [8-^{13}C]-*p*-coumaric (**250**), and [8-^{13}C]-caffeic (**56**) acids *in vivo*.[138] Based on the incorporation data obtained, (−)-*Z*-blechnic acid (**248b**) was apparently the initial coupling product (**Figure 38**), with further metabolism affording both (−)-*E*-blechnic acid (**99b**) and (−)-*E*-brainic acid (**102b**). The proteins/enzymes involved in these transformations, as well as the corresponding encoding genes, have not yet been reported. One possible is that they involve either DP- or DP-like-mediated conversions.

1.23.6.2 8–3′ (8–5′) Coupling

In *Piper regnellii*, the 8–3′-linked lignans (+)-conocarpan (**16a**) and (−)-*epi*-conocarpan (**188b**), as well as fully conjugated derivatives such as **191**, accumulate in the leaf and root tissues.[310] Of these, (+)-conocarpan (**16a**) formation was envisaged in preliminary studies to result through 8–3′ regio- and stereoselective coupling of

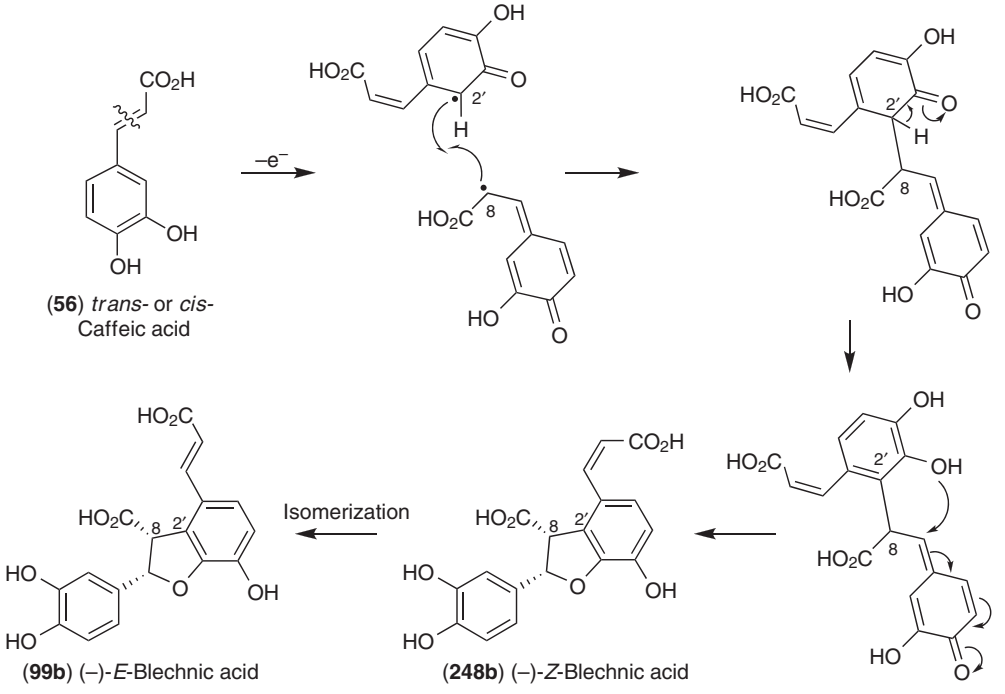
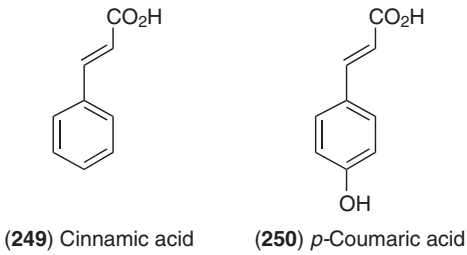


Figure 38 Putative biosynthetic pathway to 8–2′-linked (−)-*E*-blechnic acid (**99b**) in *Blechnum spicant*.[138]



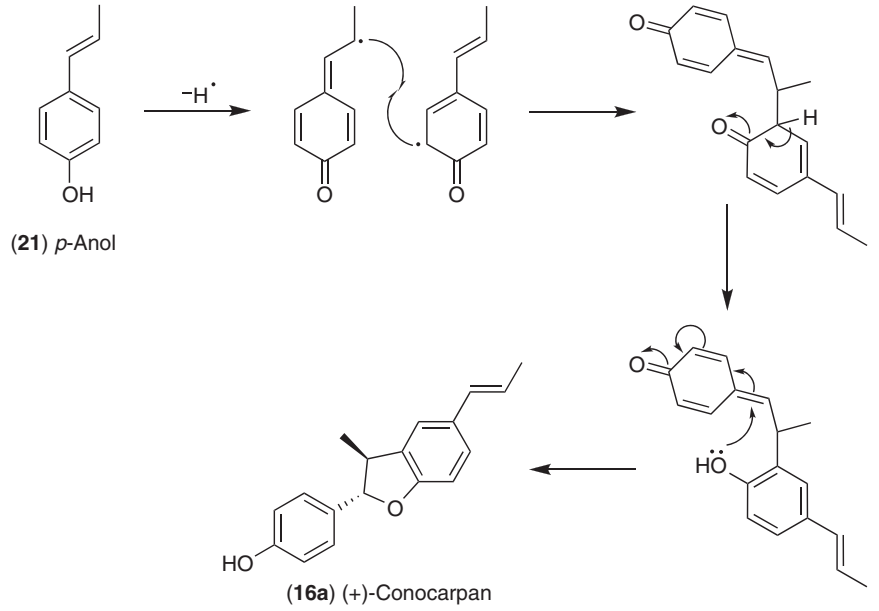


Figure 39 Putative biosynthetic pathway to 8–3′-linked (+)-conocarpan (**16a**) in *Piper regnellii*.

two *p*-anol (**21**) moieties (**Figure 39**). While of interest, these preliminary studies have not yet established the nature of the proteins and enzymes involved. Again, how this either compares or differs with the above DP-mediated processes will be important to determine.

1.23.6.3 8–*O*–4′ Coupling

Eucommia ulmoides reportedly produces diastereoselectively coupled 8–*O*–4′-linked lignans, with the stereochemistries at positions 7 and 8 being of particular interest, in terms of how they are controlled. For example, coniferyl alcohol (**91**) was preferentially converted into the optically active 8–*O*–4′-linked (+)-*erythro*- and (−)-*threo*-guaiacylglycerol coniferyl alcohol ethers (**115**, in an ∼3:2 ratio, **Figure 40**) using *E. ulmoides* cell-free extracts containing a presumed insoluble cell wall residue preparation in the presence of H_2O_2.[311] Similarly, sinapyl alcohol (**94**) was apparently also metabolized to preferentially afford both (+)-*erythro*- and (−)-*threo*-syringylglycerol sinapyl alcohol ethers (**252**), in an ∼10:1 ratio.[312] Furthermore, young excised *E. ulmoides* shoots offered both coniferyl (**91**) and sinapyl (**94**) alcohols formed guaiacylglycerol sinapyl alcohol ether (**251**) in a ∼2.4:1 *erythro*/*threo* ratio.[313] The biochemical factor(s) controlling diastereoselective coupling, including involvement of possible DP protein(s), await further clarification.

1.23.7 Allylic (Phenylpropenal) Double Bond Reductases and Phenylcoumaran Benzylic Ether Reductases

1.23.7.1 Allylic (Phenylpropenal) Double Bond Reductases: Biosynthesis of Dihydrolignans and Dihydromonolignols

As far as current chemotaxonomical information permits, the evolutionarily earliest known example of dihydrolignans (i.e., lignans with saturated side-chains) is in the ferns (e.g., **103a** and **104b** in *P. vittata*[139]), with this structural motif sporadically reoccurring in the gymnosperms and angiosperms. Metabolites with reduced side-chains now encompass not only 8–5′-linked lignans, but those that are 8–*O*–4′-linked (e.g., **116** and **117**) as well as various monomeric phenylpropanoids. That is, *P. glauca* shoots and galls accumulate, among other phenolics, dihydrodehydrodiconiferyl alcohol (**103**) and its demethylated derivative **116**, dihydroconiferyl alcohol (**119**), dihydro-*p*-coumaric acid (**253**), and dihydroferulic acid (**255**), as well as the

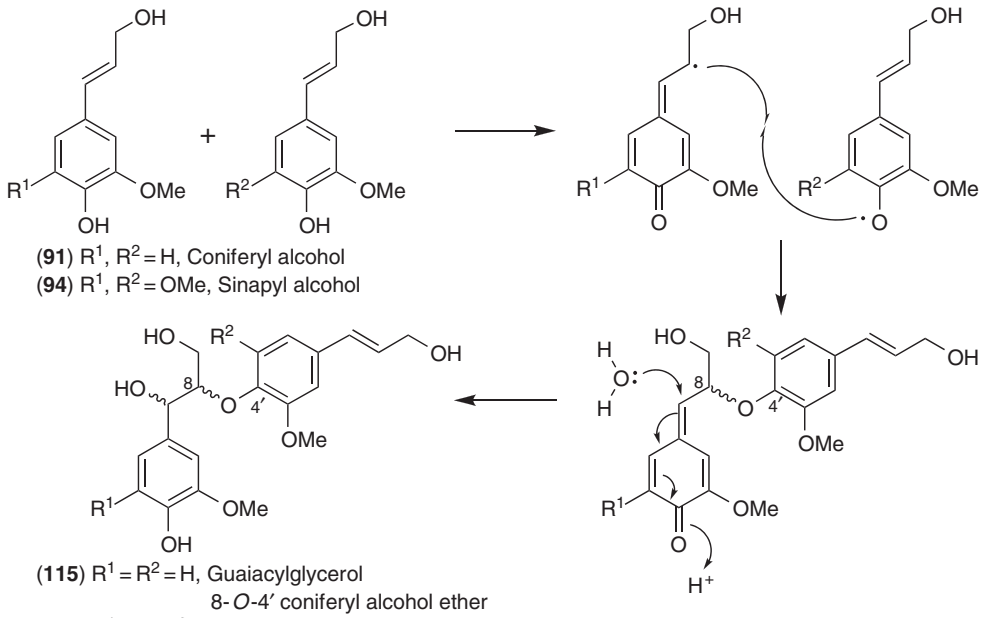
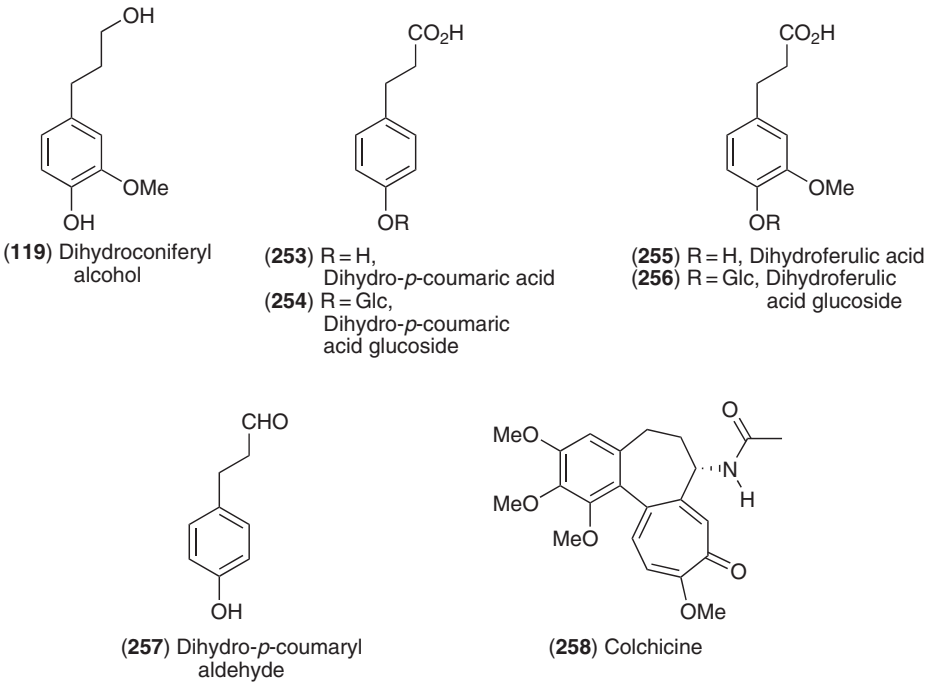


Figure 40 Putative diastereoselective 8–*O*–4′-coupling in *Eucommia ulmoides* affording lignans **115**, **251**, and **252**.

corresponding glucosides **254**/**256**.[161] While the main function of these natural products appear to be in plant defense (see Section 1.23.10.2), dihydro-*p*-coumaryl aldehyde (**257**) was also proposed as a precursor of the phenethylisoquinoline alkaloid, colchicine (**258**), in *Colchicum byzantinum* and *Colchicum autumnale*.[314]



*1.23.7.1.1 Discovery of allylic (phenylpropenal) double bond reductases and gene cloning: Loblolly pine (*Pinus taeda*)*

Biosynthetic studies to dihydrolignans and dihydromonolignols began with observations made using *P. taeda* cell cultures elicited with an 8% sucrose solution. Under these conditions, which result in phenylpropanoid pathway induction, dihydrodehydrodiconiferyl alcohol (**103**) was isolated from the complex mixture of soluble phenolics accumulating in the cell bathing medium.[147] Using a partially purified protein preparation from this source, the NADPH-dependent conversion of dehydrodiconiferyl alcohol (**14**) into dihydrodehydrodiconiferyl alcohol (**103**) was demonstrated.[315,316] Ultimately, three predominant coeluting protein bands were separated using SDS–PAGE, with trypsin digestion and peptide sequencing leading to the identification of cinnamyl alcohol dehydrogenase (CAD), the last enzymatic step involved in monolignol biosynthesis, a fructose-bisphosphate aldolase, and an ~39-kDa protein of undetermined function.

The full-length cDNA clone for the latter was obtained, which encoded an ~38.7-kDa NADPH-dependent protein that was subsequently expressed in *E. coli*, albeit mainly in an insoluble form. The reductase was, however, unable to catalyze the NADPH-dependent side-chain reduction of either dehydrodiconiferyl (**14**) or coniferyl (**91**) alcohols directly. Based on the earlier protein assays, which contained this protein and CAD, it was both deduced and then established using recombinant protein that the actual substrates for the reductive transformations in the *P. taeda* cell cultures were dehydrodiconiferyl aldehyde (**259**) and coniferyl aldehyde (**262**), respectively, to afford **260** and **263** (**Figure 41**). The gene encoding this protein was thus named *PtPPDBR* (*P. taeda* phenylpropenal double bond reductase).

This finding thus apparently differed from another preliminary report indicating that coniferyl aldehyde (**262**) could be converted *in vitro* into coniferyl (**91**) and dihydroconiferyl (**119**) alcohols, in the presence of NADPH, using a microsomal preparation from developing xylem of *Pinus strobus*.[317] That study instead suggested, based on the data obtained, the involvement of a cytochrome P-450.

Based on the data above, the functionally operative enzyme isolated from the *P. taeda* cell culture was an NADPH-dependent phenylpropenal double bond reductase (PtPPDBR).[315,316] Comprehensive kinetic parameters could not be measured, however, due to poor expression levels of recombinant PtPPDBR. Such data still need to be obtained, though, in order to assess its relative efficacy in the overall phenylpropanoid pathway.

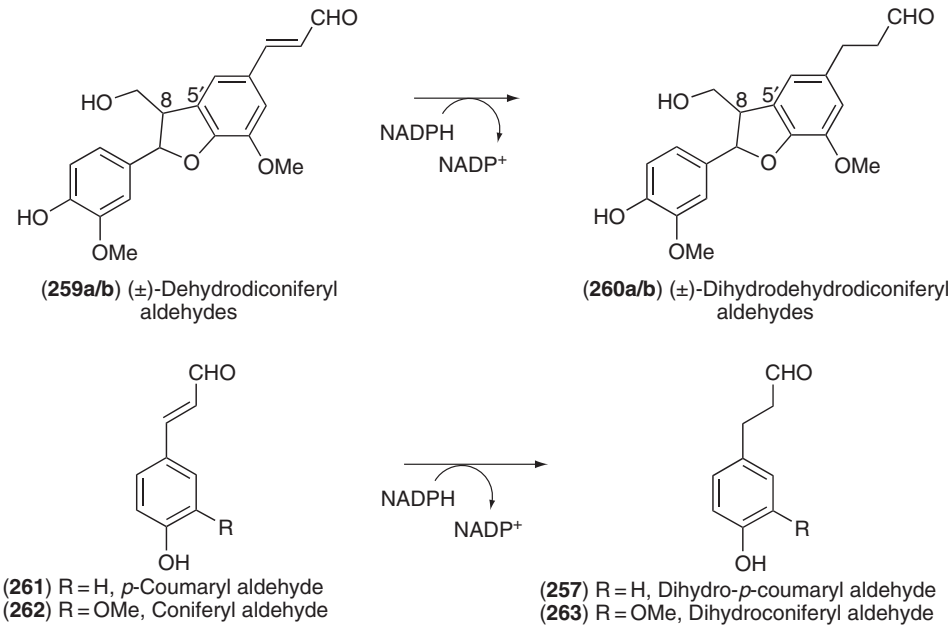


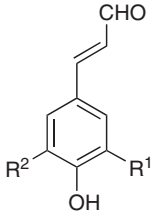
Figure 41 Reactions catalyzed *in vitro* by allylic double bond reductases from loblolly pine (*Pinus taeda*) and *Arabidopsis thaliana*.

1.23.7.1.2 mRNA tissue localization of PtPPDBR in loblolly pine

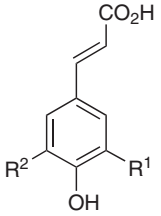
Pinus taeda PPDBR mRNAs were also localized in various stem tissues using *in situ* hybridization.[316] These were detected in the vascular cambium, radial parenchyma cells of the xylem, and axial parenchyma cells of the cortex of young (just below the apical meristem, **Figure 42(a)**) and more mature *P. taeda* stems (**Figures 42(b)** and **42(c)**). Localization to the vascular cambium and radial parenchyma cells thus correlated well with mRNA localization of DP in *F. intermedia* and *T. plicata*, as well as that of PLR in *F. intermedia*. That is, the PtPPDBR mRNA localization co-occurs in cell types known to be involved in lignan biosynthesis.

1.23.7.1.3 Allylic double bond reductase homologs: eleven-membered multigene family in Arabidopsis

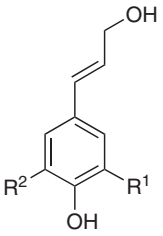
Based on sequence comparisons of PtPPDBR with homologs in the *A. thaliana* genome, a very provisional annotation of an 11-membered alkenal double bond reductase (AtDBR) multigene family was made,[318] based on an ~63/43% amino acid similarity/identity to PtPPDBR. It was, therefore, of interest to clone each of the encoding genes, in order to obtain the corresponding recombinant proteins in *E. coli* and assess their potential phenylpropenal reductase activities. Of the 11 recombinant proteins (obtained as His-tag fusion proteins), AtDBR1, 8, and 9 were catalytically competent to reduce substrates such as *p*-coumaryl (**261**) and coniferyl (**262**) aldehydes, to the corresponding (propyl) single bond derivatives, **257** and **263**, with AtDBR1 and 9 also able to reduce caffeyl aldehyde (**264**) to a lesser extent. Of these, AtDBR1 was further characterized with its kinetic parameters determined for **261** and **262**: The K_m, k_{cat}, and k_{cat}/K_m values were 0.53/0.41 mmol l^{-1}, 2.82/0.43 s^{-1}, and 5360/1060 mol^{-1} l s^{-1}, respectively. None of the other tested phenylpropanoid aldehydes (5-hydroxyconiferyl (**265**) and sinapyl (**266**) aldehydes), carboxylic acids (**56**, **249**, **250**, **267**–**269**), alcohols (**91**, **94**, **199**, **270**, **271**), or dehydrodiconiferyl aldehyde (**259**) served as substrates for side-chain reduction *in vitro*.



(**261**) R^1, R^2 = H, *p*-Coumaryl aldehyde
(**264**) R^1 = OH, R^2 = H, Caffeyl aldehyde
(**262**) R^1 = OMe, R^2 = H, Coniferyl aldehyde
(**265**) R^1 = OMe, R^2 = OH, 5-OH Coniferyl aldehyde
(**266**) R^1, R^2 = OMe, Sinapyl aldehyde



(**250**) R^1, R^2 = H, *p*-Coumaric acid
(**56**) R^1 = OH, R^2 = H, Caffeic acid
(**267**) R^1 = OMe, R^2 = H, Ferulic acid
(**268**) R^1 = OMe, R^2 = OH, 5-OH Ferulic acid
(**269**) R^1, R^2 = OMe, Sinapic acid



(**199**) R^1, R^2 = H, *p*-Coumaryl alcohol
(**270**) R^1 = OH, R^2 = H, Caffeyl alcohol
(**91**) R^1 = OMe, R^2 = H, Coniferyl alcohol
(**271**) R^1 = OMe, R^2 = OH, 5-OH Coniferyl alcohol
(**94**) R^1, R^2 = OMe, Sinapyl alcohol

Following the PtPPDBR work, AtDBR1 was later reported as able to convert 4-hydroxy-(2*E*)-nonenal (4-HNE, **272**) into 4-hydroxynonanal (4-HNA, **273**, **Figure 43**) *in vitro*.[320] However, in our hands, this occured with much lower enzymatic efficacy (K_m of 0.28 mmol l^{-1}, k_{cat} of 0.16 s^{-1}, and k_{cat}/K_m of 620 mol^{-1} l s^{-1}) relative to *p*-coumaryl aldehyde (**261**).[318] 4-HNE (**272**) was considered as a potential substrate, since

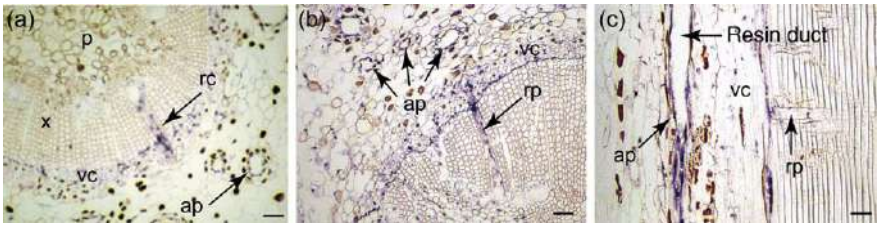


Figure 42 PtPPDBR gene expression localization in *P. taeda* stems at different developmental stages. (a) Transversal cross-section of meristematic region; (b and c) transversal (b) and longitudinal (c) cross sections of a young stem. Abbreviations: ap, axial parenchyma cells; p, pith; rp, radial parenchyma cells; vc, vascular cambium. Bars: 50 μm. Reproduced from H. Kasahara; Y. Jiao; D. L. Bedgar; S.-J. Kim; A. M. Patten; Z.-Q. Xia; L. B. Davin; N. G. Lewis, *Phytochemistry* **2006**, *67*, 1765–1780. Copyright 2006, with permission from Elsevier.

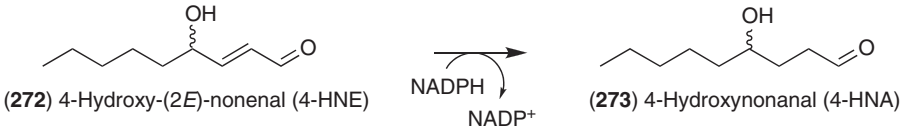


Figure 43 Reaction catalyzed by AtDBR using HNE (**272**) as substrate.

it is a well-characterized lipid peroxidation product that can induce apoptosis; reduction leads to 4-HNA (**273**), which is thought to participate in 4-HNE (**272**) detoxification *in vivo*. It is important to emphasize that, while enzymatically catalyzed conversions have been demonstrated *in vitro*, the true physiological substrate(s) of AtDBR1 in *Arabidopsis* and its role *in vivo* remain uncertain. Indeed, a similar situation currently holds for all members of this multigene family in *Arabidopsis*.

In summary, both PtPPDBR and AtDBR1 were demonstrated *in vitro* to be able to catalyze the addition of an NADPH-derived hydride to the (side chain) double bond of a conjugated propenal (e.g., **261** and **262**), but not of the monolignols (e.g. **199**, **91**). Other related double bond reductases have also been described with similar structural motif requirements. For example, *Artemisia annua* artemisinic aldehyde reductase can reduce the aldehyde-conjugated Δ11(13) double bond of artemisinic aldehyde (**274**) into **275**, but not that of artemisinic alcohol (**276**, **Figure 44(b)**),[321] and (+)-pulegone reductase from *Mentha piperita* (amino acid identities of 42% to PtPPDBR and 63% to AtDBR1) can convert **277** into **278** and **279**[322] (**Figure 44(b)**). The reported enzymes thus have similar biochemical functions, in terms of requiring the substrate to have a conjugated enone system.

1.23.7.1.4 Structural biology studies: Arabidopsis *DBR1*

AtDBR1 is a homodimer in solution, as determined by size-exclusion chromatography, as well as by static and dynamic multiangle light scattering.[318] It is a member of the zinc-independent family of medium-chain dehydrogenases/reductases (MDR), with a nucleotide-binding Gly-rich motif. Its closest structural homolog in the protein database is presently 12-hydroxydehydrogenase/15-oxo-prostaglandin 13-reductase (12-HD/PGR) from guinea pig.

Recombinant AtDBR1 was studied to establish its catalytic mode of action. In this context, crystal structures at 2.5 – 2.8 Å resolution were obtained in apo-form, as well as binary and ternary complexes with $NADP^+$ in the absence and presence of either potential substrates, *p*-coumaryl aldehyde (**261**, **Figure 45**) or 4-HNE (**272**), respectively.

Mechanistically, the highly conserved residue (Tyr260 in AtDBR1) is considered hydrogen-bonded to the aldehyde oxygen of the substrate, and is thus thought to act as a general acid during catalysis. The proposed catalytic mechanisms for AtDBR1[318] (**Figures 46(a)** and **46(b)**), and by extension PtPPDBR, as well as the related HD/PGR are thus very similar.[323] For AtDBR1, an enolate/carbocation intermediate can be envisaged to be generated via double bond migration, followed by the addition of the hydride from NADPH to the carbocation at either C7 from *p*-coumaryl aldehyde (**261**, **Figure 46(a)**) or C3 from 4-

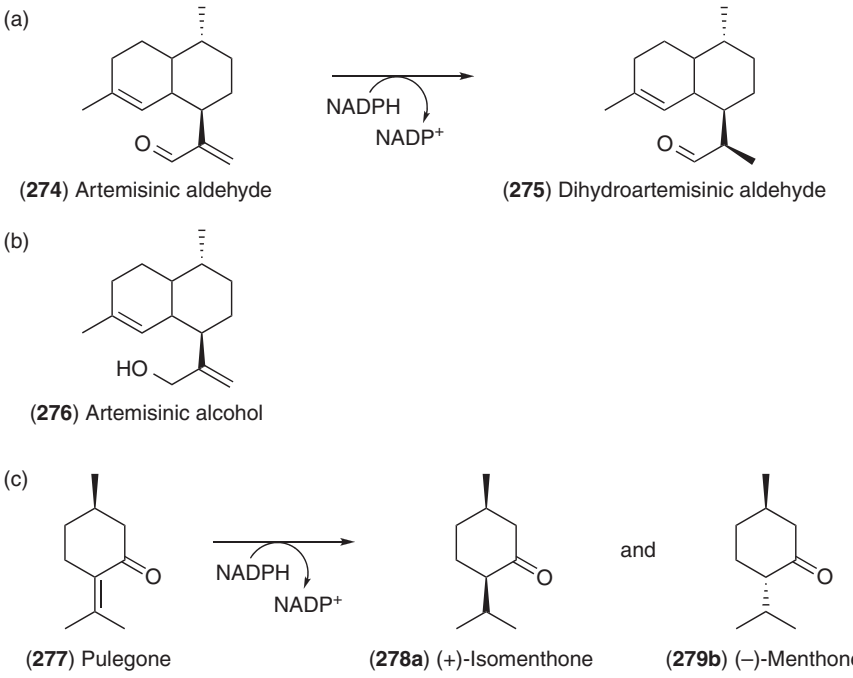


Figure 44 Reactions catalyzed by other NADPH-dependent double bond reductases *in vitro*. (a) Artemisinic aldehyde reductase from *Artemisia annua*. (b) Structure of artemisinic alcohol (**276**). (c) (+)-Pulegone reductase from *Mentha piperita*.

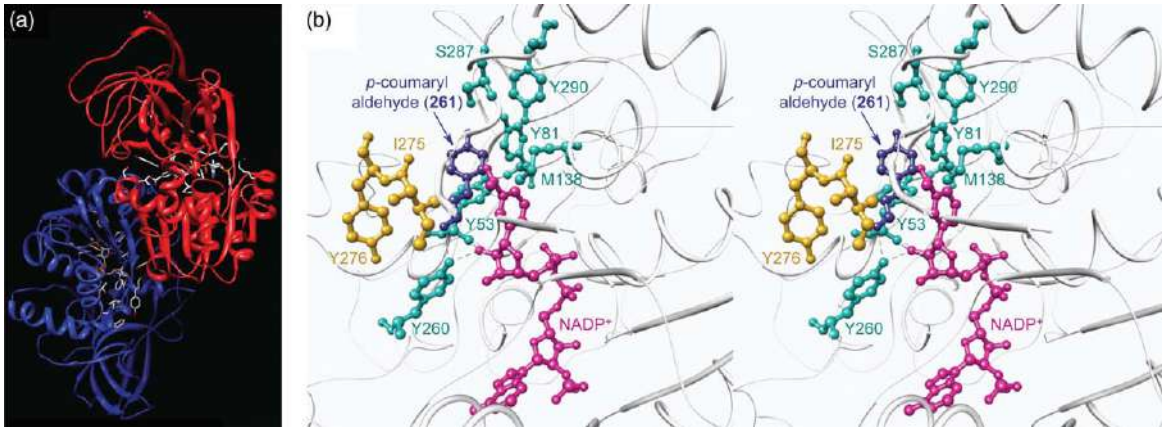


Figure 45 (a) Crystal structure of AtDBR1 ternary complex. (b) Stereoview of the substrate-binding pocket of AtDBR1 (viewed from the bulk solvent) in the ternary complex with $NADP^+$ and *p*-coumaryl aldehyde (**261**). The participating residues in the binding substrate and cofactor are shown with their residue position numbers. The residues are depicted in light blue and yellow to represent their belonging to two different subunits.

HNE (**272**, **Figure 46(b)**). Ultimately, the adjacent carbon (C8 in *p*-coumaryl aldehyde (**261**) or C2 in 4-HNE (**272**)) is protonated by solvent with re-formation of the aldehydic moiety. An alternative concerted mechanism (**Figures 46(c)** and **46(d)**), however, cannot be ruled out. Hydride addition (to either C7 of *p*-coumaryl aldehyde (**261**) or C3 of 4-HNE (**272**)) could lead directly to double bond migration and protonation of the carbonyl oxygen with H^+ from Tyr260, to form an enol as product, which would then undergo tautomerism to yield the aldehydes **257** or **273**.

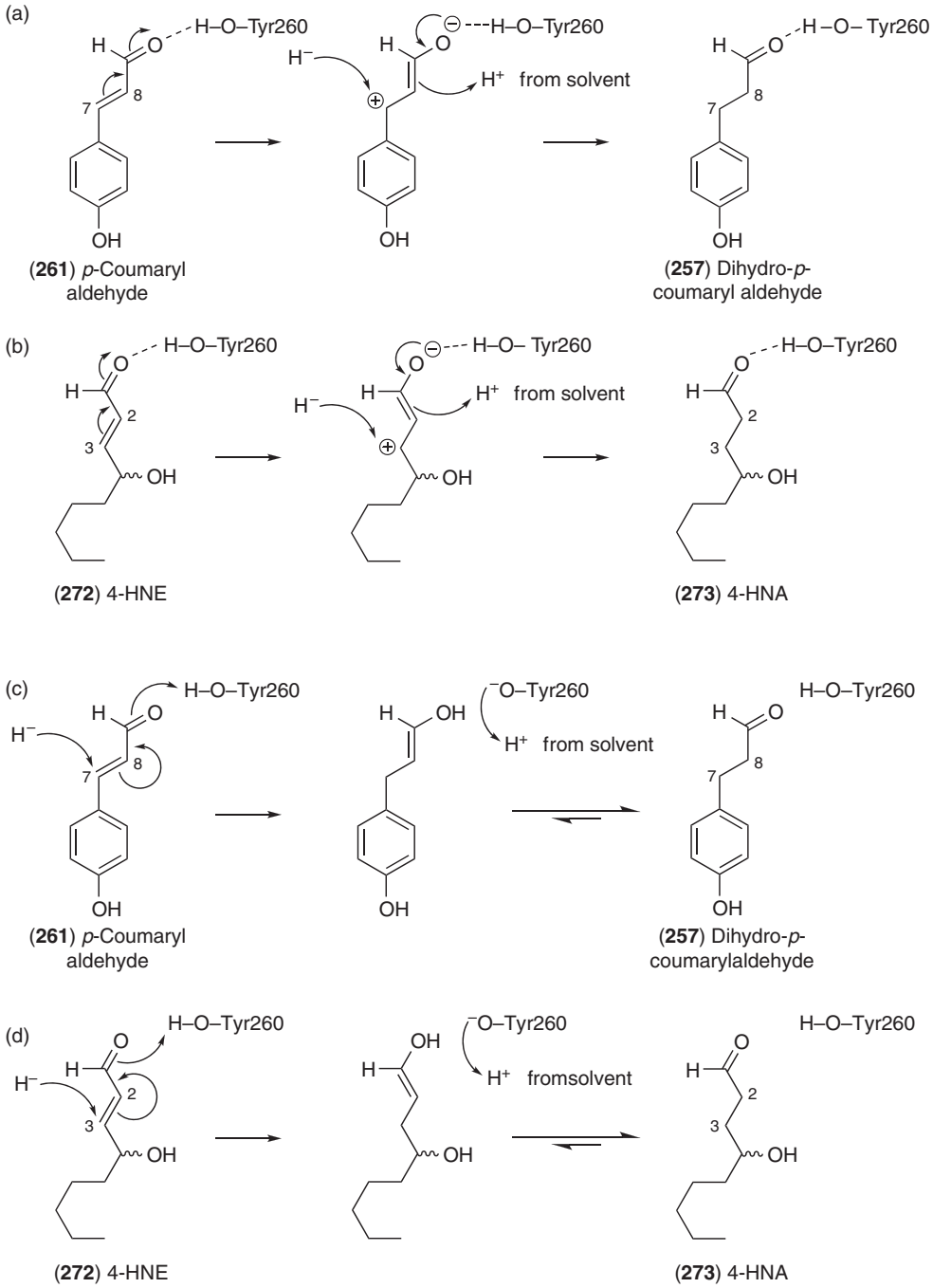


Figure 46 Proposed mechanisms for AtDBR1-catalyzed double bond reduction of *p*-coumaryl aldehyde (**261**) (a, c) and 4-HNE (**272**) (b, d). (a), (b) Hydride addition to an enolate carbocation intermediate Reproduced from B. Youn; S.-J. Kim; S. G. A. Moinuddin; C. Lee; D. L. Bedgar; A. R. Harper; L. B. Davin; G. Lewis; C. Kang, *J. Biol. Chem.* **2006**, *281*, 40076–40088. (c), (d) Concerted hydride addition to give the enol form of the product, with subsequent tautomerization to afford the final aldehyde.

In summary, while these data provide good insight into the overall catalytic mechanism of allylic double bond reductases, it must be emphasized that the actual physiological roles of the AtDBR multigene family (11 members) still remain to be fully established.

1.23.7.2 PLR Homologs: Phenylcoumaran Benzylic Ether Reductases, Isoflavone Reductases, and Pterocarpan Reductases

8–5′-Linked lignans (e.g., dihydrodehydrodiconiferyl alcohol triacetate (**280**, **Figure 47(a)**)) can co-occur with the phenylcoumaran benzylic ether-reduced homolog, tetrahydrodehydrodiconiferyl alcohol tetraacetate (**118**) in the needles of the gymnosperm *C. japonica*.[160] Reduction of the 7–*O*–4′ interunit linkage can thus be envisaged to occur through a similar biochemical process to that of PLR, that is, via presumed quinone methide intermediacy with hydride attack at C7.

Using *Forsythia* PLR cDNA as a probe, a PLR homolog was obtained from a *P. taeda* cell suspension cDNA library, this having ~45% amino acid identity to the *Forsythia* PLR_Fi1. Its full-length cDNA (PCBER_Pt1), which encoded for a 33.6-kDa protein was obtained, with the corresponding recombinant native protein expressed in *E. coli*.[265] This was demonstrated to catalyze *in vitro* the NADPH-dependent reduction of the racemic phenylcoumaran lignans, (±)-dehydrodiconiferyl alcohols (**14a**/**b**), and (±)-dihydrodehydrodiconiferyl alcohols (**103a**/**b**), into the corresponding racemic 7–*O*–4′-reduced products, (±)-isodihydrodehydrodiconiferyl alcohols (IDDC, **229a**/**b**), and (±)-tetrahydrodehydrodiconiferyl alcohols (TDDC, **281a**/**b**) (**Figure 47(b)**). The protein was thus 'provisionally' named phenylcoumaran benzylic ether reductase (PCBER), to reflect the mode of

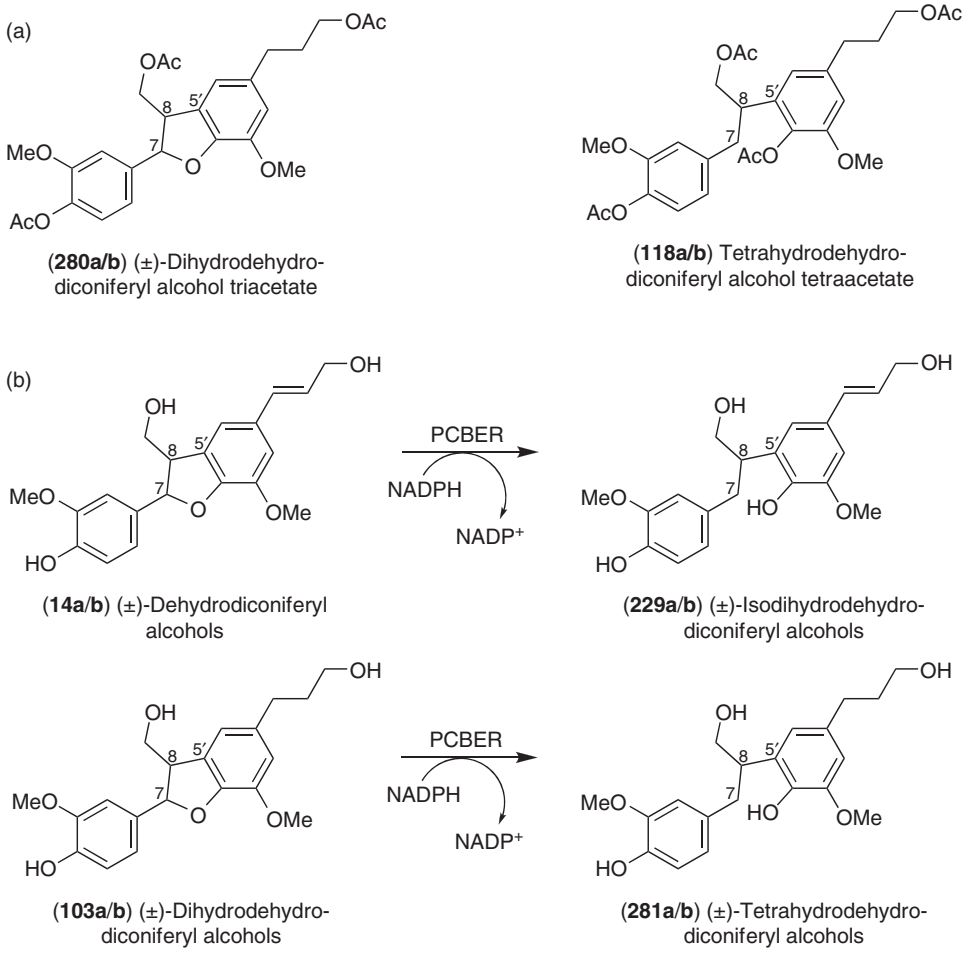


Figure 47 (a) (±)-Dihydrodehydrodiconiferyl alcohol triacetate (**280a/b**) and (±)-tetrahydrodehydrodiconiferyl alcohol tetraacetates (**118a/b**) from *Cryptomeria japonica* needles. (b) Reactions catalyzed by *Pinus taeda* phenylcoumaran benzylic ether reductase, PCBER.

catalysis detected, and was established to be a type A reductase (as for PLR and IFR). Interestingly, the recombinant PCBER was able to reduce both enantiomers of the respective substrates tested into their racemic products, thereby differing from PLR in terms of being regioselective, but not enantiospecific.[265] Two homologs (*PCBER_Cj79* and *PCBER_Cj80*) were also obtained from a *C. japonica* leaf cDNA library using *PCBER_Pt1* as a probe, these having ~83 and 87% similarity and ~72 and 80% identity to PCBER_Pt1, respectively (H. Kasahara, unpublished results).

While the recombinant PCBER product formation matched that present in *P. taeda* cell suspension cultures and *C. japonica*, the kinetic parameters obtained made the physiological function or functions tentative as repeatedly emphasized.[8,265] That is, kinetic parameters using (±)-dehydrodiconiferyl alcohols (**14a**/**b**) and (±)-dihydrodehydrodiconiferyl alcohols (**103a**/**b**) as potential substrates were very low for PCBER_Pt1 with K_m, V_{max}, and k_{cat}/K_m values of 0.61/1.95 mmol l^{-1}, 0.029/0.016 pkat μg^{-1} protein, and 1.59/0.27 $mol^{-1} l s^{-1}$ respectively. PLR substrates, (±)-pinoresinols (**13a**/**b**), however, were not efficiently processed by PCBER from either the *P. taeda* source or when using another PCBER homolog from *Populus trichocarpa*.[265]

1.23.7.2.1 Tissue localization

Localization of the putatively annotated PCBER was carried out at the tissue and cellular levels using polyclonal antibodies raised against both the *P. taeda*[291] and *P. trichocarpa*[324] enzymes. In *P. taeda* seedlings, it was apparently localized in the vascular cambium/differentiating secondary xylem regions of young developing shoots (tip of seedlings, **Figure 48(a)**), and in fully differentiated axial and radial parenchyma cells. It was also detected in the vascular cambium of mature woody stems (base of seedling, **Figure 48(b)**).[291] In *P. trichocarpa*, the putative PCBER was also localized in differentiating xylem, in young differentiating fibers and ray parenchyma cells both in juvenile (greenhouse-grown poplar) and mature (field-grown tree) wood.[324] Interestingly, it was considered to be the most abundant protein in *P. trichocarpa* secondary xylem, as demonstrated by two-dimensional gel electrophoresis.[325]

These data were thus again provisionally consistent with a role for the putatively annotated PCBER in lignan deposition, although other possibilities could not be ruled out. As repeatedly emphasized,[8,265] other substrates and functions for PCBERs also need to be considered. In this respect, there is the possibility that PCBERs might have multiple catalytic functions (see Section 1.23.9).

1.23.7.2.2 Structural biology studies of PLR homologs: PCBER, IFR, and pterocarpan reductases

Availability of recombinant PCBER provided an opportunity to consider its catalytic mechanism, as well as that of the related IFR. A PCBER_Pt1 crystal structure was thus obtained in apo-form (at 2.2 Å resolution)

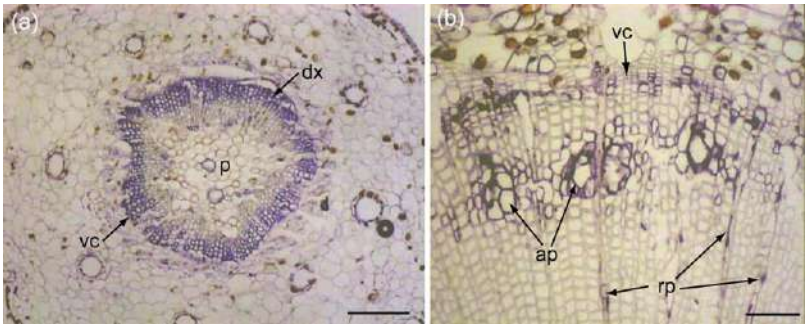


Figure 48 PCBER localization in *P. taeda* (1-year-old) seedlings. (a) PCBER in developing xylem regions of young green shoots. (b) PCBER was mainly present in axial and radial parenchyma cells, as well as in vascular cambium regions of basal stem sections. Abbreviations: ap, axial parenchyma; dx, developing xylem; p, pith; rp, radial parenchyma; vc, vascular cambium. Bars: 70 μm (a) and 400 μm (b). Reproduced from M. Kwon; L. B. Davin; N. G. Lewis, *Phytochemistry* **2001**, *57*, 899–914. Copyright 2001, with permission from Elsevier.

(Figure 30(b)), and not unexpectedly was similar to those of PLR_Tp2 and PLR_Tp1 (Figure 30(a)): PCBER_Pt1 also contained two domains separated by a cleft, with a conserved nucleotide cofactor-binding Rossman fold motif. As before, the highly conserved Lys134 residue was present in the active site, indicative of its conserved role during catalysis and the intermediacy of a presumed quinone methide (Figure 31(b)).

Another related protein, the *Medicago sativa* isoflavone reductase (IFR_Ms1) had high sequence homology to both PLR_Tp1 and PCBER_Pt1 (67/72% similarity and 44/55% identity). X-ray crystal structures from both were used to model, *in silico*, the 3D IFR_Ms1 structure.[266] Interestingly, energy minimization did not lead to significant changes from initial structural coordinates, with the extra residues of IFR falling into disordered loop regions. The structural features for PLR_Tp1 and PCBER_Pt1 were also evident in IFR_Ms1, for example, a Rossman fold motif, as well as the conserved Lys144 residue in the active site. The X-ray structure has also been confirmed by others (Figure 30(c)).[326]

The proposed IFR mechanism does not, however, involve a quinone methide intermediate via abstraction of a phenolic proton (as with PLR and PCBER). On the other hand, the conjugated enone system in the substrate (e.g., **230**) can undergo hydride addition, presumably to give an enol product, which then affords the more stable ketone tautomer (**231**) in solution (Figure 31(c)).

Interestingly, however, the isoflavonoid (−)-medicarpin (**232b**) can undergo a NADPH-dependent reduction to afford (−)-vestitol (**233b**) (Figure 31(d)). This reaction is catalyzed by pterocarpan reductase (PTR), and is considered to proceed via a quinone methide intermediate (Figure 31(d)).[327] All of the above reductases (PLRs, PCBERs, IFRs, and PTRs), together with leucoanthocyanidin reductases (LACRs), apparently form a phylogenetically and structurally related family of enzymes.

1.23.8 Norlignan Biosynthesis

While C8-decarboxylated norlignans, such as compounds **82**–**87** (Section 1.23.3.2.1), are found in bryophytes, most norlignans reported so far are from gymnosperms and angiosperms. For example, the gymnosperm *C. japonica* accumulates various 8–7′-linked norlignans, such as (*E*)-hinokiresinol (**123**), agatharesinol (**121**), cryptoresinol (**122**), and sequirin B (**124**).[142] The angiosperm asparagus, *Asparagus officinalis*, on the contrary, biosynthesizes the isomeric (*Z*)-hinokiresinol (**174**), as well as the presumed norlignan derivatives, the acetylenic alkyl–phenyl ethers, asparenydiol (**178**) and its *O*-methylated derivatives, asparenyol (**179**) and asparenyn (**180**).[207,328]

1.23.8.1 Hinokiresinol: Discovery of Biochemical Pathway, Encoding Genes, and Enzymes

The biosynthesis of (*E*/*Z*)-hinokiresinols (**123**/**174**) is now quite well understood, beginning with findings from radiolabeled precursor experiments using cell-free protein extracts from both *A. officinalis* and *C. japonica*.[168,207,329,330] Initially, the hinokiresinol isomers (**123**/**174**) were considered derived from two non-identical phenylpropanoid moieties, *p*-coumaryl alcohol (**199**) and a *p*-coumaroyl derivative (i.e., CoA ester **282**),[329,330] with this later being extended to *p*-coumaryl coumarate (**235**) as the presumed precursor.[168] Following consideration of possible biosynthetic mechanisms from **235** to (*Z*)-hinokiresinol (**174**) (Figure 49) initially from our laboratory[2] and then the Umezawa laboratory,[207] it was envisaged that a somewhat analogous process to allyl-/propenylphenols might be occurring (see Section 1.23.9.1).

Following initial precursor administration experiments, hinokiresinol synthase (HRS) was later purified from *A. officinalis* elicitor-treated cells.[207] After a six-step column chromatographic protocol, native PAGE suggested that it had been purified to apparent homogeneity, although SDS–PAGE gave two bands at ~21 and 23 kDa, respectively. Peptidase treatment, followed by HPLC purification and amino acid sequencing of the fragments obtained, suggested that both proteins were similar, these being named HRSα and HRSβ.

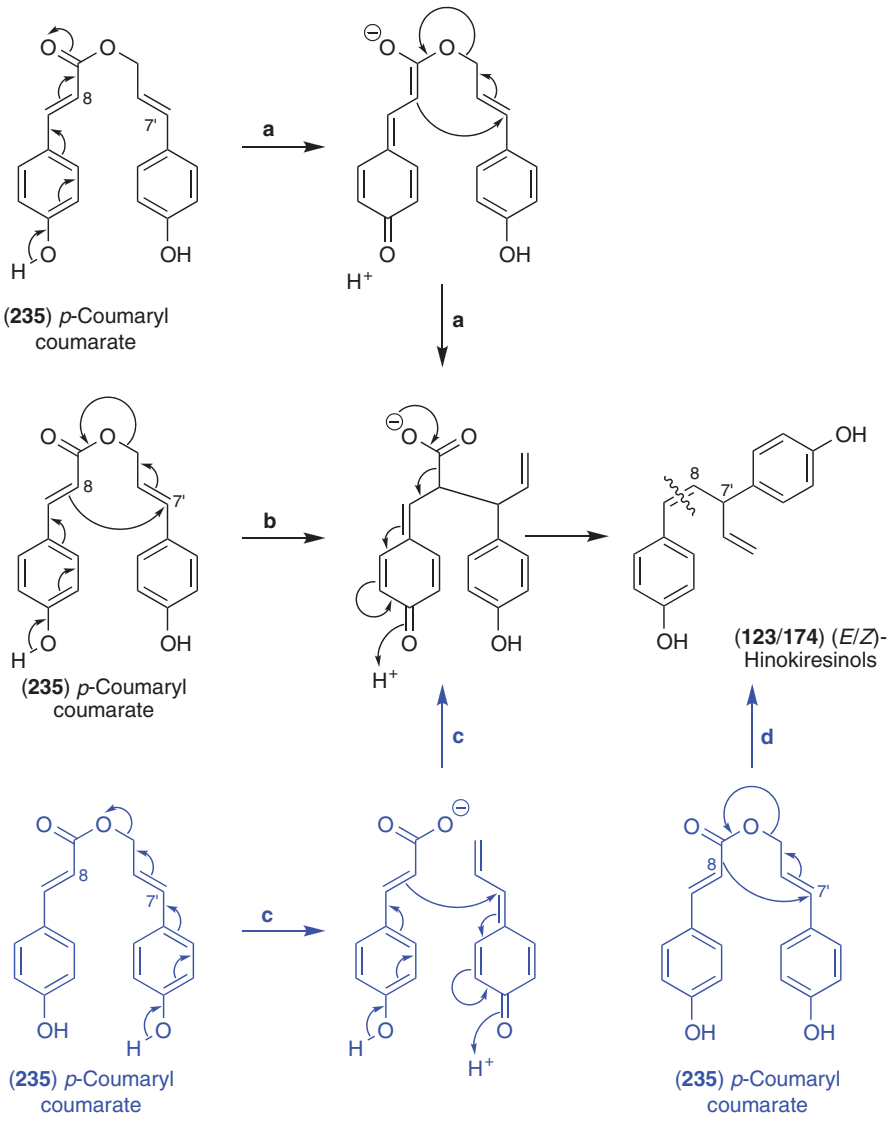
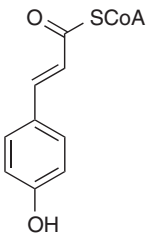


Figure 49 Proposed mechanisms for the formation of (*E*/*Z*)-hinokiresinols (**123**/**174**) from *p*-coumaryl coumarate (**235**) (Vassão *et al.*[2] in blue, Suzuki *et al*.[207] in black). Both (a) ester enolate Claisen rearrangement[207] and (b) two-step intramolecular rearrangement[207] involve the formation of a quinone methide in the *p*-coumarate moiety of the substrate, while route (**c**) involves ester link breakage with the formation of a quinone methide in the *p*-coumaryl (alcohol) moiety, followed by bimolecular coupling[2] to give the same *p*-coumarate quinone methide intermediate as (a) and (b). In all the three (a–c) proposed mechanisms, the *p*-coumarate quinone methide (center) undergoes decarboxylation with re-aromatization to give either (*E*/*Z*)-hinokiresinols (**123**/**174**). Route (d) involves direct intramolecular rearrangement, with concerted decarboxylation and C7′–C8 bond formation without intermediacy of a quinone methide (i.e., no participation of the phenolic rings).[2]



(**282**) *p*-Coumaroyl CoA

The corresponding sequenced fragments were then used to obtain full-length cDNAs, with these encoding ~20.4- and 19.8-kDa proteins for HRSα and HRSβ, respectively. *In silico* searches indicated that several homologs in literature databases were annotated as either phloem protein 2 (PP2) or PP2-like, but with no known physiological/biochemical function(s). Nevertheless, these form a gene superfamily of which HRSs is part of a clade: HRSα and HRSβ were also ~49% identical to each other based on amino acid sequence, and the corresponding fully functional recombinant proteins were expressed in *E. coli* as His-tag proteins.

Interestingly, recombinant HRSα and HRSβ when individually assayed with *p*-coumaryl coumarate (**235**), afforded (*E*)-hinokiresinol (**123**). By contrast, *A. officinalis* accumulated the corresponding *Z*-isomer **174**, and the native HRS preparation also catalyzed the formation of this isomer as well. The (*Z*)-isomer was, however, produced when both HRSα and HRSβ isoforms were assayed together in an 1:1 ratio. These data, together with gel filtration analysis, suggested that *A. officinalis* HRS was a heterodimer of HRSα and HRSβ. Kinetic parameters for the HRS heterodimer gave an apparent K_m of 0.44 μmol l^{-1}, a V_{max} of ~0.075 pkat μg^{-1} protein, a k_{cat} of ~1.5×10^{-3} s^{-1}, and a k_{cat}/K_m of ~3400 mol l^{-1} s^{-1} for *p*-coumaryl coumarate (**235**) (considering one active site per monomer). The kinetic data were thus also consistent with those of monolignol-related pathway enzymes as indicated earlier.

From a mechanistic perspective, HRS catalysis (**Figure 49**) involves decarboxylation and 8–7′ bond formation, and requires no additional cofactors. Four mechanisms have now been proposed as possible: an ester enolate Claisen rearrangement (**Figure 49(a)**); either a concerted (**Figure 49(d)**) or a two-step (**Figure 49(b)**) molecular rearrangement; and/or an ester cleavage followed by bimolecular coupling (**Figure 49(c)**).[2,207] In three of these mechanisms (**Figures 49(a)**–**49(c)**), a putative quinone methide intermediate is considered to be generated.

1.23.8.2 Agatharesinol

Agatharesinol (**121**) is a norlignan closely related structurally to (*E*)-hinokiresinol (**123**), differing only by the addition of two hydroxyl groups to the pendant olefinic moiety. It is constitutively found in *C. japonica* heartwood, but its formation can also be induced in sapwood tissue, since sapwood from freshly felled trees, when allowed to stand at room temperature, slowly generates agatharesinol (**121**).

Somewhat surprisingly, agatharesinol (**121**) biosynthesis has been reported as not involving (*E*)-hinokiresinol (**123**) as an intermediate,[331] although this was not actually proven in these studies. Using the norlignan-inducible sapwood tissue, administration of [ring-^{2}H]-Phe (**2**), [ring-^{13}C]-(*E*)-cinnamic acid (**249**), [9-$^2H_2$]-*p*-coumaryl alcohol (**199**), and [9-$^2H_2$]-(*E*)-hinokiresinol (**123**) gave somewhat conflicting results.[331,332] While [ring-^{2}H]-Phe (**2**) and [ring-^{13}C]-(*E*)-cinnamic acid (**249**) were intactly incorporated into agatharesinol (**121**), [9-$^2H_2$]-*p*-coumaryl alcohol (**199**) was only converted into [9′-$^2H_2$]-(*E*)-hinokiresinol (**123**) but not agatharesinol (**121**). On the other hand, uptake of [7-^{2}H]-(*E*)-hinokiresinol (**123**) did not afford deuterium-labeled agatharesinol (**121**), prompting these researchers to conclude that a different biosynthetic pathway to (*E*)-hinokiresinol (**123**) must be operative.

This conclusion cannot, however, be made from such data: For example, both [9-$^2H_2$]-*p*-coumaryl alcohol (**199**) and [7-^{2}H]-(*E*)-hinokiresinol (**123**) may simply not have been efficiently translocated to the site(s) of agatharesinol (**121**) biosynthesis, whereas both [ring-^{2}H]-Phe (**2**) and [ring-^{13}C]-(*E*)-cinnamic acid (**249**) were.

Interestingly, immunolocalization of agatharesinol (**121**) was examined in *C. japonica* heartwood tissue,[333] using bovine serum albumin-linked agatharesinol (**121**) to raise polyclonal antibodies from rabbits.[334] The antiserum so obtained detected agatharesinol (**121**) in ray parenchyma cells of heartwood, as well as in inner walls of some of the adjacent tracheid cells[333] (**Figure 50**). This finding was the first immunolocalization study of a heartwood (norlignan) compound, even though it is well known that lignans accumulate in such cell types.

1.23.8.3 Acetylenic Norlignans

The biosynthetic pathways to the more unusual acetylenic norlignans found in *A. officinalis* and *Hypoxis* spp. are less well understood, although preliminary radiolabeling precursor studies have been performed. *Hypoxis* corms

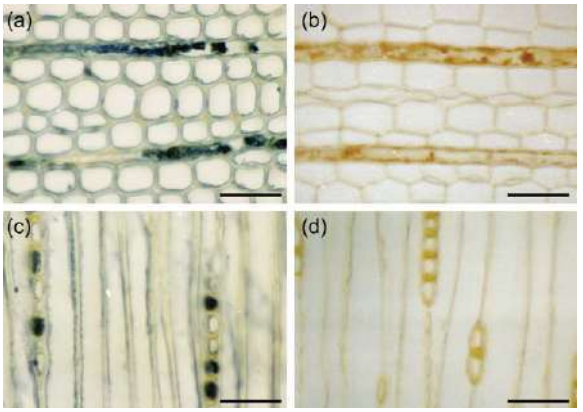


Figure 50 Immunolocalization of agatharesinol (**121**) in *Cryptomeria japonica*. (a), (b) Cross-section of heartwood tissue treated with (a) antiagatharesinol antiserum and (b) pre-immune serum. (c) (d) Tangential section of heartwood tissue treated with (c) antiagatharesinol antiserum and (d) pre-immune serum. Reproduced from T. Nagasaki; S. Yasuda; T. Imai, *Phytochemistry* **2002**, *60*, 461–466. Copyright 2002, with permission from Elsevier.

were established to be the main sites of hypoxoside (**182**) biosynthesis[335] with whole plants able to incorporate radiolabeled Phe (**2**), cinnamic acid (**249**), *p*-coumaric acid (**250**), caffeic acid (**56**), and acetate into hypoxoside (**182**) – albeit with different efficacies.[336,337] Detailed incorporation studies with stable isotope-labeled precursors in *A. officinalis* also established that both aromatic rings (as well as the butyne moiety) of asparenyn (**180**), asparenyol (**179**), and asparenydiol (**178**) were shikimic acid-derived through Phe (**2**).[211,338,339] A preliminary biosynthetic sequence for their formation was proposed (but without experimental confirmation) to involve a hinokiresinol (**123**/**174**)-like skeleton and a spiro intermediate.[211] To date, no enzymes or genes have yet been reported for their formation.

1.23.9 Allyl-/Propenylphenol Biosynthesis

As indicated earlier (Section 1.23.3.1), allyl-/propenylphenol biosynthetic pathways are present in a wide range of plant species from algae (at least provisionally), to the hornworts and liverworts, to the ferns, gymnosperms, and angiosperms. The apparently sporadic distribution of their occurrence, however, raises the possibility that their biochemical pathway or pathways may have evolved independently at different times. This section now addresses deduction of the biochemical pathway(s) to this most interesting class of monomeric – and frequently dimeric – natural products, and some of the twists and turns that occurred.

As previously discussed, one potential complication, at least from a chemotaxonomical perspective, is that monomeric allyl-/propenylphenols are often found in near trace amounts. Indeed, such occurrences are generally only reported in tabular form as very minor constituents of complex mixtures of terpenoids and the like. This is a complication because such reports are currently more difficult to trace in the literature when documented in this way. Nevertheless, the recent discovery of their biochemical pathways raises another interesting issue. This is in terms of some of the enzymes involved being rather substrate versatile, and thus perhaps able to participate in apparently unrelated (multiple) biochemical pathways. This potentially adds yet another level of complexity to this branch of plant metabolism.

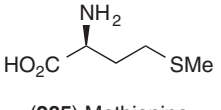
1.23.9.1 Deduction of Allyl-/Propenylphenol (Monomeric and Dimeric) Biosynthetic Pathways

Over the past few years, the biosynthetic route leading to allyl-/propenylphenols *in planta* has been fully elucidated.[2–5] This settled a question that had remained unresolved for more or less five decades of enquiry, including whether the C9 (terminal) carbon was lost or not.

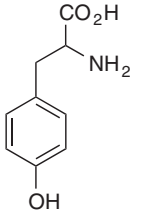
Interest in allyl/propenyl biosynthetic pathways can be traced back to the chemical studies by Birch and Slaytor in 1956.[340] These researchers very briefly reported the chemical conversion of 3,4-methylenedioxycinnamyl alcohol (**283**) into a mixture of safrole (**22**) and isosafrole (**24**) (~60% yield in an ~55:45 ratio), by action of $LiAlH_4$ in boiling ether in the presence of $AlCl_3$ (Lewis acid) (**Figure 51(a)**). This conversion was presumed to occur via generation of a resonance-stabilized allylic carbocation intermediate following loss of the terminal hydroxyl group, with concomitant reduction via hydride addition. A similar conversion did not occur, however, with cinnamyl alcohol (**284**, **Figure 51(b)**), suggesting to these researchers that a *para*-oxygenated aromatic ring moiety (although not necessarily a free phenol) was needed for the displacement of the terminal hydroxyl group. A somewhat analogous generation of a resonance-stabilized carbocation intermediate was thus initially postulated for the corresponding biosynthetic processes.

1.23.9.1.1 Radiolabel tracer studies: controversy over intact incorporation of monolignol pathway intermediates and scientific judgment?

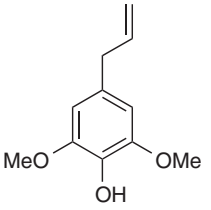
The first biosynthetic studies toward gaining an understanding of allyl-/propenylphenol biosynthesis were reported in the early 1960s, with Kaneko[341–344] demonstrating, via a combination of radiolabeling precursor administration and chemical degradation approaches, that [2-^{14}C]-Phe (**2**), [9-^{14}C]-*p*-coumaric acid (**250**), and [Me-^{14}C]-Met (**285**) were incorporated into anethole (**5**) in fennel (*F. vulgare*). Further demonstration of phenylpropanoid pathway intermediates being involved was also reported by others in banana[345] and basil,[346] respectively. Specifically, [9-^{14}C]-caffeic acid (**56**) and [1-^{14}C]-Phe (**2**) were incorporated into eugenol (**4**), methyleugenol (**42**), 5-methoxyeugenol (**287**), and elemicin (**45**) in banana disks. Administration of double-labeled [9-^{14}C, OC^{3}H$_3$]-ferulic acid glucoside (**288**), as well as [9-^{14}C, OC^{3}H$_3$]- and [9-^{14}C, ^{3}H$_2$OH]-coniferyl alcohol glucoside (**289**) in basil also gave ^{3}H-/^{14}C-labeled eugenol (**4**) and methyleugenol (**42**) with only minor changes in ^{3}H:^{14}C ratios between administered substrates and products. Confirmation of the C9 label in the allylphenols **4** and **42** was determined by degradation studies. These data were thus consistent with retention of the intact phenylpropanoid skeleton into allyl-/propenylphenols.



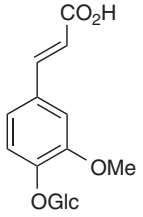
(**285**) Methionine



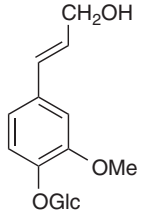
(**286**) Tyrosine



(**287**) 5-Methoxyeugenol



(**288**) Ferulic acid glucoside



(**289**) Coniferyl alcohol glucoside

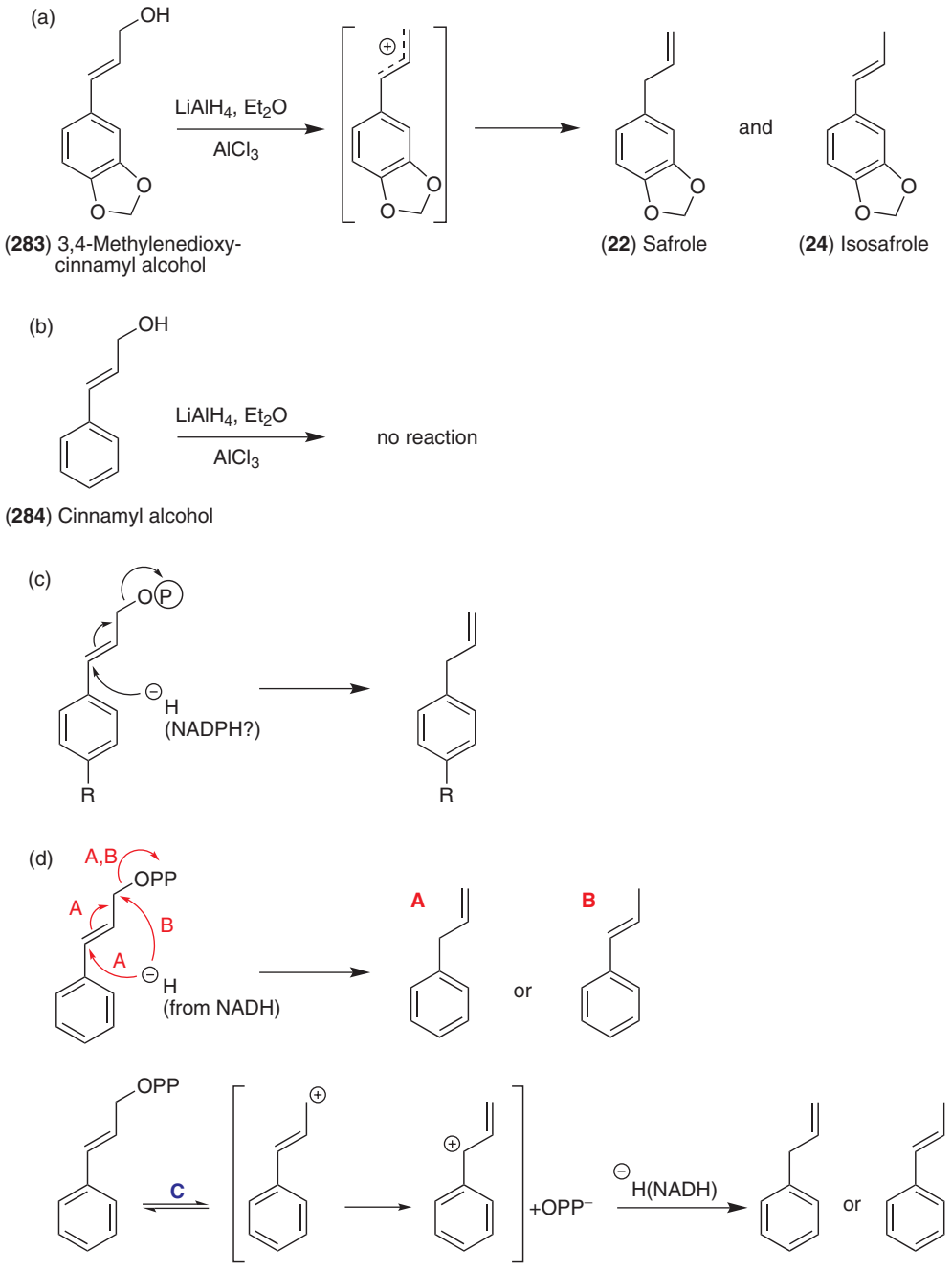


Figure 51 Proposed chemical and biochemical mechanisms leading to allyl-/propenylphenols. (a), (b) Chemical formation of (a) safrole (**22**) and isosafrole (**24**) from 3,4-methylenedioxycinnamyl alcohol (**283**), in presence of $LiAlH_4$ and $AlCl_3$, and (b) lack of reaction when cinnamyl alcohol (**284**) was used, as reported by Birch.[340] (c) Birch's refined biochemical mechanism for the formation of allylphenols (R = not specified), with nucleophilic (allylic) displacement of a phosphorylated group by H^- (presumed derived from NADPH).[347] (d) Biochemical mechanisms proposed by Geissman and Crout,[348] via either allylic (A) or direct (B) nucleophilic displacement of pyrophosphate by H^- (from NADH) to afford the corresponding allyl- or propenylphenol, or (C) through a S_N1 mechanism and delocalized carbocation intermediate with hydride reduction.

In further postulates to account for allyl/propenyl formation, Birch in 1963[347] next proposed a mechanistic rationale for displacement of the terminal hydroxyl group (**Figure 51(c)**). Based partly on comparison with other metabolic pathways, the terminal hydroxyl group was now anticipated to require activation for more

facile displacement. This consideration was based on parallels to terpenoid biochemistry, that is, involving the formation and metabolism of isopentenyl pyrophosphates. Accordingly, a monolignol-derived phosphate ester derivative was envisaged, with allylic addition of a hydride leading to displacement thereby generating the corresponding allylphenol. Geissman and Crout[348] also later proposed a nucleophilic addition (of a hydride from NADH) to a (hydroxy)cinnamyl pyrophosphate, via either direct or allylic displacement of the pyrophosphate group, or alternatively via an S_N1 mechanism with the formation and reduction of a delocalized allylic carbocation (**Figure 51(d)**).

Conversely, there were also various accounts of C9 carbon loss occurring during the formation of eugenol (**4**) in both basil[349,350] and cinnamon (*Cinnamon zeylanicum*).[351] That is, during the 1970s, administration of double-labeled Phe (**2**) (nonspecifically tritiated, as well as containing either 1-, 2-, or 3-^{14}C) was carried out using basil, with the corresponding radiolabeled forms of eugenol (**4**) individually isolated. A much higher ^{3}H:^{14}C radiolabel ratio (~20 times than that of the ^{3}H:^{14}C precursor) was reported for eugenol (**4**) when the ^{14}C label was in the carboxylic acid group of Phe (**2**). This suggested the loss of the terminal C9 carbon during metabolism of Phe (**2**) into eugenol (**4**). Additionally, in cinnamon, the radiolabel from [1-^{14}C]-Phe (**2**) was reportedly poorly incorporated into eugenol (**4**, 0.003%), relative to marginally better incorporations with either [2-^{14}C]- or [3-^{14}C]-Phe (**2**) (0.047 and 0.098%, respectively). [Me-^{14}C]-Met (**285**) also allegedly led to eugenol (**4**, 0.1% incorporation) radiolabeled at the C9 position (based on chemical degradation), suggesting a methionine (**285**) origin for this particular carbon.

It must be emphasized that most of these radiotracer studies generally suffered from low incorporations, at best 0.5% and most often much less than 0.1%. Such low incorporation data (i.e., <0.1%) are, however, generally suspect. Nevertheless, in spite of the rather flimsy radiochemical data in support of C9 loss, this pathway was still being considered by some researchers as late as 2001,[352] that is, whereby an additional carbon (presumed to be derived from *S*-adenosyl methionine, SAM) was considered covalently attached to a *p*-coumaric acid (**250**) moiety to form cyclopropyl intermediate **290** (**Figure 52**). This putative intermediate was then envisaged to undergo C9 decarboxylation with ring opening to directly afford the allyl/propenyl side chains of chavicol (**20**) and *p*-anol (**21**). This, however, seemed an unlikely possibility from a biochemical standpoint. Indeed, the very weak evidence for C9 loss can be considered as a question of scientific judgment, when compared to the evidence for intact incorporation.

1.23.9.1.2 Intermediacy of monolignol esters in allyl-/propenylphenol biosynthesis: clues from norlignans?

Our long-term interests in allyl-/propenylphenol biosynthesis encompass how the formation of both the monomeric and dimeric natural products occurs. Accordingly, particular attention was given to plant species such as basil, creosote bush, and various *Piper* spp. These distinct plant species can differentially harbor allyl-/propenylphenols in either monomeric, dimeric, or both forms, and thus provide the means to establish any parallels and differences for biosynthesis of the monomers and dimers. The biosynthetic pathway findings made to date are summarized below, beginning with basil, which typically accumulates the monomeric derivatives, methylchavicol (**3**) and eugenol (**4**).

The Thai basil variety was chosen for biosynthetic studies as it largely accumulates methylchavicol (**3**) in its glandular trichomes;[2] basil is not known, however, to accumulate any lignans in those cell types. Various labeled

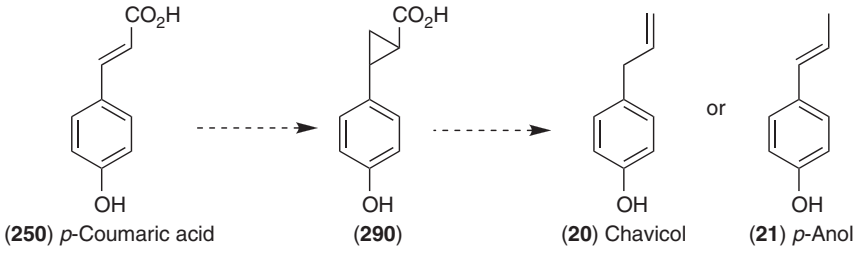


Figure 52 Putative biochemical pathway to chavicol (**20**)/*p*-anol (**21**) from *p*-coumaric acid (**250**) via the unusual cyclopropyl intermediate (**290**) by Gang *et al*.[352] (No evidence for pathway exists.)

precursor administration/incorporation experiments were thus performed using young apical basil leaves,[2] which we had established to be the most biosynthetically active in terms of producing methylchavicol (**3**). Specifically, [U-^{14}C]-Phe (**2**), [8-^{14}C]-cinnamic acid (**249**), and [9-^{3}H]-*p*-coumaryl alcohol (**199**), but not [U-^{14}C]-Tyr (**286**), were metabolized into methylchavicol (**3**) *in vivo*, albeit to differing levels of efficacy (~1–4% by 24 h). These data thus again eliminated the notion of a need for C9 loss, such as via a cyclopropyl intermediate (**290**, **Figure 52**), from further consideration as had been previously proposed.[352] Using cell-free extracts, however, *p*-coumaryl alcohol (**199**) (either unlabeled or [9-^{3}H]-labeled) was not directly converted to afford chavicol (**20**) *in vitro*, suggesting that additional enzymatic steps and/or then unknown cofactors were needed.

Three routes for *p*-coumaryl alcohol (**199**) transformation were next concurrently considered as being possibly viable (**Figure 53**): (a) side-chain double bond reduction initially to afford dihydro-*p*-coumaryl alcohol (**291**) (i.e., as for the double bond reductase described above in Section 1.23.7.1), followed by dehydration; (b) *O*-methylation of the phenolic group to afford *p*-methoxyl cinnamyl alcohol (**292**), with this possibly preceding deoxygenation (as proposed in Kaneko[344]); and (c) activation of the side-chain hydroxyl group to afford a *p*-coumaryl alcohol (**199**) derivative bearing a better leaving group. Routes a and b were eliminated following *in vivo* precursor administration experiments, as well as using *in vitro* incubations with crude cell-free enzyme preparations.[2]

Potentially activated forms of the monolignol side-chain hydroxyl group were next examined using cell-free extracts *in vitro*, for example, by coincubation of monolignols with ATP, GTP, UTP, Glc-6-phosphate, and/or glutathione, but with no success. At about the same time, however, the monolignol ester *p*-coumaryl coumarate (**235**) was demonstrated to serve as substrate in norlignan biosynthesis, to afford *E*- and *Z*-hinokiresinols (**123** and **174**) in *A. officinalis* and *C. japonica*,[168,329,330] respectively (Section 1.23.8.1). Although this conversion apparently took place in the absence of any external cofactor, no biochemical/organic chemical mechanism was proposed. Therefore, a possible mechanism[2] was considered (**Figure 54(a)**) to explain the formation of **123**/**174**, which, in turn, led to the formulation of another mechanism for allyl-/propenylphenol production (**Figure 54(b)**).

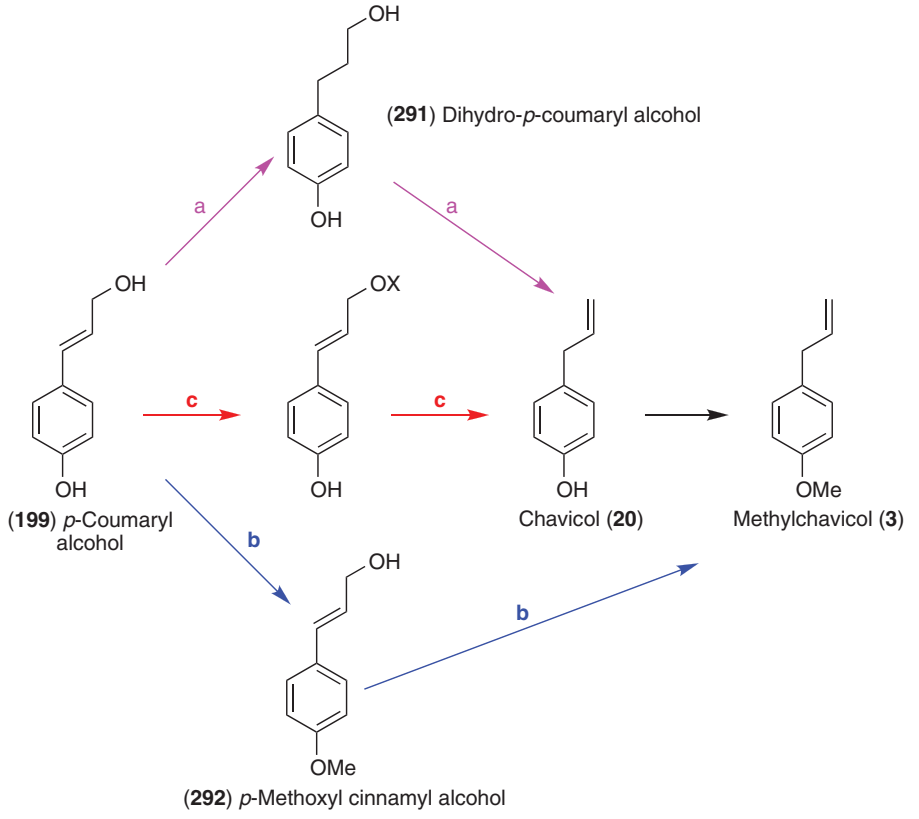


Figure 53 Possible biosynthetic pathways to chavicol (**20**) and methylchavicol (**3**). (X = facile leaving group.)

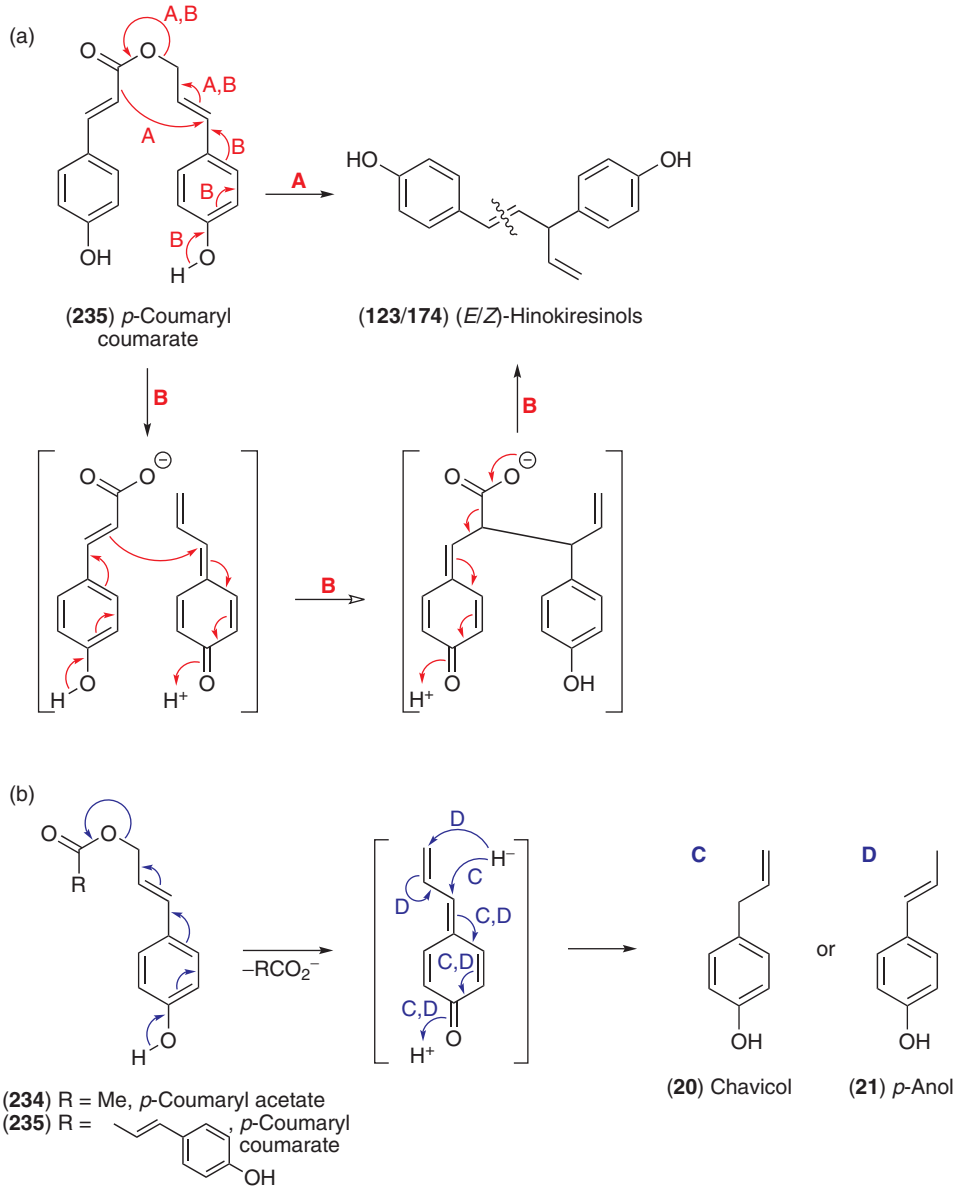


Figure 54 Possible mechanisms for conversion of *p*-coumaryl esters into (a) (*E*/*Z*)-hinokiresinols (**123**/**174**) and (b) chavicol (**20**)/*p*-anol (**21**). (A) Concerted; (B) ester cleavage, followed by cyclization, decarboxylation, and rearomatization; (C) and (D) ester displacement, putative quinone methide formation with subsequent reduction by hydride (from NAD(P)H), and rearomatization to form (C) chavicol (**20**) and (D) *p*-anol (**21**). In (C) and (D), the acid moiety may be interchangeable.

That is, hydride addition to a putative activated monolignol (e.g., *p*-coumaryl coumarate (**235**)) bound in the enzyme active site could potentially lead to C9 deoxygenation through ester displacement (e.g., to afford chavicol (**20**) and *p*-anol (**21**), respectively, and/or homologs thereof) (**Figure 54(b)**). In this way, the (*p*-coumarate) ester functionality might, therefore, serve as a good leaving group, perhaps being displaced by an incoming hydride through nucleophilic addition directly at either C7 or C9, or through an intermediary quinone methide, as for PLR (**Figure 31(a)**), to form the isomeric allyl-/propenylphenols. Such a process would, however, also require a reductive step, for example, with NADH/NADPH as cofactor.

Crude cell-free protein extracts from basil glandular trichomes were thus next demonstrated to be able to convert *p*-coumaryl coumarate (**235**) into chavicol (**20**) *in vitro*, in the presence of either NADPH or NADH.[2]

Additional experiments[3] (Vassão, unpublished results) established that acetylated derivatives (*p*-coumaryl acetate (**234**), coniferyl acetate (**236**)) were also utilized as substrates to afford chavicol (**20**) and eugenol (**4**), respectively, that is, thereby demonstrating, at a minimum, the substrate versatile nature of these transformations. Analogous processes occurred in the creosote bush,[4] as discussed below using the very same substrates (**234**–**236**) (**Figure 31**). These data, therefore, resolved the nature of the biochemical pathway to this class of molecules. Taken together, the findings demonstrated that allyl-/propenylphenol formation occurred through a two-step process: Monolignol activation by acylation, and ester displacement by means of an NAD(P)H-dependent conversion, presumably involves quinone methide intermediacy, as for PLR, PCBER, and so on.[2] The studies did not, however, unambiguously clarify the precise chemical identity of the ester-cleaved moiety (leaving group).

1.23.9.2 Allyl-/Propenylphenol Synthases

1.23.9.2.1 Bifunctional chavicol/eugenol and p*-anol/isoeugenol synthases (CES and AIS): The twists and turns to biochemical clarity*

Subsequent identification of genes encoding chavicol/eugenol and *p*-anol/isoeugenol synthases (CES and AIS) was somewhat facilitated by recognition that suppression of a transcription factor (*ODORANT1*) in petunia upregulated expression of a PLR/PCBER/IFR homolog of unknown biochemical function.[353] A full-length sequence for this homolog was obtained from a petunia petal EST collection, and encoded an ~36 kDa peptide with amino acid sequence similarities/identities of ~59/40% to that of PLR_Tp2 from *T. plicata*, ~62/42% to a PLR homolog isolated from the creosote bush in 1999, ~61/41% to the *P. taeda* PCBER, as well as ~59/37% to an IFR from *M. sativa*.[3]

A creosote bush PLR homolog, isolated from a cDNA library obtained from leaves actively producing allyl-/propenylphenol-derived lignans, also encoded a protein of ~35 kDa, but whose biochemical function was not initially unambiguously established when first isolated in 1999. Additionally full-length cDNA for a basil homolog was obtained (based on the petunia PLR/PCBER/IFR homolog above) using data from a basil trichome EST database, and which encoded a 35.6 kDa protein.[3]

Escherichia coli cells were then next transformed to obtain the corresponding petunia and basil recombinant proteins, in addition to that already available for the creosote bush PLR homolog. Each was examined for its ability to afford allyl- or propenylphenols when incubated with either *p*-coumaryl or coniferyl acetate/*p*-coumarate esters (**234**–**236**).[2–4]

1.23.9.2.1(i) Creosote bush chavicol/eugenol synthase (CES) Although this ultimately represented the second open literature report of a functional chavicol/eugenol synthase (CES), for the benefit of the reader, this is described first. As discussed earlier in Section 1.23.5.4, it is of interest that the creosote bush massively accumulates numerous lignans, all 8–8′-linked, such as NDGA (**143**). However, there are no reports of other skeletal types, including 8–5′- or 8–*O*–4′-linked lignans from this species. Nor is there any 'substantial' accumulation of monomeric allyl-/propenylphenols, such as chavicol (**20**)/eugenol (**4**) and *p*-anol (**21**)/isoeugenol (**6**). As indicated earlier though, monomeric allyl-/propenylphenols often occur in trace amounts in different plant species and thus may go long undetected from a chemotaxonomical perspective.

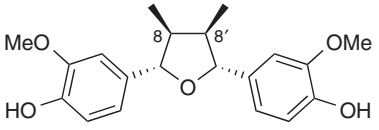
The creosote bush PLR homolog efficiently converted, in the presence of NADH/NADPH, all of the esters assayed to afford either chavicol (**20**) or eugenol (**4**), respectively, and thus became the first substrate versatile chavicol/eugenol synthase (CES) to be characterized. It efficiently converted *p*-coumaryl coumarate (**235**), *p*-coumaryl acetate (**234**), and coniferyl acetate (**236**); apparent K_m values of 210, 350, and 290 $\mu mol\,l^{-1}$, V_{max} values of 75, 200, and 190 $pkat\,\mu g^{-1}$ protein, and k_{cat}/K_m of 12 000, 19 500, and 22 000 $mol^{-1}\,l\,s^{-1}$, were obtained for substrates **235**, **234**, and **236**, respectively (**Table 1**). These data were thus comparable to the catalytic efficacies noted earlier for PLR, and for other monolignol pathway enzymes.

On the contrary, the creosote bush CES showed at best marginal activity toward a number of other potential substrates, including pinoresinol (**13**), larreatricin (**200**), nectandrin B (**293**), licarin A (**88**), dehydrodiconiferyl alcohol (**14**), as well as the isoflavones genistein (**294**) and biochanin A (**295**), as

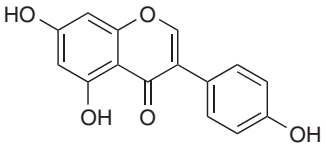
Table 1 Reported kinetic properties of CES and AIS

Enzyme function	*Trivial name*	*Source*	*Substrate*	K_m (μM)	V_{max} (*pkat* μg^{-1} *prot.*)	k_{cat} (s^{-1})	k_{cat}/K_m ($mol^{-1}\ l\ s^{-1}$)
CES	LtCES1[4]	Creosote bush	**234**	350	200	6.8	19 500
CES	LtCES1[4]	Creosote bush	**235**	210	75	2.55	12 000
CES	LtCES1[4]	Creosote bush	**236**	290	190	6.46	22 000
CES	ObEGS1[3]	Basil	**236**	5100	20	0.7	160
CES	CbEGS1[354]	*Clarkia breweri*	**236**	93	7.3	0.26	2 800
CES	CbEGS2[354]	C. *breweri*	**236**	311	6.9	0.25	800
CES	PhEGS1[354]	Petunia	**236**	245	18.4	0.6	2 400
AIS	PhIGS1[3]	Petunia	**236**	1600	7	0.3	136
AIS	PhIGS1[354]	Petunia	**236**	226	35.7	1.3	5 700
AIS	CbIGS1[354]	C. *breweri*	**236**	212	27.6	0.99	4 700
AIS	PaAIS1[356]	Anise	**236**	230	28.2	1.02	4 440
AIS	PaAIS1[356]	Anise	**234**	135	1.9	0.07	520

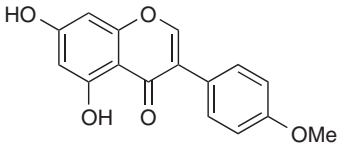
previously reported.[4] These data therefore demonstrated that the creosote bush CES (LtCES1) was unable to function as an efficacious PLR, PCBER, or IFR.[4] Surprisingly, a later report[354] stated that the creosote bush CES had not been assayed for PCBER activity. These researchers[354] were in error with such assertions, as this had been comprehensively examined and described,[4] with compelling evidence obtained only for CES activity (see above).



(**293**) Nectandrin B



(**294**) Genistein



(**295**) Biochanin A

In addition to CES, the creosote bush contains a *p*-anol/isoeugenol synthase, AIS (LtAIS1) with ~63/43% similarity/identity to LtCES1, which is capable of efficiently converting *p*-coumaryl (**234**, **235**)/coniferyl (**236**) esters to afford *p*-anol(**21**)/isoeugenol (**6**), respectively (S.-J. Kim, unpublished results). This demonstrated that this organism harbors genes encoding proteins capable of forming both allyl- and propenylphenols.

1.23.9.2.1(ii) **Piper regnellii** This *Piper* species biosynthesizes both allylphenols (e.g., myristicin (**23**), apiol (**36**), and dillapiol (**38**)), as well as presumed propenylphenol-derived lignans, such as (+)-conocarpan (**16a**), (−)-*epi*-conocarpan (**188b**), and eupomatenoid-6 (**191**).[218] It was, therefore, of interest to also investigate whether this organism was capable of biosynthesizing allyl- and/or propenylphenols in either monomeric or dimeric form, or both. As indicated below, this has also been established: For example, two recombinant *P. regnellii* enzymes, PrCES1 and PrCES2, with ~82/71% similarity/identity to LtCES1, are able to efficiently catalyze monomeric allylphenol formation efficiently *in vitro* (S.-J. Kim, unpublished results), thereby providing the needed insight into the biochemical processes affording each class of metabolites (monomers and dimers) within this organism.

1.23.9.2.1(iii) Basil and petunia chavicol/eugenol and* p*-anol/isoeugenol synthases (CES and IGS): Kinetic parameter inconsistencies The initial report[3] on basil and petunia CES and AIS was unfortunately both ultimately misleading and, in parts, in error. The basil enzyme was mistakenly restricted to being a eugenol synthase (ObEGS1) and the petunia enzyme as an isoeugenol synthase (PhIGS1).[3] This was mistakenly stated even though the earlier basil work[2] had already established that chavicol (**20**) formation could result from the incubation of cell-free extracts with either *p*-coumaryl acetate (**234**) or *p*-coumaryl coumarate (**235**) and either NADPH or NADH. It is more accurate to describe these enzymes as being, (at a minimum, substrate-versatile) as for the creosote bush CES.[4]

More problematic in the original account[3] was the kinetic data reported for recombinant CES/AIS from basil/petunia. That is, the kinetic values obtained for coniferyl acetate (**236**) were indicative of very poor efficacies: These had apparent K_m values of 5.1 and 1.6 mmol l^{-1}, V_{max} of ~20 and 7 pkat μg^{-1} protein, and k_{cat}/K_m of 160 and 136 $mol^{-1}\,l\,s^{-1}$ for the basil (ObEGS1) and petunia (PhIGS1) enzymes, respectively (**Table 1**). The overall k_{cat}/K_m values for **236** were therefore lower by nearly 2 orders of magnitude than those of the creosote bush CES (LtCES1), that is, 22 000 $mol^{-1}\,l\,s^{-1}$ for the latter *versus* 160 and 136 $mol^{-1}\,l\,s^{-1}$ for the basil and petunia enzymes. Although cognizant at the time of the report[3] that the basil/petunia kinetic data were not compelling (relative to PLR, etc.), this was provisionally rationalized due to uncertainty of the true identity of the actual ester form being used *in vivo* (i.e., *p*-coumarate, acetate, or some other ester).

The petunia PhIGS1 kinetic measurements were ultimately corrected by the same researchers[354] upon our prompting (see **Table 1**). The corrected data, without explanation/clarification and embedded within other reports, were now somewhat closer to the creosote bush CES values (i.e., k_{cat}/K_m of 5700 vs 22 000 $mol^{-1}\,l\,s^{-1}$, but still being of lower efficacy by a factor of ~4). This k_{cat}/K_m of 5700 $mol^{-1}\,l\,s^{-1}$ was, however, in stark contrast to the earlier petunia report of ~136 $mol^{-1}\,l\,s^{-1}$. Surprisingly, the corrected petunia data had a K_m of 0.23 mmol l^{-1} (down by a factor of 7) and a V_{max} increased by a factor of ~5. Additionally, the K_m for the basil CES was also corrected to 0.57 mmol l^{-1} (also being reduced by nearly an order of magnitude). Kinetic data for other compounds, such as *p*-coumaryl acetate (**234**) and *p*-coumaryl coumarate (**235**), were apparently not measured, even though the basil CES (ObEGS1), in our hands, was apparently able to utilize both to form chavicol (**20**) with an efficiency somewhat comparable to that toward coniferyl acetate (**236**) (D. G. Vassão, unpublished data). These data thus contradict later claims[355] that the basil CES was of 'limited ability' in terms of being able to use *p*-coumaryl acetate (**234**) as a potential substrate, since no comparable substrate effects were observed in our work.

1.23.9.2.1(iv) Other species Other CES/AIS homologs able to carry out monolignol ester reductions *in vitro* to give either allylphenols or propenylphenols have also been preliminarily characterized in *Clarkia breweri* (CbIGS1, CbEGS1, CbEGS2),[354] petunia (PhEGS1),[354] and anise (*P. anisum*, PaAIS1).[356] Kinetic data for the *Clarkia* CES homologs (the so-called CbEGS1 and CbEGS2) and a putative petunia CES were again lower than those of the creosote bush CES when using coniferyl acetate (**236**) as substrate, that is, with k_{cat}/K_m values of factors from ~5 to more than an order of magnitude lower (**Table 1**). Our nomenclature of CES and AIS[4] has, however, finally begun to be adopted, that is, for the presumed *Pimpinella* AIS (PaAIS1).[356] This protein, though, also has a slow turnover relative to the creosote bush CES (**Table 1**).

1.23.9.2.2 Chemotaxonomy, kinetic properties, and homology comparisons of CES/AIS with PCBER, PLR, IFR (-like) annotations in the plant kingdom: caveats on incomplete analyses

The recent explosion in genomics has often facilitated the correct annotation of gene function, as well as providing some additional insights into the existence of putative homologs (e.g., PLR-like, PCBER-like, etc.), based on sequence comparisons and predictions.

On the contrary, as repeatedly emphasized by ourselves, such database annotations and phylogenetic classifications must be treated with great care and caution as such analyses still represent the application of inexact science. For instance, 17 genes were previously annotated as cinnamyl alcohol dehydrogenase (CAD) and/or CAD-like in The *A. thaliana* Information Resource (TAIR) database. However, a detailed *in silico* analysis established that eight of these had very low homology to *bona fide* CADs, and that they also lacked the Zn catalytic center and Zn-binding signatures of CADs. Of the remaining nine NADPH-dependent proteins,

only six were catalytically competent to reduce the cinnamyl aldehydes **261**, **262**, **264**–**266** to the corresponding alcohols. Of those, only AtCAD4 and AtCAD5 were catalytically the most active,[357] with both having this physiological function *in vivo*.[358] Another example is the Klee *et al.*[359] reinterpretation of the presumed function of a CAD homolog in tobacco (CAD1),[360] which was apparently instead a gene involved in phenylethanol (**296**) biosynthesis, as shown using tomato, etc. In short, many researchers are currently relying too heavily on computationally derived phylogenetic analyses and on homology comparisons, both of which can often be limited in scope and/or give misleading indications.

By contrast, the following are required to establish function: demonstration of comparable *in vitro* enzymatic efficacies relative to the *bona fide* enzyme in question; colocalization of biochemical pathway enzyme and substrates/products *in planta*; demonstration (e.g., by overexpression and/or RNAi inhibition/gene expression suppression) that the pathway in question can be enhanced or suppressed (perhaps involving more than one member of a multigene family). Unfortunately, this is not often done, leading to significant levels of confusion and inaccuracies in the literature, such as in the aforementioned reports.[354,356]

Some additional examples will suffice for the need for circumspection, these being drawn from either CES/AIS and/or PCBER, PLR, and IFR (homolog) comparisons.

1.23.9.2.2(i) **Piper regnellii** This species, as aforementioned, produces both allylphenols and 8–8′-linked phenylpropene-type dimeric lignans, that is, myristicin (**23**), apiol (**36**), and dillapiol (**38**), as well as (+)-conocarpan (**16a**), (−)-*epi*-conocarpan (**188b**), and eupomatenoid-6 (**191**).[218] The *P. regnellii* CES (PrCES1 and PrCES2) amino acid sequences have ∼82% similarity and ∼71% identity to that of the creosote bush CES (LtCES1), and both currently fall within the same clade (**Figure 55**), as do several PCBER and PCBER-like genes from other plant species. This additionally puts into question the true physiological role(s) of PCBER homologs in some of these organisms, as we have repeatedly emphasized.[8,265] However, *P. regnellii* does not contain, to our knowledge, any 8–3′-/8–5′-linked lignans that have undergone 7–*O*–4′ interunit linkage reduction. *P. regnellii* CES1 and CES2 are thus unlikely to have both CES and PCBER functions *in vivo*, and presumably only serve as a CES. This again underscores the limitations of only using phylogenetic analyses to attempt to gauge potential biochemical function(s).

1.23.9.2.2(ii) ***Creosote bush*** This species, as also aforementioned, essentially exclusively accumulates 8–8′-linked lignans, but apparently not others with, for example, 8–5′ linkages, and so on. Interestingly, both CES (LtCES1 and LtCES2) and AIS (LtAIS1) genes are present in this plant, whose corresponding proteins are of ∼63/43% similarity/identity to each other. Their presence, therefore, also provides a biochemical explanation for both allyl- and propenylphenol formation in this species. Additionally, even though both CES genes cluster relatively close to various PCBER/PCBER homologs (**Figure 55**), only marginal (if any) PCBER activity could be detected for LtCES1 under the assay conditions tested.

Interestingly, the creosote bush AIS (LtAIS1) amino acid sequence also has ∼83/68% and ∼78/61% similarity/identity to those of petunia IGS (PhIGS1) and anise AIS (PaAIS1), respectively, and all group together separately as a subclade (**Figure 55**). Segregation as such may, however, not reflect independent evolution, as the creosote bush CES and AIS also fall into different subclades. Many more *bona fide* plant AIS and CES are thus needed to consider evolutionary processes than currently available.

1.23.9.2.2(iii) ***Basil*** This organism produces chavicol (**20**) and eugenol (**4**) in its glandular trichomes. The gene encoding the basil CES (ObEGS1) currently falls into a subclade with the petunia AIS (PhlGS1), as well as the creosote bush AIS (LtAIS1), and the *Clarkia* CES/AIS (CbEGS1 and CblGS1) genes. It segregates, however, from the creosote bush CES (LtCES1 and LtCES2) (**Figure 55**). More information is thus also required before any meaningful explanation can be made with regard to biochemical pathway evolution, that is, whether the evolutionary processes leading to CES activity in different plants were of either convergent or divergent nature.

1.23.9.2.2(iv) **Pinus taeda** Interestingly, according to the phylogenetic depiction in **Figure 55**, the PtPCBER falls into a separate clade from CES and/or AIS gene families. As indicated earlier, however, PtPCBER was isolated from a cell line harboring 8–5′-linked lignans, such as **14**, **103**, and **116**.[147] Although cognizant of its low kinetic parameters, PtPCBER is, nevertheless, expressed in the vascular cambium/ray

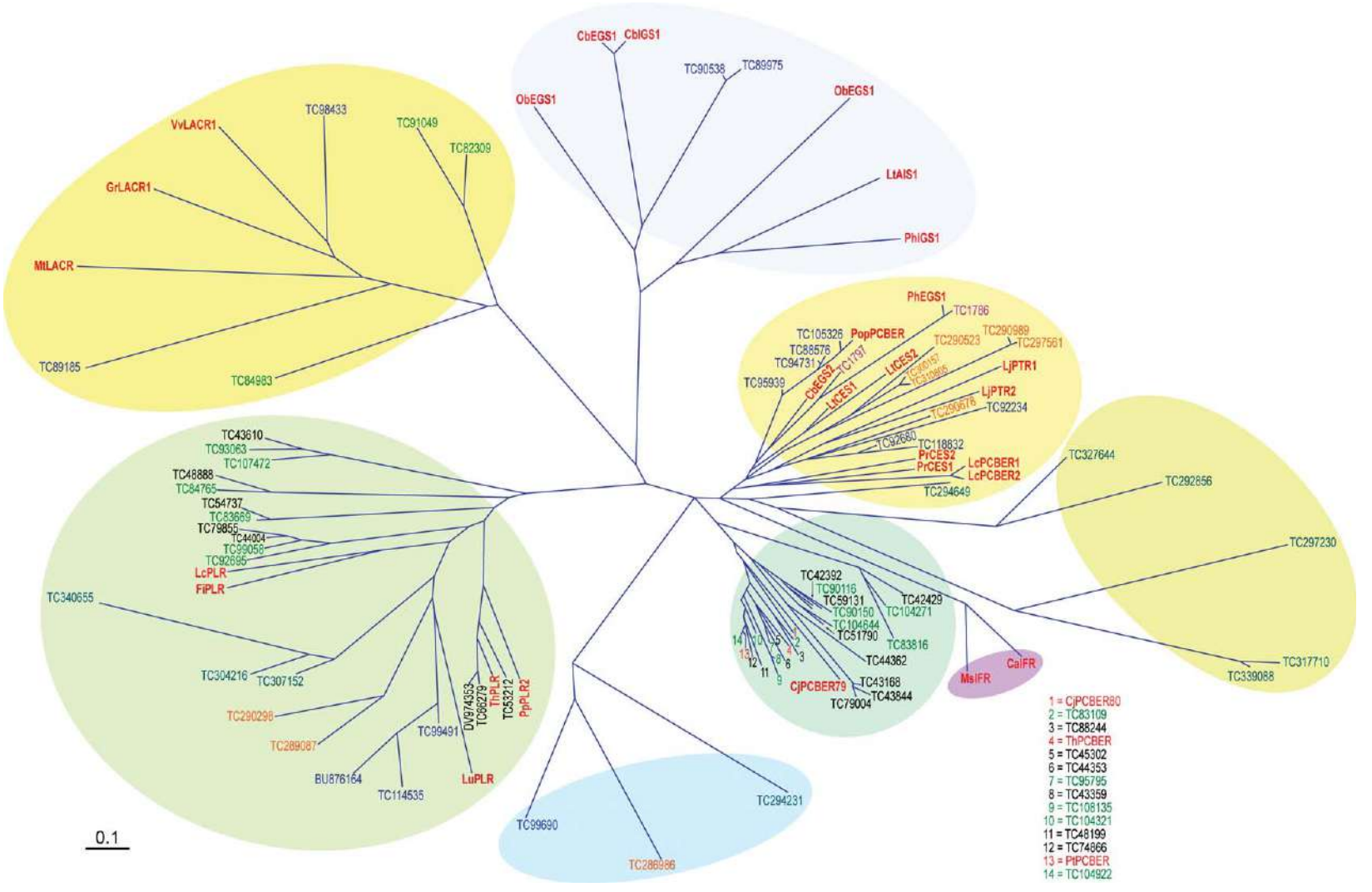


Figure 55 Current phylogenetic dendrogram of tentative consensus sequences of allyl-/prophenylphenol synthases, CES/AIS (*Arabidopsis* (orange), petunia (fushia), rice (light blue), poplar (dark blue), pine (green) and spruce (black)) and *Petunia hybrida* AIS (PhIGS1, DQ372813) and CES (PhEGS1, EF467241), *Ocimum basilicum* CES (ObEGS1, DQ372812), *Clarkia breweri* AIS (CbIGS1, EF467238), CES (CbEGS1 (EF467239) and CbEGS2 (EF467240)), *Larrea tridentata* CES (LtCES1 and LtCES2) and AIS (LtAIS1), and *Piper regnellii* CES (PrCES1 and PrCES2), as well as *Cicer arietinum* isoflavone reductase IFR (CaIFR, Q00016), *Medicago sativa* IFR (MsIFR, CAA1106), *Lotus japonicus* pterocarpan reductase (LjPTR1 and LjPTR2, AB265589 and AB265590), *Gossypium raimondii* leucoantocyanidine reductase (GrLACR, CAI56324), *Medicago truncatula* LACR (MtCAI56327), *Vitis vinifera* LACR (VtCAI26309), *Pinus taeda* PCBER (PtPCBER, AAC32591), *Populus trichocarpa* PCBER (PopPCBER, CAA06706), *Tsuga heterophylla* PCBER (ThPCBER, AAF64177), *Linum strictum* PCBERs (LsPCBER1 (ACA60729) and PCBER2 (ACA60730)) and *Forsythia intermedia* PLR (FiPLR, AAC49608), *Linum usitatissimum* PLR (LuPLR, CAH60858), *Thuja plicata* PLR (TpPLR2, AF242504) and *T. heterophylla* PLR (ThPLR, AAF64184) (red). Amino acid sequences were aligned using ClustalW. To reconstruct phylogenetic tree, maximum likelihood was carried out using Phylip 3.68.

parenchyma cells,[291] this also being consistent with a lignan-related metabolic function. It needs to be established though as to whether this protein is bifunctional, that is, having CES and/or AIS properties *in vitro* and *in vivo*, as well as the PCBER properties already documented. As noted earlier, members of the Pinaceae can also contain very minor amounts of allyl-/propenylphenols in their essential oils, so this (dual) possibility must be considered.

1.23.9.2.2(v) **Cryptomeria japonica** This organism contains at least two PCBER genes of ~85 and 72–80% similarity and identity to that of the PtPCBER. *Cryptomeria japonica* is known to biosynthesize various 8–5′-linked lignans, including one with a reduced 7–*O*–4′ interunit linkage, the tetra-acylated derivative **118**.[160] There are, therefore, PCBER-like enzymatic transformations also occurring in this species. On the contrary, there is currently no chemotaxonomic evidence for the presence of allyl-/propenylphenols in *C. japonica*. It can be considered that the formation of tetra-acetylated lignans, such as **118**, may also involve intermediacy of monolignol esters (e.g., **236**). If correct, then the occurrence of the tetra-acetylated derivative is indicative of only PCBER-like activities.

1.23.9.2.2(vi) ***Petunia CES*** It was quite surprising that a petunia CES (PhEGS1) was considered for its potential/efficacy as a PCBER,[354] particularly because of the following: (1) 8–5′-linked lignans with 7–*O*–4′ interunit linkage-reduced metabolites have never been reported chemotaxonomically from this species; and (2) the gene family is not part of the clades containing PCBER and/or PCBER-like genes. Nevertheless, although the petunia CES (PhEGS1) had marginal PCBER activity, it is well known that many enzyme classes display varying levels of enzymatic activities with a variety of substrates that are of no physiological relevance. Thus, the significance of comparing the petunia CES with PtPCBER was questionable, particularly since petunia apparently does not biosynthesize both classes of metabolites.

1.23.9.2.2(vii) **Clarkia *CES*** The CbEGS2 protein was also considered as a potential PCBER,[354] which in this case provisionally falls into a CES and PCBER-like subclade (**Figure 55**). Its CES activity was, however, more than an order of magnitude lower than that of the creosote bush CES (LtCES1), that is, indicative of fairly low catalytic activity overall. Again, though, the relevance of testing this protein for PCBER activity was questionable, as, to our knowledge, *Clarkia* flowers are not known to produce 7–*O*–4′-reduced 8–5′-linked lignans.

1.23.9.2.3 CES (AIS) structural and mechanistic studies: comparison to PLRs, PCBERs, and IFRs

The closest homologs to the allyl-/propenylphenol synthases are the aforementioned, comprehensively studied, PLRs, [263,264,266,269] PCBERs,[265,266] and IFRs.[361,362] All, as previously indicated, are members of the short-chain dehydrogenase/reductase (SDR) family. Of these, the PLRs and PCBERs were mainly discovered and characterized during the 1990s. Accordingly, determination of PLR and PCBER X-ray crystal structures, as well as that of the modeled IFR,[266] provided incisive and key insights into CES/AIS catalytic function.[4]

In this regard, PLRs, PCBERs, IFRs, and CES/AIS all catalyze the NADPH-dependent hydride transfer to carbons derived from a phenylpropanoid side-chain within their respective phenolic substrates. As for this enzyme class, the creosote bush CES (LtCES1) has a nucleotide-binding Rossman motif (residues 11–17 (GxxGxxG)), as well as the corresponding conserved catalytic Lys residue (Lys133 in LtCES1). The latter residue was mutagenized to afford the recombinant K133A mutant, which lacked catalytic activity, thereby providing further support that Lys133 functioned as a general base during catalysis.[4] Based on our previous studies, Tyr15 and Ile16 were also predicted to interact with the NADPH pyrophosphate group, whereas Phe155 was envisaged to be stacked against the NADPH nicotinamide group.

Taken together, these data further supported the involvement of a quinone methide as the enzyme-bound intermediate in the CES catalytic mechanism[4] (**Figure 31(e)**), as had already been proposed for PLR, PCBER, and PTR (**Figures 31(a), 31(b), and 31(d)**). In the LtCES1 catalytic mechanism, the Lys133 residue can be considered to abstract the phenolic proton from the substrate, thereby facilitating the formation of the putative quinone methide intermediate, with loss of the corresponding ester functionality as a leaving group. Addition of the incoming NAD(P)H hydride to C7 of the intermediate for CES, with concomitant rearomatization of the phenolic ring, affords the corresponding allylphenol

products. Analogously, hydride addition to C9 of the putative quinone methide intermediate in AIS results in the formation of the conjugated propenylphenol product. Note, however, that in LtCES1, various substrates (**234**–**236**) were efficiently converted into chavicol (**20**) and eugenol (**4**), indicative of its substrate versatility.[4]

An X-ray crystal structure was also recently reported for the basil CES (ObEGS1), which confirmed our findings. The basil CES was obtained in its apo-form (**Figure 30(d)**) at ~1.8 Å, as well as complexes with either $NADP^+$, NADPH, or $NADP^+$ in combination with the inhibitor EMDF ((7*S*,8*S*)-ethyl-(7,8-methylene)-dihydroferulate, **297**).[355] However, the putative (but still unproven) 'true' substrate coniferyl acetate (**236**) could not be stably incorporated into the crystals or readily modeled into the active site of the basil CES (ObEGS1). Nevertheless, these data thus provided additional confirmation to the above earlier study[4] of CES catalysis.

The structures so reported, as fully anticipated, closely resembled those previously described by ourselves[266] for PLR, PCBER, and IFR (**Figures 30(a)**–**30(c)**), as well as CES.[4] That is, the basil CES had a Rossman fold motif at residues 14–20 (GXXGXXG), relative to that previously reported for LtCES1 at residues 11–17; the Tyr and Ile residues interacting with the pyrophosphate group of NADPH were at residues 18/19, whereas the earlier study for LtCES1 had these at 15/16. Other confirmatory findings included Phe154 (Phe155 in LtCES1), stacking against the NADPH nicotinamide group, as well as that of Lys132 (Lys133 in LtCES1) for general base catalysis. In further confirmation of our findings,[4] mutation of Lys132 residue to afford K132A resulted in the abolition of enzymatic activity. The X-ray data therefore confirmed our prior results from site-directed mutagenesis and modeling, as well as several other aspects of overall structure prediction.

Furthermore, given that PLRs, PCBERs, CESs, and AISs display considerable homology, and have a similar catalytic mechanism involving transfer of a hydride onto a phenylpropanoid-derived side chain, it should not be too surprising to observe some degree of substrate versatility (degeneracy) among each of the enzymes with regard to potential substrates *in vitro*. This would not be unexpected, as the monolignol esters **234**–**236** are either smaller or of similar size than, for example, PLR substrates, and also share some common structural features (e.g., a free phenolic group). That is, both sets of substrates can potentially bind to the enzyme active site in a similar orientation and presumably undergo similar C7 hydride addition. Following the same rationale, one can therefore expect that similar versatility may be displayed by some of the other PLR, PCBER, and IFR reductases. Conversely, CESs might not necessarily be able to utilize the larger (and less conformationally flexible) lignans as substrates, as their active sites might present greater steric hindrance to efficient binding.

1.23.9.2.4 Allyl-/propenylphenol downstream metabolism

Although allyl-/propenylphenols are present both as monomers, dimers, or both in many plant species, most downstream metabolism for the monomers largely duplicates that of other conversions already discussed for lignan biogenesis. For the monomers, downstream reactions can also precede both accumulation and emission of the typically volatile compounds. In this regard, floral emission to the atmosphere of volatiles in petunia is thought to occur passively, based on their physiological concentrations and boiling points.[363] The most commonly observed metabolic derivatization is *O*-methylation of the phenolic groups, catalyzed by the SAM-dependent *O*-methyltransferases (OMTs): for example, allyl-/propenylphenol OMTs have been isolated from *C. breweri*,[364,365] and more recently from anise[356] and basil.[366] Both enzymes were shown to have strong substrate preferences for either chavicol (**20**) or eugenol (**4**), with these activities being interchanged by mutations in single amino acid residues (F260S in chavicol OMT, S261F in eugenol OMT).[367]

Other ring substitution patterns can be effectuated by the action of specific P-450 enzymes, as previously noted for the sesame lignans[225] (Section 1.23.5.1.1). These include, for the monomers, a preliminary description of the formation of the methylenedioxy bridge in safrole (**22**) from eugenol (**4**) in *Illicium parviflorum* by action of a cytochrome P-450.[368] Other derivatization reactions include phenol conjugation, for example, *O*-acylation.

The most usual phenol ring substitution patterns in both the general phenylpropanoid pathway and the allyl-/propenylphenols are the same (i.e., where carbons 3, 4, and 5 bear H, OH, or OMe substituents, whereas carbons 2 and 6 are unsubstituted). Other more unusual patterns are present in some plant species, as indicated in the chemotaxonomical analysis section. For the latter, these conversions can result through additional (i.e., C2 and/or C6) hydroxylation/*O*-methylation of existing ring substructures deriving from the phenylpropanoid pathway.

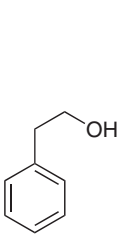
On the contrary, some unusual propenylphenols from *Pimpinella* spp. have the so-called pseudoisoeugenol carbon skeletons, for example, **8**, in which the ring substitution patterns (1-propenyl-2-hydroxy-5-methoxybenzene) deviate from the typical 4-hydroxylated patterns with the three carbon side-chain at C1. Pseudoisoeugenol biosynthesis was initially studied in the late 1980s using isotope labeling/tracer incorporation studies,[369,370] with *P. anisum* cell cultures. Demonstration of an NIH shift of the propenyl side chain during hydroxylation of anethole (**5**) to afford the pseudoisoeugenol skeleton was the later demonstrated.[6,371] However, the enzyme responsible for side-chain migration, as well as the remaining enzymatic steps involved in further derivatization (e.g., epoxidation and acylation), remain still to be described.

1.23.9.3 Monolignol Acyltransferases: Incomplete Characterization and Substrate Degeneracy

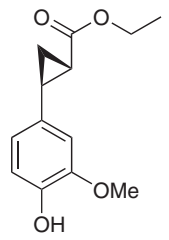
After monolignol esters were identified as substrates for allyl-/propenylphenol synthases, an *in silico* analysis of a petunia petal EST database led to the identification of a cDNA, by the Clark laboratory,[372] which encoded a protein homologous to several known BAHD acyltransferases. These enzymes constitute a class of acyltransferases named for the first four enzymes characterized in this class:[373] benzyl alcohol *O*-acetyltransferase (BEAT), anthocyanin *O*-hydroxycinnamoyl transferase (AHCT), anthranilate *N*-hydroxycinnamoyl/benzoyltransferase (HCBT), and deacetylvindoline 4-*O*-acetyltransferase (DAT). BAHD acyltransferases are known to be involved in the biosyntheses of several aroma and flavor compounds, although amino acid identities among homologous enzymes can vary enormously, with the petunia BAHD homolog being only 22–26% identical to benzyl alcohol/phenylethanol benzoyl transferase (BPBT) from petunia, BEAT from *C. breweri*, and a rose (*Rosa hybrida*) alcohol acyltransferase (AAT1), respectively.

The corresponding putative acyltransferase gene was found to be expressed primarily in petal limbs, with a circadian rhythm profile correlating well with floral volatile emission (i.e., highest amount of transcripts were observed at night). Additionally, RNAi-induced gene silencing led to a decrease in accumulation and emission of isoeugenol (**6**) (along with decreases in levels of 2-phenylethanol (**296**), phenylacetaldehyde (**298**), phenylethyl acetate (**299**), phenylethyl benzoate (**300**), and benzyl acetate (**301**)), as well as an approximately sevenfold increase in the amounts of coniferyl aldehyde (**262**). These effects, however, encompass several different pathways.

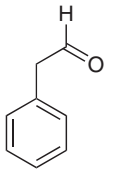
Nevertheless, the full-length cDNA was heterologously expressed in *E. coli*, with the recombinant protein assayed for acyltransferase activity *in vitro*. At pH 7.5, it efficiently acetylated coniferyl alcohol (**91**) as its preferred substrate among the compounds tested (with apparent K_m of 27.5 μmol l^{-1}, k_{cat} of 0.81 s^{-1}, and k_{cat}/K_m of 29 450 mol^{-1} l s^{-1}). Other test substrates such as cinnamyl alcohol (**284**), sinapyl alcohol (**94**), octanol (**302**), and geraniol (**303**) were processed ~2–4 times slower, with activities 30–60 times lower being observed using *p*-coumaryl alcohol (**199**) and 2-phenylethanol (**296**) at the same substrate concentrations. At pH 6.0, the enzyme had an ~2-fold higher K_m for coniferyl alcohol (**91**) (56.5 μmol l^{-1}), but with a ~2.5-fold higher k_{cat} (2.05 s^{-1}), thereby resulting in apparently a similar overall catalytic efficiency (k_{cat}/K_m of 36 280 mol^{-1} l s^{-1}). The enzyme was thus named acetyl-CoA:coniferyl alcohol acetyltransferase (PhCAAT), although it clearly shows extensive substrate versatility. By contrast, a basil ObCAAT enzyme has been preliminary reported to acetylate *p*-coumaryl alcohol (**199**) as its preferred substrate *in vitro*, while also using caffeyl alcohol (**270**) and coniferyl alcohol (**91**) to afford *p*-coumaryl acetate (**234**), caffeyl acetate (**304**), and coniferyl acetate (**236**), respectively.[374]



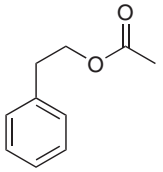
(**296**) 2-Phenylethanol



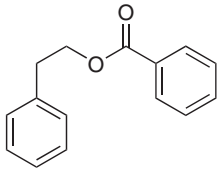
(**297**) EMDF [(7S,8*S*)-ethyl-(7,8-methylene)-dihydroferulate]



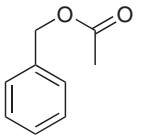
(**298**) Phenylacetaldehyde



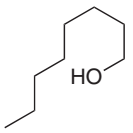
(**299**) Phenylethyl acetate



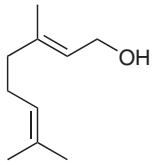
(**300**) Phenylethyl benzoate



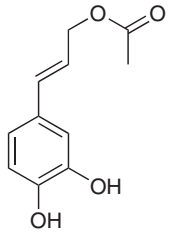
(**301**) Benzyl acetate



(**302**) Octanol



(**303**) Geraniol



(**304**) Caffeyl acetate

Although initial studies[2,3] were not conclusive about the true identity of the monolignol esters involved in regiospecific reduction (i.e., regarding their acyl moieties), subsequent studies[354,356] have essentially only tested coniferyl esters (e.g., **234**) for allyl-/propenylphenol synthesis. However, petunia also accumulates and emits benzoyl esters and phenylethanol derivatives. Using RNAi for the so-called PhCAAT, their levels were also adversely affected as noted above. Surprisingly, benzoyl-CoA and other potential substrates were apparently not tested with the petunia PhCAAT that is now implicated in allyl-/propenylphenol biosynthesis. Nor were other putative substrates, such as *p*-coumaroyl- and feruloyl-CoAs examined either. Alternate acyl donors were apparently also not tested during the preliminary characterization of the basil acyltransferase.

By contrast, an *in vitro* biochemical characterization of an acyltransferase from *L. tridentata* (LtCAAT1, having ~71/57% amino acid similarity/identity to PhCAAT) is currently in progress in our laboratory, and indicates that both acetyl-CoA and benzoyl-CoA can also serve as acyl donors with all the monolignols tested (S.-J. Kim, unpublished results); that is, LtCAAT1 apparently forms, at least *in vitro*, different classes of monolignol esters. Additionally, the acyltransferases appear to efficiently process different monolignols as acyl acceptors, for example, *p*-coumaryl (**199**), caffeyl (**270**), coniferyl (**91**), 5-hydroxyconiferyl (**271**), and sinapyl (**94**) alcohols. Given the observed substrate versatilities being noted, it is thus considered that the depiction of a 'true' *in vivo* substrate (e.g., coniferyl alcohol (**91**) and acetyl-CoA) lacks experimental rigor.

1.23.10 Biological Properties *in Planta* and in Human Usage

Allyl-/propenylphenols, lignans, and to a lesser extent, norlignans are of increasing scientific interest, not only for their utility by humans, but also for their important roles *in planta*. These are now considered in terms of those that have established properties and uses, followed by those whose bioactivities are of a more preliminary

character. In many of the latter cases, these represent promising avenues of research. On the contrary, more often than naught their potentially interesting pharmacological findings are seldom pursued further. In other instances, some of their properties are only effective at very high dose levels/concentrations that would generally preclude utility as medicinals. Such limitations in current natural compound pharmacological studies should thus be kept in mind during the coming subsections.

1.23.10.1 Allyl-/Propenylphenols

1.23.10.1.1 Antimicrobial properties

The most common allylphenol in Nature, eugenol (**4**), has been long used in cloves for its antimicrobial properties, both in food preparations and as a constituent of dental resins. Indeed, for millennia, humanity has used allyl-/propenylphenol-containing spices to protect foods against microbial spoilage. This is because spices were, in addition to salting and drying, the main means of food preservation, especially in warmer climates (i.e., in tropical and equatorial regions), where bacterial growth and contamination are frequently more favored.[375] Furthermore, plants growing in such warmer/more humid climates are typically challenged with numerous and diverse microbial pathogens, and the formation of their chemical defense compounds (e.g., allyl-/propenylphenols, terpenoids, alkaloids) presumably helps withstand such attacks. Other allyl-/propenylphenols (e.g., methylchavicol (**3**) and anethole (**5**)) also have antimicrobial properties, and plants accumulating them (e.g., tarragon, anise, and fennel) are cultivated and used as herbs and spices. Humanity has long treasured the flavors and fragrances of spices, perhaps culturally linking their presence to such beneficial qualities and which result in their uses as food seasonings to the current time. In many societies, however, they have lost much of their relevance as food preservatives *per se*.

Centuries ago, spices were brought to Europe, where they gained large economic importance and became expensive and desirable trade items. Indeed, eugenol (**4**)-rich cloves, which grow in tropical climates (until recently, almost exclusively in the so-called 'Spice Islands' in Indonesia), subsequently became important for their antimicrobial and analgesic properties. Cloves were one of the main spices (along with cinnamon (*Cinnamomum verum*, syn. *C. zeylanicum*) and black pepper (*P. nigrum*)) imported by Europeans since Roman times. The Aksumite Empire (based in nowadays Ethiopia) controlled the main trading routes between the Mediterranean and India, through Egypt and the Red and Arabian Seas, until its decline with the rise of Islam during the seventh century, when these trade routes were effectively blocked. Eventually, trade was reestablished and controlled by Arabs until the fifteenth century, when the rise of the Turk Ottoman Empire again interrupted trade by land. This led to the 'age of exploration' and the creation of early maritime spice trade routes by Europeans around the African coast in the fifteenth/sixteenth centuries. This, in turn, resulted in successive periods of maritime domination by different nations, that is, where Portugal established several coastal colonies during the sixteenth century, many of which were later under Dutch, French, and British control during the rise of their own colonial periods. The development of these routes led to the 'discovery' and colonization of the American continent by the Europeans. Nowadays, Indonesia produces again most of the world's clove supply, after a decrease in production in the past decades in Zanzibar and Pemba (Tanzania), where they had been introduced.

Eugenol (**4**) is thought to exert its antimicrobial effects by affecting either glucose uptake/metabolism[376] or cellular membrane shape/integrity.[377,378] Nevertheless, its antimicrobial effects are apparently attained only at relatively high concentrations. Reported minimal inhibitory concentration (MIC) values against bacteria typically are ~1–5 mmol l^{-1} (e.g., against *Staphylococcus mutans*,[379] *E. coli*, and *Staphylococcus aureus*,[380] as well as *Bacillus subtilis* and *Listeria innocua*[381]); these values are more than 1000 times higher than those of the conventional antibacterial nisin (1 and 2 μmol l^{-1}, respectively).

Although such antibacterial effects are observed only at relatively high eugenol (**4**) concentrations, such functions are presumably attained *in planta*. Some volatile oil components accumulate in very high concentrations, for example, basil contains ~10 mg eugenol (**4**) per gram (fresh weight).[352] Accordingly, utilization of such spices during food preparation can potentially lead to sufficiently high concentrations for antibacterial effect, for example, millimolar eugenol (**4**) concentrations can result from the use of a few hundred milligram cloves, as it comprises ~15% of the dry weight.

In addition to the effects on bacteria, eugenol (**4**) reportedly acts as a fungicidal/fungistatic agent, albeit with relatively high MIC values ranging between 1 and 5 mmol l^{-1} against *Trametes versicolor* (white rot fungus),

Coniophora puteana (brown rot fungus),[382] *Drechslera sorokiniana* (a root rot fungus), *Colletotrichum graminicola* (antrachinose pathogen), *Fusarium solani* (a plant and human fungal pathogen), and *Macrophomina phaseolina* (charcoal rot fungus).[383] It also affects the growth of the yeast *Saccharomyces cerevisiae* with an MIC of 1.8 mmol l^{-1}, where it was demonstrated to cause membrane damage/leakage.[384] Interestingly, membrane damage was of enough utility to allow for eugenol (**4**) use in cell lysis during extraction of genomic DNA from *S. cerevisiae* and *Candida albicans*, instead of SDS/zymolyase. Additionally, *in vivo* studies using oral *C. albicans* infection in immunosuppressed rats reported an MIC of 12 mmol l^{-1} for eugenol (**4**), whereas that for the reference treatment nystatin was 5.8 μmol l^{-1}, or ~2000 times lower.[385] Again, as discussed above for the antibacterial effects of eugenol (**4**), such dosage levels/concentrations are too high for consideration as a potential drug. Nevertheless, beneficial effects may be achieved following food preparation and ingestion.

1.23.10.1.2 Anesthetic properties

In addition to being a useful antimicrobial, eugenol (**4**) is a topical anesthetic, and combination of these properties led to its use in dental polymers as pulp-capping agents.[386–389] It has also been reported to be an efficient anesthetic in fish (as well as rats), facilitating capture, transport/handling, and also surgical procedures, although it becomes lethal at higher doses.[388,390,391] In rainbow trout, eugenol (**4**) concentrations of 50 mg l^{-1} result in short-term anesthesia (<5 min), with 'some deaths' observed at 100 mg ml^{-1}.[392] Cloves are also added to clove ('kretek') cigarettes, originally as a means of clove oil delivery directly to the lungs, to help combat chest pains.[393] The anesthetic activities of eugenol (**4**) are thought to derive from its effects on Na^+ and Ca^{2+} ion channels, inhibiting voltage-gated currents/ion potentials,[394–396] although its effect on GABA receptors may also be involved.[392]

1.23.10.1.3 Other reported activities

Other reports of eugenol (**4**) biological/medicinal properties in the scientific literature are generally more preliminary, and in need of further exploration from a potential medicinal perspective. For example, it is preliminarily reported as having antioxidant, antitumor, antiviral, antileishmanial, anthelmintic, and hypotensive properties in humans/animals.[86,397–399] It has also been shown to modulate ion channel activities as described above[400–403] and inhibit a number of enzymes, including monoamine oxidase,[404] 5-lipoxygenase,[405] and inducible nitric oxide synthase,[406] with reported reduction of NF-κB activity.[407] A number of enzymes are also apparently activated by eugenol (**4**), these including p38 MAP kinases[408] and UDP-glucuronyl transferases.[409] Additionally, micromolar concentrations of eugenol (**4**) have been reported to induce apoptosis in HL-60 leukemia cells (40 μmol l^{-1})[410] and also in mast cells, whereas at higher concentration (700 μmol l^{-1}) it leads to the translocation of phospho-Ser15-p53 into mitochondria and its interaction with Bcl-2 and Bcl-xL.[411] Eugenol (**4**) (at 0.5–2.5 μmol l^{-1} concentration) also inhibits melanoma growth by inhibition of E2F1 transcription.[412]

Other reported allyl-/propenylphenol biological activities include the anti-inflammatory action of anethole (**5**) through inhibition of tumor necrosis factor (TNF)-mediated signaling,[413] for example, via NF-κB activation, the same mechanism proposed for the anti-inflammatory activity of isoeugenol (**6**).[414] Safrole (**22**) and isosafrole (**24**) reportedly protect liver tissue from CCl_4 carcinogenicity by inhibition of its CYP2E1-dependent activation.[415] On the contrary, hydroxychavicol (**50**) potentializes the deleterious effects of benzopyrene in cigarette smoke leading to oral squamous cell carcinoma,[416] and asaricin (**32**) can kill moth larvae at submilligram quantities,[417] the latter being potentially useful as a pesticide and fungicide.[418,419]

1.23.10.1.4 *Effects* in planta

Besides the antimicrobial effects described above, allyl-/propenylphenols have other reported roles *in planta*. Methyleugenol (**42**) is present in a number of floral scents, and is the major attractive component of *Bactrocera* fruit fly pollinators in the orchid *B. cheiri*. It is present in amounts of ~400 μg per flower, about 10 times more than the other allyl-/propenylphenols present.[80] Indeed, methyleugenol (**42**) is commercially used in insect lures and traps, thus reducing the use of conventional insecticides, for example, against Oriental fruit flies (*Bactrocera dorsalis*) in Hawaii.[420–424]

1.23.10.1.5 Mutagenicity

Allyl-/propenylphenols have also long been studied for possible mutagenic potential, as they are typical components of several prepared foods and some can be carcinogenic in test animals at very high doses.

FEMA acceptable/safe daily intakes for methylchavicol (**3**) and methyleugenol (**42**) have been estimated at 1–10 mg kg^{-1} body weight per day, which is 100–1000 times the average human dietary consumption.[425] In low doses, such as those typically found within normal diets, they are metabolized *in vivo* to form harmless products through, for example, demethylation, glucuronidation, and epoxidation/epoxide hydrolysis.[426] In laboratory experiments utilizing much larger doses (100–2000 mg kg^{-1} day^{-1}), allyl-/propenylphenols were also metabolized through P-450-catalyzed C7 side-chain oxidation, followed by the action of sulfotransferases.[427–431] The sulfooxyalkenyl phenol derivatives thus formed were found to be less stable in aqueous media and generated, upon sulfate loss, carbocations that reacted with proteins and DNA to afford covalent adducts. This, however, perhaps accounts for their mutagenic/tumorigenic potential at very high doses.[431–433] Another possibility for their reported toxicity at very high doses may be the reaction of their quinone methide derivatives with basic amino acids (i.e., Lys and Arg).[434] Indeed, such protein adducts have been detected in rat hepatocytes,[435] and more stable carbonium ions/quinone methides have been found to be more genotoxic.[436] Nevertheless, dosages utilized in these studies are abnormally high and orders of magnitude higher than those typically observed from dietary intake.

1.23.10.1.6 Potential future uses as commodity chemicals/biofuels

Allyl-/propenylphenols are now also being considered as potentially useful petrochemical substitutes.[5] Aromatic hydrocarbons (e.g., ethylbenzene) are already present in automotive and aviation fuels, and contribute favorably to their combustion characteristics. Most allyl-/propenylphenols are liquid at room temperature, and release relatively high energy upon combustion (i.e., they have relatively high carbon to oxygen ratios, with higher energy densities especially when compared to ethanol). Another important potential niche for these compounds as petrochemical substitutes is in polymer industries. More than 10% of the current petroleum uses are for nonenergy purposes, including synthesis of polymers and polymer intermediates such as styrene; these facets of the current petroleum crisis are not addressed by any of the main proposed solutions, such as cellulosic ethanol and biodiesel. Allylphenols can be readily converted into the corresponding propenylphenols through alkaline isomerization, and the resulting propenylphenols are structurally substituted/functionalized styrenes. These can undergo similar polymerization reactions to afford polymer chains of relatively high molecular weight, depending on reaction conditions, or may be further refined into other useful products. Therefore, such properties create the exciting possibility for their consideration as (bio)fuels/fuel additives and/or polymer intermediates, if technologies for production at the commodity level become available and economically viable. Additionally, the remaining biomass, if partly reduced in lignin contents, without adversely affecting physiological properties *in planta*, could facilitate cellulose fermentation for ethanol production and also wood processing in pulp/paper industries.

1.23.10.2 Lignans

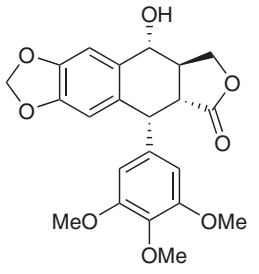
Several lignans have either achieved continued success as pharmaceuticals and/or are showing considerable promise in cancer treatment. Others have been implicated in beneficial effects for human health either in various foodstuffs or in traditional medicine; for example, *S. chinensis*, a plant used in Traditional Chinese Medicine (TCM) for its liver-protecting properties, produces the lignin, gomisin A (**157**),[437–439] whereas secoisolariciresinol diglucoside (SDG, **217**) present in flaxseed (*L. usitatissimum*) has been proposed to prevent onset of various cancers[284] and to modulate fat metabolism,[440] and NDGA (**143**) is thought to be the main bioactive component of *Larrea* species used in traditional Native American medicine.[245] On the contrary, there are many recent but preliminary bioactivity reports of phytochemical studies of lignans of purported medicinal value. In general, these have either not been pursued in more detailed studies, or whose 'efficacious' activities are only achieved at very high concentrations (dosages). These properties are discussed in more detail in the following subsections.

1.23.10.2.1 Lignans in cancer chemotherapy and cancer prevention

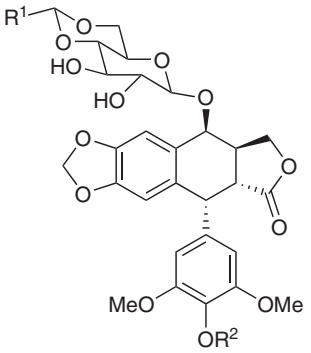
1.23.10.2.1(i) Podophyllotoxin and derivatives Perhaps the most important use of a lignan for medicinal purposes is that of the (−)-podophyllotoxin (**1b**) derivatives in chemotherapy. In this regard, (−)-podophyllotoxin (**1b**) was the first use of a lignan as a lead compound in the semisynthesis of antitumoral agents. However, it proved too toxic for successful use on its own as a therapeutic agent. Derivatives such as etoposide (**305**) and teniposide

(**306**) are now extensively used instead in treatment of various cancers (e.g., testicular, ovarian, brain, breast, and small- and large-cell lung cancers, and lymphomas),[441,442] usually in combination with other antitumoral drugs. By itself, (−)-podophyllotoxin (**1b**) inhibits microtubule assembly, thus leading to cell cycle arrest, whereas its derivatives such as etoposide (**305**) and teniposide (**306**) act on DNA topoisomerase II.[442] This enhanced efficacy, among other structural differences, has been correlated with the presence of bulkier *epi*-C-4 equatorial substituents (e.g., glycosides).[443,444] Such inhibitors stabilize a covalent complex between transient double-stranded DNA breaks and DNA topoisomerase II (which occur normally during DNA recombination and replication), inhibiting DNA strand religation and thus leading to lasting DNA damage.[445]

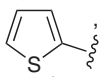
Development of (−)-podophyllotoxin (**1b**) analogs, such as **305** and **306**, still required either large solution volumes or addition of solubilizers, such as polysorbate 80 and polyethylene glycol, for administration of larger doses thereby limiting therapeutic use. Through extensive derivatization studies, however, the etoposide (**305**) C-4′ phosphate derivative etopophos (**307**) was later developed as a more water-soluble prodrug, as it is converted into etoposide (**305**) *in vivo* by action of endogenous phosphatases. As a result, etopophos (**307**) is now a preferred drug for clinical administration, having been approved for intravenous use by the FDA in 1996.[446] Other promising analogs include the epimeric picropodophyllin (**308**), an inhibitor of insulin-like growth factor-1 receptor that inhibits vascular endothelial growth factor (VEGF) secretion and neovascularization, as well as uveal melanoma growth.[447,448] Another derivative of interest is the dual topoisomerase inhibitor tafluposide (F 11782, **309**),[449–451] which is currently under phase I clinical safety studies as a chemotherapeutic agent.[452] Likewise, other semisynthetic analogs have been developed with lower cytotoxicity ID_{50} values, for example, the 4′-*O*-demethyl diamine derivative **310**, with ID_{50} of 0.027 μmol l^{-1} (compared to 0.2 μmol l^{-1} for etoposide (**305**)).[444]



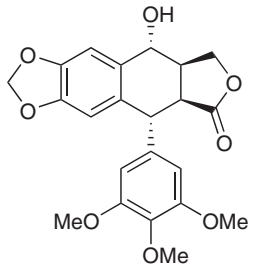
(**1b**) (–)-Podophyllotoxin



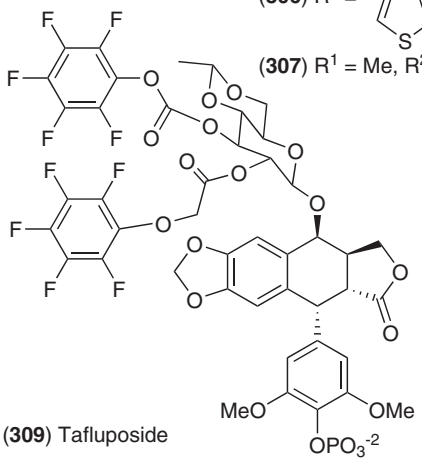
(**305**) R^1 = Me, R^2 = H, Etoposide

(**306**) R^1 =  , R^2 = H, Teniposide

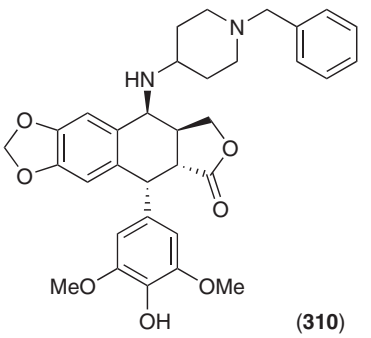
(**307**) R^1 = Me, R^2 = PO_3^{-2}, Etopophos



(**308**) Picropodophyllin



(**309**) Tafluposide



(**310**)

As the therapeutic uses of these compounds continue to grow, new sources have been sought to account for the increased demand; US etoposide (**305**) sales tripled in 1995 and have risen more than 10% annually in the

following years.[441] (−)-Podophyllotoxin (**1b**) is normally extracted from podophyllin, the resin of *Podophyllum* spp. (e.g., *P. emodi*, syn. *P. hexandrum*, and *P. peltatum*) rhizome. By comparison, its content in *P. peltatum* is only 0.25% of dry weight, whereas *P. emodi* accumulates up to 4.3% of dry weight. However, with glucosidase action prior to extraction, (−)-podophyllotoxin (**1b**) yields can be increased up to 5.2% of dry weight of *P. peltatum* leaves and rhizomes.[453]

Synthetically prepared alternatives for their production are disfavored due to typically low yields, resulting from the large number of steps and the challenges in obtaining the correct enantiomeric forms.[441] *In vitro* plant cultures are also currently unable to generate (−)-podophyllotoxin (**1b**) at a low enough cost, with among the highest amounts obtained being 130 mg l^{-1} in 10 days.[454] According to Verpoorte *et al.*[455] productions rates of 300 mg l^{-1} within 14 days would still result in prices 10 times higher than those obtained through traditional production methods, and even elicitation or precursor administration methods may not lower prices enough for biotechnological production of (−)-podophyllotoxin (**1b**) or derivatives.[454] Therefore, *in vitro* production protocols must be considerably improved before traditional extraction methods are economically surpassed. These efforts perhaps may be facilitated by the ongoing elucidation of (−)-podophyllotoxin (**1b**) biosynthetic pathway steps.

1.23.10.2.1(ii) Nordihydroguaiaretic acid and derivatives Another lignan whose derivatives are gaining increased interest as chemotherapeutic agents is NDGA (**143**). Prior to this, it was used as an antioxidant and as an additive in rubber production. The actual bioactive entity in chemotherapy is not completely established, with NDGA (**143**) perhaps serving in the capacity of a prodrug. This is being considered since NDGA (**143**) was recently shown to auto-oxidize in aqueous media (with a reaction half-life of ~3 h at pH 7.4)[456] to form a schizandrin-like dibenzocyclooctadiene lignan (**311**); the latter is perhaps the actual bioactive agent responsible for the biological activities discussed below. NDGA (**143**) has also been shown to undergo auto-oxidation to form quinones that are reactive toward thiols, thus leading to GSH depletion by adduct formation, this being another of its potential mechanisms of action.[457]

In recent years, NDGA (**143**) has become a promising potential pharmaceutical due to the anticancer properties reported for itself and some of its derivatives, and it has also been licensed for treatment of actinic keratosis (Actinex, Chemex Pharmaceuticals, Denver, CO).[245,458] Its reported anticancer properties on varied cancer cell lines may result from a combination of different activities. Examples include arresting cells on G1 phase (at least partially by reactivation of p16 INK4a, a gene that undergoes methylation in some cancers),[459] activation of TRAIL-induced apoptosis,[460] and inhibition of IGF signaling[461–464] as well as of lipoxygenase[462,465] and topoisomerase II alpha[466] activities. Among cancer cell lines that NDGA (**143**) has shown promising activity against are colorectal tumors (IC_{50} 1.9 μmol l^{-1}),[467] breast cancers,[468] adenocarcinomas,[469] and neuroblastomas (blocking growth at 15–30 μmol l^{-1}).[463] M4N (Terameprocol, **312**), the tetra-*O*-methylated derivative of NDGA (**143**), is also active against melanomas with similar IC_{50} values (1–20 μmol l^{-1}),[470] and is particularly suitable for intratumoral injection due to its low aqueous solubility, as it remains concentrated at the injection site.[471] Indeed, M4N (**312**) is currently undergoing phase I/II NIH trials[452] as a promising potential treatment against treatment-refractory solid tumors.[472,473] Another derivative, tetra-*O*-acetyl NDGA (**313**), has also been reported to be effective against adenocarcinomas.[469] As more structural studies are performed with these lignans, as for the podophyllotoxin (**1**) derivatives described above, NDGA (**143**) or derivatives of potentially higher efficacy may find similar extensive pharmacological applications as chemotherapeutic drugs.

1.23.10.2.1(iii) Enterolignans and cancer prevention Other classes of lignans showing some potential as cancer-preventative drugs/nutraceuticals are next discussed, albeit with the caveat that much more detailed studies are still necessary. These include the so-called 'mammalian' or entero-lignans, enterodiol (**316**) and enterolactone (**317**), which are produced by colonic microflora. These can be formed from many plant lignans and their glucosides. Both possess weakly estrogenic and antiestrogenic activities on different cell lines *in vitro*, and have been proposed to protect against the growth of some human cancers.[474,475] Higher levels of these lignans *in vivo* are found after intake of various foods, most prominently flaxseed (*L. usitatissimum*).[476] They have also been implicated in the antitumor activities of wheat bran on colon cancer cells,[477] of rye bran on diabetes,[478] and of flaxseed on atherosclerosis-related processes.[479]

Epidemiological studies suggest a protective effect of enterolignans against lower digestive tract cancers. These compounds were reported to inhibit growth of prostate cancer cell lines *in vitro*,[480–482] and significant inverse associations were observed between serum concentrations of enterolignans (but not of genistein, daidzein or equol) and prostate cancer.[483] Different *in vivo* studies reported protective effects on colorectal adenomas (considered to be cancer precursors)[484] but no effect on colorectal cancers.[485] Similarly, enterolignans were reported to have no correlation with either adenoma formation in mice[486] or lower myocardial infarction risk,[487] and a positive correlation was observed between enterodiol (**316**) plasma levels and risk of premalignant lesions of the cervix.[488] Enterolactone (**317**), on the contrary, was reported to reduce growth and metastasis of hepatomas in rats.[489]

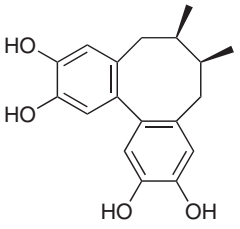
Using *in vitro* experiments with varied cell lines, enterolignans were found to decrease the activities of aromatase and 17β-hydroxysteroid dehydrogenase in MCF-7 breast cancer cells, thereby reducing cell proliferation,[490] whereas other researchers found a proliferation-stimulatory estrogenic effect.[491] Other studies have indicated that enterolactone (**317**) differentially modulates estrogen receptors α and β signaling, which could account for these discrepancies.[492,493]

Although the study of such hormonal effects is inherently difficult, most epidemiological studies correlate higher levels of plasma enterolactone (**317**) with reduced breast cancer risk and/or metastasis,[494–498] with some studies narrowing the beneficial effects to particular woman or cancer genotypes.[495,499] Again, these effects are not universally observed and/or accepted, and it has been reported that a flaxseed diet during gestation or lactation can increase the susceptibility of rat offspring to mammary tumorigenesis.[500] From these results, it is evident that the metabolism, mechanism of action, and potential of these compounds must be better understood before their potential as either cancer-preventative nutraceuticals or medicinal products are fully established.

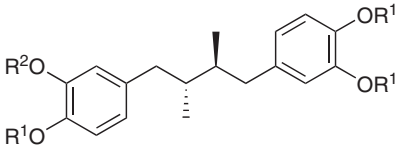
The main *in vivo* precursors of enterodiol (**316**) and enterolactone (**317**) are considered to be various plant lignans, such as matairesinol (**10**), secoisolariciresinol (**110**), as well as sesamin (**11**), pinoresinol (**13**), syringaresinol (**96**), arctigenin (**197**), 7-hydroxymatairesinol (**114**), lariciresinol (**105**), and isolariciresinol (**133**), as well as glucosides thereof, for example, secoisolariciresinol diglucoside (SDG, **217**) from flax.[474,475,501–503] The biosynthetic steps from these precursors is thought to involve deglycosylation (when necessary, e.g., from SDG (**217**)), followed, apparently as a necessary first step, by *O*-demethylation of one of the methoxyl groups of, for example, **110**. Ring dehydroxylation then probably occurs by loss of aromaticity and dehydration.[284,474,475,504,505] Additionally, precursor lignans such as pinoresinol (**13**) can be reduced to afford lariciresinol (**105**) and secoisolariciresinol (**110**) *in vivo*, whereas enterodiol (**316**) can be oxidized to afford enterolactone (**317**). All these reactions have been reported to be performed by strains of phylogenetically diverse colonic bacteria, for example, *Bacteroides*, *Eubacterium*, *Clostridium*, *Peptostreptococcus*, *Enterococcus*, *Eggerthella*, and *Ruminococcus* spp.[506–510] The enterolignans **316**/**317** thus formed are taken up and further metabolized by colon epithelial cells (and also perhaps by liver enzymes), affording the corresponding sulfates and glucuronides.[511] Indeed, it is mainly those metabolites that are later secreted in urine and bile, whereas the free enterolignans are only found in feces and (together with their conjugated derivatives) in plasma.[474,475] Typically, these lignans appear in plasma 8–10 h after intake, and have a mean elimination half-life of ~4–12 h,[512] with considerable inter-individual variations in the levels of lignans observed both in plasma and in secreted fluids, probably due to differences in gut microflora. Interestingly, while most lignans found in plants are optically active (including the (+)-secoisolariciresinol (**110**) moiety of SDG (**217**) in flaxseed),[275] the enterolactone (**317**) found in urine is racemic;[475] the reaction(s) potentially leading to racemization are not yet understood, and these observations may thus suggest a lignin origin as well.

These lignans have been reported to exert varied other activities *in vivo*, including protection against oxidants,[513] lowering diet-induced fat accumulation in mice,[440] lowering of plasma cholesterol and glucose concentrations in high-cholesterol subjects,[514] alleviating lower urinary tract symptoms in subjects with benign prostatic hyperplasia,[515] and also aiding in glycemic control in type-2 diabetic patients.[516]

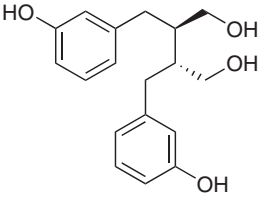
1.23.10.2.1(iv) Other lignans and norlignans with anticancer potential Some other lignans have also been reported to suppress either tumor growth or tumorigenicity of known carcinogens *in vitro*, but these results are still very preliminary. For instance, several recent studies have suggested that honokiol (**318**) from *Magnolia officinalis* is a promising potential chemotherapeutic agent for treatment of several cancers (e.g., colorectal and ovarian), based on antioxidant, proapoptotic, and antiangiogenic activities both *in vivo* and *in vitro* against several cancer cell lines.[517–524] Although clinical studies are still lacking, honokiol (**318**) was considered nearly as effective an antitumoral agent against colorectal carcinoma cells as the common chemotherapeutic drug adriamycin.[517]



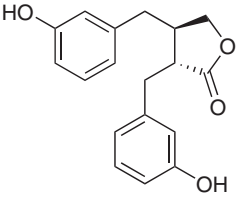
(**311**)



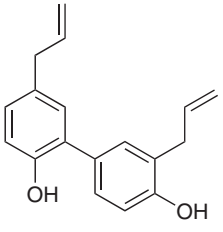
(**143**) $R^1 = R^2$ = H, Nordihydroguaiaretic acid
(**312**) $R^1 = R^2$ = Me, Tetra-*O*-methyl NDGA (M4N)
(**313**) $R^1 = R^2$ = Ac, Tetra-*O*-acetyl NDGA
(**314**) $R^1 = R^2$ = Gly, Tetra-*O*-glycyl NDGA
(**315**) R^1 = H, R^2 = Me, 3-*O*-methyl NDGA



(**316**) Enterodiol



(**317**) Enterolactone



(**318**) Honokiol

The unusual acetylenic norlignan hypoxoside (**182**) from African potato (*Hypoxis hemerocallidea*), along with its aglycone rooperol (**181**), may also have good potential as pharmaceuticals.[525–527] The latter (**181**) has been reported as an antioxidant and a cancer-cell growth inhibitor at concentrations of 2–10 μg ml^{-1}. Its glucoside (**182**) can also serve as a nontoxic prodrug (up to 100 μg ml^{-1}) that is deconjugated *in vivo* into its aglycone form (**181**),[528] but circulates mainly as phase II metabolites (glucuronides and sulfates).[212,525–527,529] Furthermore, *H. hemerocallidea* is possibly the best-known medicinal plant in South Africa, whose hypoxoside (**182**)-rich extracts are used for the treatment of urinary diseases, prostate hypertrophy, and cancer.[530] Although promising, the reported bioactivities for these African potato norlignans are still quite preliminary, and will require numerous additional *in vitro* and *in vivo* toxicity studies and structure–activity analyses before their potential as medicinal compounds can be better understood.

1.23.10.2.2 Antiviral lignans

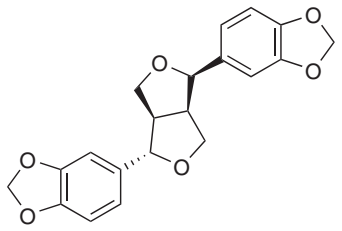
Another set of interesting lignan biological activities either currently used or being considered for medicinal purposes is that of antivirals. In addition to the previously discussed antitumor activities, (−)-podophyllotoxin (**1b**) and its derivatives also inhibit viral replication, for example, against measles, human papilloma HPv, cytomegalovirus, and herpes simplex type I viruses, being used medicinally against genital herpes.[442] NDGA (**143**) and derivatives also have antiviral activities, particularly against HIV and papillomavirus, although with relatively high concentrations being necessary. NDGA (**143**) has an IC_{50} of 25 μmol l^{-1} against HIV, while its tetra-*O*-methylated derivative (M4N, **312**) has an IC_{50} of 11 μmol l^{-1};[531] NDGA (**143**) and M4N (**312**) were also found to inhibit HIV-1 replication by suppressing proviral and HIV Tat-transactivated transcription.[531] The tetra-*O*-glycyl NDGA derivative (**314**) has an IC_{50} values of 12 μmol l^{-1} against HIV[532] and 3–5 μmol l^{-1} against herpes simplex virus-1.[533] Tetra-*O*-acetyl NDGA (**313**) and 3-*O*-methyl NDGA (**315**) are also effective against human papillomavirus and induce apoptosis in cervical tumor cells in *in vitro* assays at low micromolar concentrations,[534,535] whereas M4N (**312**) has shown promising results in Phase I/II trials as a vaginal ointment in treatment of HPV-linked cervical cancers.[536,537]

1.23.10.2.3 Nutraceutical lignans: sesame

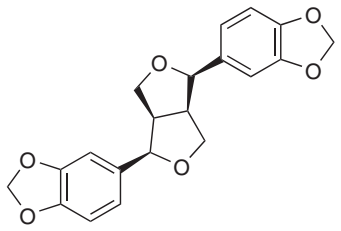
Many lignans have reported antioxidant properties, which in large part can be ascribed to the presence of free phenolic groups within their structures. Nevertheless, sesame represents perhaps one of the best-studied dietary lignan sources, as the lignans in both its seed and oil have beneficial properties for humans.

Sesame seed is consumed either raw or roasted, and its oil is extracted from either raw or roasted seed as well, leading to products with different sensory properties. South Korea has the highest *per capita* consumption (~6–7 g day^{-1}), with the Japanese consuming about 3 g day^{-1}, mainly as sesame oil.[249] Sesame oil is considered one of the best vegetable oils for deep-frying, as it is more resistant to high-temperature oxidation than soybean and canola oils, and also helps increase the durability of fried foodstuffs. Those antioxidant properties are thought to derive mainly from the presence of lignans in sesame and sesame oil, and the antioxidant potential of sesame oil apparently increases after roasting.

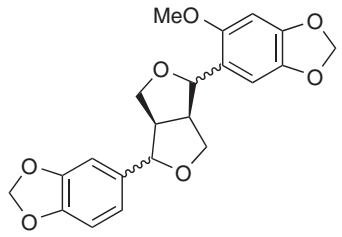
The most abundant sesame phenylpropanoids characterized are (+)-sesamin (**11a**), sesamolin (**210**), sesaminol (**206**), and sesamolinol (**211**) as previously discussed, together with *epi*-sesamin (**319**), diasesamin (**320**), sesangolin (**321**), 2-episesalatin (**322**), simpleoxide aglycone (**323**), and (+)-pinoresinol (**13a**) (from which the majority of sesame lignans are thought to be derived). Mono-, di-, and tri-glucosides of those lignans are also present.[249] Their levels vary among different *Sesamum* species, with sesamin (**11a**) and sesamolin (**210**) contents in sesame oil being typically 0.4% and 0.3%, respectively (with up to 2.4% (+)-sesamin (**11a**) in the oil of *S. radiatum*).[249]



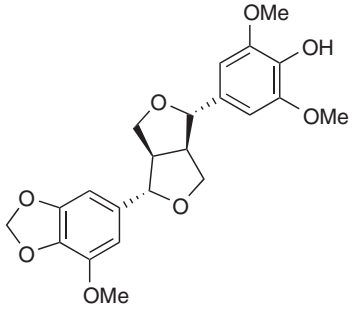
(**319**) Episesamin



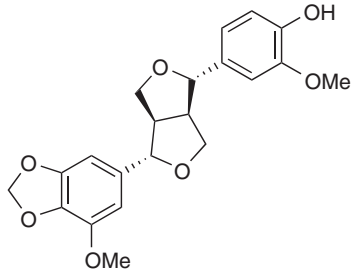
(**320**) Diasesamin



(**321**) Sesangolin



(**322**) 2-Episesalatin



(**323**) Simpleoxide aglycone

Sesaminol (**206**) is thought to be the main antioxidant lignan present in sesame oil, and its available levels are thought to increase (concomitant to sesamolin (**210**) level decrease) upon roasting, oil processing, or in the case of whole sesame seed or sesame meal, through hydrolysis of sesaminol (**206**) glucosides. (+)-Sesamin (**11a**) has no free phenol group and shows weak or no antioxidant effects; however, it is thought to be metabolized *in vivo* to afford the aforementioned enterolignans, for example, **313** and **314**.[503]

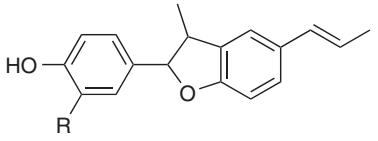
In addition to their antioxidant properties, sesame lignans affect tocopherol/vitamin E metabolism, increasing tocopherol levels in plasma, brain, and liver (two- to four-fold in the latter, where these lignans inhibit tocopherol metabolism).[249,538] Fatty acid metabolism is also affected by sesame lignans in microorganisms and animals, for example, sesamin (**11**) and other sesame lignans inhibit $\Delta 5$ desaturase activity in the fungus *Mortierella alpina* and rats, with IC_{50} values for sesamin (**11**) of 5.6 and 72 μmol l^{-1}, respectively (other fatty acid desaturases were not inhibited to the same extent).[249,539] Sesame lignans were also shown to increase hepatic fatty acid oxidation and decrease hepatic fatty acid synthesis in rats, which leads to decreased levels of serum fatty acids.[540,541] A similar trend was observed regarding cholesterol metabolism, with sesamin (**11**) in combination with α-tocopherol lowering the concentration of serum and liver cholesterol in rats and humans,

thus indicating a preventive role against atherosclerosis.[542] Additionally, sesamin (**11**) was reported to have potential beneficial effects on hypertension[543] and to increase the rate of alcohol metabolism in rats and humans.[249]

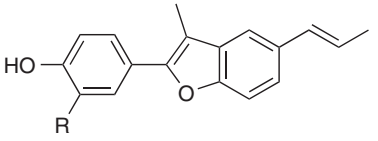
It is thus considered that sesame seed lignans can exert a number of different potentially beneficial effects in humans and other animals, with mechanisms for their actions having been proposed in some cases. Some of these activities are apparently specific to sesame lignans, while others can be potentially ascribed to their downstream metabolic products, for example, enterolignans such as **313** and **314**. Further detailed *in vivo* studies regarding these different effects and the true active principles must still be performed. Nevertheless, sesame seed continues to gain well-deserved attention as a nutraceutical dietary component.

1.23.10.2.4 Antichagasic lignans

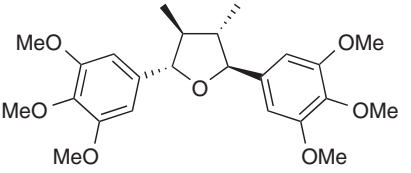
Some lignans have shown somewhat promising *in vitro* activities in cases where currently available treatments are not efficient (e.g., requiring high drug concentrations), thus potentially serving as useful lead compounds for the development of pharmaceutical compounds. For example, eupomatenoids (**191**, **192**), their possible *in vivo* precursors conocarpan (**16**) and **324**, as well as grandisin (**325**), have been isolated from a few plant species including *Conocarpus erectus*,[544] *P. regnellii*,[218] and *Piper solmsianum*,[545] and are active against some tropical parasites and insect larvae.[546] More importantly, though, they have shown activity against epimastigote forms of *Trypanosoma cruzi*, the causative agent of Chagas' disease, with IC_{50} values of 3–9 $\mu g\,ml^{-1}$, and being ~20% more inhibitory to *T. cruzi* growth than the typical medicaments benznidazole (which is less effective during the chronic part of the disease) and gentian violet (which has an IC_{50} value of 28 $\mu g\,ml^{-1}$).[547–549] Eupomatenoid-5 (**192**) also acts on the amastigote (intracellular) forms of *T. cruzi* at similar concentrations. Nevertheless, while these compounds may have some promising potential applications in Chagas' disease treatment, further pharmacological studies are lacking regarding their toxicities or activities of semisynthetic derivatives.



(**16**) R = H, Conocarpan
(**324**) R = OMe



(**191**) R = H, Eupomatenoid-6
(**192**) R = OMe, Eupomatenoid-5



(**325b**) (–)-Grandisin

1.23.10.2.5 Properties in planta

As discussed in detail before,[1] functional assignment of lignans (as well as most other specialized metabolites) *in planta* is very sparse and often circumstantial. Nevertheless, in most described cases, their functions seem to be mainly defensive, for example, antimicrobial, antioxidant, antifeedant, and allelopathic. For instance, in agreement with a putative defense application, the levels of dihydromonolignols (e.g., **291**, **119**) apparently increase upon aphid attack (e.g., *Adelges abietis* attack on *P. glauca*).[161] In other cases, they may serve to attract microbes or insects, sometimes with deleterious effects. One example is olivil (**12**), which was isolated from the resin of olive trees in Europe. Olive trees were later introduced in Japan in 1908, and olivil (**12**) has been recently found to act as a feeding stimulant to a species of weevils in Japan. These weevils, which originally used other Oleaceae plants (e.g., *Ligustrum japonicum* and *Ligustrum obtusifolium*) for their diet, have since changed their host preferences to the olive tree. These so-called 'olive weevils' now constitute the most serious pest of the olive tree in Japan,[550] indicating the delicate balance in Nature that can exist.

Strong antioxidant lignans are present in some of the most valuable plant oils, for example, sesame seed oil, where they are thought to prevent degradation and help stabilize these oils at high cooking temperatures. NDGA (**143**), as mentioned before, is also a very strong antioxidant and is found in large amounts in the creosote bush (*L. tridentata*), perhaps helping its survival in high-temperature desert areas. Of course, such distinctive properties are employed for human benefit, leading to the use of sesame oil in cooking, and of NDGA (**143**) as a stabilizer in polymer and rubber applications.

Some lignans are also reported to possess allelopathic properties. NDGA (**143**) is thought to be the main contributor to the phytotoxic effects of *L. tridentata*, which is often present as the single plant species in large colonized areas; indeed, early uses of this plant included sprinkling over train beds to stop weed growth.[551] Interestingly, on the contrary, dihydroconiferyl alcohol (**119**) was studied during the 1970s as a plant growth regulator, inducing hypocotyl growth in several plant species.[552–556] More recently, both dihydroconiferyl alcohol (**119**) and dihydro-*p*-coumaryl alcohol (**291**) contents were noted to be lower in the extractive contents of *Pinus laricio* after prescribed burnings;[557] dihydroconiferyl alcohol (**119**) is also found in pine weevil (*Hylobius abietis*) feces deposited to close holes bored for egg deposition, where it reportedly functions as an antifeedant to other pine weevils signaling the presence of the egg cavities under the trunk surface.[558]

As another example, the previously mentioned, *G. officinale* (ironwood, palo-santo, *lignum vitae*), a member of the Zygophyllaceae (which also includes the creosote bush, *L. tridentata*), accumulates in its heartwood a resin containing the 8–8′-linked 9–9′-deoxygenated lignan (−)-guaiaretic acid (**9b**). Its massive deposition is thought to contribute to the exceptional durability of this wood, one of the densest and hardest among commercial timbers. Indeed, the presence of large amounts of resin results in water resistance and 'self-lubricating' properties, which have made it ideal for use in ship construction. Other past uses of its wood have included manufacturing of bowling balls, mallets, and bearings.[559,560] Lignans contributing to heartwood durability and texture in other plant species include (−)-plicatic acid (**112**) and its derivatives in *T. plicata*,[156] 7-hydroxymatairesinol (**114**) and (−)-α-conidendrin (**113b**) in *T. heterophylla*,[154] and (nor)lignans such as **116**, **122**–**124** in *C. japonica*,[166,167] as discussed before in Section 1.23.3.2.3.

Acknowledgments

The authors thank the US Department of Energy (DE FG03-97ER20259), the US National Science Foundation (MCB-0417291), the National Institute of General Medical Sciences (5 R01 GM066173-02), the United States Department of Agriculture (Agricultural Plant Biochemistry #2006-03339), the BioEnergy Science Center, the US Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science, McIntire Stennis, and the G. Thomas and Anita Hargrove Center for Plant Genomic Research for generous financial support.

Dedicated to Professor David G. I. Kingston on the occasion of his 65th birthday.

References

1. N. G. Lewis; L. B. Davin, Lignans: Biosynthesis and Function. In *Comprehensive Natural Products Chemistry Vol.1: Polyketides and Other Secondary Metabolites, Including Fatty Acids and Their Derivatives*; Sir D. H. R. Barton, K. Nakanishi, O. Meth-Cohn, Eds.; Elsevier: Oxford, UK, 1999; pp 639–712.
2. D. G. Vassão; D. R. Gang; T. Koeduka; B. Jackson; E. Pichersky; L. B. Davin; N. G. Lewis, *Org. Biomol. Chem.* **2006**, *4*, 2733–2744.
3. T. Koeduka; E. Fridman; D. R. Gang; D. G. Vassão; B. L. Jackson; C. M. Kish; I. Orlova; S. M. Spassova; N. G. Lewis; J. P. Noel; T. J. Baiga; N. Dudareva; E. Pichersky, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10128–10133.
4. D. G. Vassão; S.-J. Kim; J. K. Milhollan; D. Eichinger; L. B. Davin; N. G. Lewis, *Arch. Biochem. Biophys.* **2007**, *465*, 209–218.
5. D. G. Vassão; L. B. Davin; N. G. Lewis, Metabolic Engineering of Plant Allyl/Propenyl Phenol and Lignin Pathways: Future Potential for Biofuels/Bioenergy, Polymer Intermediates, and Specialty Chemicals? In *Advances in Plant Biochemistry and Molecular Biology*; H. J. Bohnert, H. Nguyen, N. G. Lewis, Eds.; Elsevier: Oxford, UK, 2008; Vol. 1, Bioengineering and Molecular Biology of Plant Pathways, pp 385–428.
6. R. Martin; J. Reichling, *Phytochemistry* **1992**, *31*, 511–514.
7. N. G. Lewis; L. B. Davin; S. Sarkanen, The Nature and Function of Lignins. In *Comprehensive Natural Products Chemistry*; Sir D. H. R. Barton, K. Nakanishi, O. Meth-Cohn, Eds.; Elsevier: Oxford, UK, 1999; Vol. 3, pp 617–745.
8. L. B. Davin; M. Jourdes; A. M. Patten; K.-W. Kim; D. G. Vassão; N. G. Lewis, *Nat. Prod. Rep.* **2008**, *25*, 1015–1090.
9. L. B. Davin; A. M. Patten; M. Jourdes; N. G. Lewis, Lignins: A Twenty-First Century Challenge. In *Biomass Recalcitrance. Deconstructing the Plant Cell Wall for Bioenergy*; M. E. Himmel, Ed.; Blackwell Publishing: Oxford, UK, 2008; pp 213–305.
10. O. Doebner; E. Lücker, *Archiv. Pharm.* **1896**, *234*, 590–610.
11. G. Schroeter; L. Lichtenstadt; D. Irineu, *Chem. Ber.* **1918**, *51*, 1587–1613.
12. R. D. Haworth; C. R. Mavin; G. Sheldrick, *J. Chem. Soc.* **1934**, 1423–1429.
13. T. H. Easterfield; J. Bee, *J. Chem. Soc.* **1910**, *97*, 1028–1032.
14. V. Villavecchia; G. Fabris, *Ann. Lab. chim. centr. delle Gabelle* **1896**, *3*, 13–26.
15. J. Böeseken; W. D. Cohen, *Biochem. Z.* **1928**, *201*, 454–463.
16. B. L. Vanzetti, *Gazz. Chim. Ital.* **1929**, *59*, 373–378.

17. B. L. Vanzetti; P. Dreyfuss, *Gazz. Chim. Ital.* **1934**, *64*, 381–399.
18. D. C. Ayres; S. E. Mhasalkar, *J. Chem. Soc.* **1965**, 3586–3589.
19. R. D. Haworth, *Ann. Rep. Progr. Chem.* **1937**, *33*, 266–279.
20. O. R. Gottlieb, *Phytochemistry* **1972**, *11*, 1537–1570.
21. O. R. Gottlieb; M. Yoshida, Lignans. In *Natural Products of Woody Plants – Chemical Extraneous to the Lignocellulosic Cell Wall*; J. Wand, C. H. Kirk, Eds.; Springer Verlag: Berlin, 1989; pp 439–511.
22. G. P. Moss, *Pure Appl. Chem.* **2000**, *72*, 1493–1523.
23. F. S. El-Feraly; S. F. Cheatham; R. L. Breedlove, *J. Nat. Prod.* **1983**, *46*, 493–498.
24. B. Y. Hwang; J.-H. Lee; J. B. Nam; Y.-S. Hong; J. J. Lee, *Phytochemistry* **2003**, *64*, 765–771.
25. A. Ichihara; S. Kanai; Y. Nakamura; S. Sakamura, *Tetrahedron Lett.* **1978**, *19*, 3035–3038.
26. A. Ichihara; Y. Nakamura; H. Kawagishi; S. Sakamura, *Tetrahedron Lett.* **1979**, *20*, 3735–3736.
27. P. Kenrick; P. Crane, *Embryophytes. Land Plants. Version 01 January 1996 (temporary). http://tolweb.org/Embryophytes/20582/1996.01.01 in The Tree of Life Web Project*, http://tolweb.org/, 1996.
28. P. Crane, *Spermatopsida. Seed Plants. Version 01 January 1996 (temporary). http://tolweb.org/Spermatopsida/20622/1996.01.01 in The Tree of Life Web Project*, http://tolweb.org/, 1996.
29. Atta-ur-Rahman; M. I. Choudhary; S. Hayat; A. M. Khan; A. Ahmad; S. Malik, *Phytochemistry* **1999**, *52*, 495–499.
30. N. E. Awad, *Bull. Fac. Pharm. (Cairo Univ.)* **2002**, *40*, 169–174.
31. M. Y. Moussa, *Bull. Fac. Pharm. (Cairo Univ.)* **2003**, *41*, 139–144.
32. N. E. Awad, *Bull. Fac. Pharm. (Cairo Univ.)* **2002**, *40*, 233–243.
33. K. Bouarab; F. Adas; E. Gaquerel; B. Kloareg; J.-P. Salaün; P. Potin, *Plant Physiol.* **2004**, *135*, 1838–1848.
34. S. S. Merchant; S. E. Prochnik; O. Vallon; E. H. Harris; S. J. Karpowicz; G. B. Witman; A. Terry; A. Salamov; L. K. Fritz-Laylin; L. Maréchal-Drouard; W. F. Marshall; L.-H. Qu; D. R. Nelson; A. A. Sanderfoot; M. H. Spalding; V. V. Kapitonov; Q. Ren; P. Ferris; E. Lindquist; H. Shapiro; S. M. Lucas; J. Grimwood; J. Schmutz; P. Cardol; H. Cerutti; G. Chanfreau; C.-L. Chen; V. Cognat; M. T. Croft; R. Dent; S. Dutcher; E. Fernandez; H. Fukuzawa; D. Gonzalez-Ballester; D. Gonzalez-Halphen; A. Hallmann; M. Hanikenne; M. Hippler; W. Inwood; K. Jabbari; M. Kalanon; R. Kuras; P. A. Lefebvre; S. D. Lemaire; A. V. Lobanov; M. Lohr; A. Manuell; I. Meier; L. Mets; M. Mittag; T. Mittelmeier; J. V. Moroney; J. Moseley; C. Napoli; A. M. Nedelcu; K. Niyogi; S. V. Novoselov; I. T. Paulsen; G. Pazour; S. Purton; J.-P. Ral; D. M. Riano-Pachon; W. Riekhof; L. Rymarquis; M. Schroda; D. Stern; J. Umen; R. Willows; N. Wilson; S. L. Zimmer; J. Allmer; J. Balk; K. Bisova; C.-J. Chen; M. Elias; K. Gendler; C. Hauser; M. R. Lamb; H. Ledford; J. C. Long; J. Minagawa; M. D. Page; J. Pan; W. Pootakham; S. Roje; A. Rose; E. Stahlberg; A. M. Terauchi; P. Yang; S. Ball; C. Bowler; C. L. Dieckmann; V. N. Gladyshev; P. Green; R. Jorgensen; S. Mayfield; B. Mueller-Roeber; S. Rajamani; R. T. Sayre; P. Brokstein; I. Dubchak; D. Goodstein; L. Hornick; Y. W. Huang; J. Jhaveri; Y. Luo; D. Martinez; A. Ngau Wing Chi; B. Otillar; A. Poliakov; A. Porter; L. Szajkowski; G. Werner; K. Zhou; I. V. Grigoriev; D. S. Rokhsar; A. R. Grossman, *Science* **2007**, *318*, 245–251.
35. C. H. Wellman; P. L. Osterloff; U. Mohiuddin, *Nature* **2003**, *425*, 282–285.
36. B. Crandall-Stotler; R. E. Stotler, Morphology and Classification of the Marchantiophyta. In *Bryophyte Biology*; A. J. Shaw, B. Goffinet, Eds.; Cambridge University Press: Cambridge, UK, 2000; pp 21–70.
37. L. L. Forrest; E. C. Davis; D. G. Long; B. J. Crandall-Stotler; A. Clark; M. L. Hollingsworth, *The Bryologist* **2006**, *109*, 303–334.
38. R. J. Duff; J. C. Villarreal; D. C. Cargill; K. S. Renzaglia, *The Bryologist* **2007**, *110*, 214–243.
39. D. L. Nickrent; C. L. Parkinson; J. D. Palmer; R. J. Duff, *Mol. Biol. Evol.* **2000**, *17*, 1885–1895.
40. Y.-L. Qiu; L. Li; B. Wang; Z. Chen; V. Knoop; M. Groth-Malonek; O. Dombrovska; J. Lee; L. Kent; J. Rest; G. F. Estabrook; T. A. Hendry; D. W. Taylor; C. M. Testa; M. Ambros; B. Crandall-Stotler; R. J. Duff; M. Stech; W. Frey; D. Quandt; C. C. Davis, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15511–15516.
41. F. Němejc; B. Pacltová, *Palaeobotanist* **1974**, *21*, 23–26.
42. B. A. Thomas, *Ann. Bot.* **1972**, *36*, 155–161.
43. A. V. Troitsky; M. S. Ignatov; V. K. Bobrova; I. A. Milyutina, *Biochemistry (Moscow)* **2007**, *72*, 1368–1376.
44. Y. Asakawa; N. Tokunaga; M. Toyota; T. Takemoto; C. Suire, *J. Hattori Bot. Lab.* **1979**, *45*, 395–407.
45. Y. Asakawa; N. Tokunaga; T. Takemoto; S. Hattori; M. Mizutani; C. Suire, *J. Hattori Bot. Lab.* **1980**, *47*, 153–164.
46. S. R. Gradstein; R. Matsuda; Y. Asakawa, *J. Hattori Bot. Lab.* **1981**, *50*, 231–248.
47. S. R. Gradstein; R. Matsuda; Y. Asakawa, *Nova Hedwigia* **1985**, *80*, 63–86.
48. J. D. Connolly, Recent Advances in the Chemistry of the Hepaticae. In *Natural Products Chemistry*; R. I. Zalewski, J. J. Skolik, Eds.; Elsevier Science Publishers: Amsterdam, 1985; pp 3–14.
49. A. C. Figueiredo; M. Sim-Sim; J. G. Barroso; L. G; Pedro; P. A. G. Santos; S. S. Fontinha; J. Schripsema; S. G. Deans; J. J. C. Scheffer, *J. Essent. Oil Res.* **2002**, *14*, 439–442.
50. R. B. Rickards, *Geol. Mag.* **2000**, *137*, 207–209.
51. N. Radulović; G. Stojanović; R. Palić, *Phytother. Res.* **2006**, *20*, 85–88.
52. S.-M. Chaw; A. Zharkikh; H.-M. Sung; T.-C. Lau; W.-H. Li, *Mol. Biol. Evol.* **1997**, *14*, 56–68.
53. S.-M. Chaw; C. L. Parkinson; Y. Cheng; T. M. Vincent; J. D. Palmer, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 4086–4091.
54. S.-M. Chaw; T. W. Walters; C.-C. Chang; S.-H. Hu; S.-H. Chen, *Mol. Phylogen. Evol.* **2005**, *37*, 214–234.
55. K. D. Hill; D. W. Stevenson; R. Osborne, *Bot. Rev.* **2004**, *70*, 274–298.
56. H. Azuma; M. Kono, *J. Plant Res.* **2006**, *119*, 671–676.
57. A. Scott, *Nature* **1974**, *251*, 707–708.
58. A. C. Scott; W. G. Chaloner, *Proc. R. Soc. London Ser. B* **1983**, *220*, 163–182.
59. F. A. Bisby; Y. R. Roskov; T. M. Orrell; D. Nicolson; L. E. Paglinawan; N. Bailly; P. M. Kirk; T. Bourgoin; J. van Hertum, Species 2000 & ITIS Catalogue of Life: 2008 Annual Checklist. Digital resource at www.catalogueoflife.org/annual-checklist/2008/. Reading, U.K.
60. O. Faix; D. Meier; I. Grobe, *J. Anal. Appl. Pyrolysis* **1987**, *11*, 403–416.
61. S. Camarero; P. Bocchini; G. C. Galletti; A. T. Martínez, *Rapid Commun. Mass Spectrom.* **1999**, *13*, 630–636.
62. D. M. Shrimpton, *Bi-Monthly Research Notes – Canada, Forestry Service* **1974**, *30*, 12.
63. K. Snajberk; E. Zavarin, *Phytochemistry* **1975**, *14*, 2025–2028.

64. S. Shatar; R. P. Adams, *J. Essent. Oil Res.* **1996**, *8*, 549–552.
65. A. Wibe; A.-K. Borg-Karlson; T. Norin; H. Mustaparta, *J. Comp. Physiol. A* **1997**, *180*, 585–595.
66. S. Rezzi; A. Muselli; A. Bighelli; J. Casanova, *Rivista Italiana EPPOS* **1998**, 766–772.
67. G. D. L. Boyd. The composition of terpenes and 4-allylanisole in loblolly southern yellow pine oleoresin and their potential impact on host selection by the female southern pine beetle. Ph.D. Thesis. Mississippi State University, MS, USA, 1998.
68. J. Velásquez; M. E. Toro; O. Encinas; L. Rojas; A. Usubillaga, *Flavour Fragr. J.* **2000**, *15*, 432–433.
69. E. M. Pettersson; B. T. Sullivan; P. Anderson; C. W. Berisford; G. Birgersson, *J. Chem. Ecol.* **2000**, *26*, 2507–2525.
70. M. Krauze-Baranowska; M. Mardarowicz; M. Wiwart; L. Poblocka; M. Dynowska, *Z. Naturforsch. C* **2002**, *57*, 478–482.
71. M. W. Ghosn; N. A. Saliba; S. Y. Talhouk, *J. Essent. Oil Res.* **2006**, *18*, 445–447.
72. K. Kurose; D. Okamura; M. Yatagai, *Flavour Fragr. J.* **2007**, *22*, 10–20.
73. J. J. Brophy; R. J. Goldsack; A. C. Rozefelds, *J. Essent. Oil Res.* **2003**, *15*, 217–220.
74. R. P. Adams, *Biochem. Syst. Ecol.* **1998**, *26*, 637–645.
75. Z.-Q. Wang, *Ann. Bot.* **2004**, *94*, 281–288.
76. Z. Zhou; S. Zheng, *Nature* **2003**, *423*, 821–822.
77. G. Sun; Q. Ji; D. L. Dilcher; S. Zheng; K. C. Nixon; X. Wang, *Science* **2002**, *296*, 899–904.
78. R. F. Thorne, *Taxon* **2002**, *51*, 511–512.
79. J. Kałużna-Czaplińska, *Acta Chromat.* **2007**, *19*, 279–282.
80. R. Nishida; K.-H. Tan; S.-L. Wee; A. K.-W. Hee; Y.-C. Toong, *Biochem. Syst. Ecol.* **2004**, *32*, 245–252.
81. G. M. Nano; A. Martelli; P. Sancin, *Rivista Italiana Essenze, Profumi, Piante Officinali, Aromi, Saponi, Cosmetici, Aerosol* **1966**, *48*, 409–412.
82. O. Vostrowsky; K. Michaelis; H. Ihm; R. Zintl; K. Knobloch, *Z. Lebensm.-Unters. Forsch.* **1981**, *173*, 365–367.
83. W. Lenz, *Zeit. anal. Chem.* **1894**, *33*, 193–200.
84. H. Thoms, *Arch. Pharm.* **1903**, *241*, 592–603.
85. A. K. Srivastava; S. K. Srivastava; K. V. Syamsundar, *Flavour Fragr. J.* **2005**, *20*, 51–53.
86. K. Chaieb; H. Hajlaoui; T. Zmantar; A. B. Kahla-Nakbi; M. Rouabhia; K. Mahdouani; A. Bakhrouf, *Phytother. Res.* **2007**, *21*, 501–506.
87. E. H. A. Andrade; L. M. M. Carreira; M. H. L. da Silva; J. D. da Silva; C. N. Bastos; P. J. C. Sousa; E. F. Guimarães; J. G. S. Maia, *Chem. Biodivers.* **2008**, *5*, 197–208.
88. A. J. Amonkar; M. Nagabhushan; A. V. D'Souza; S. V. Bhide, *Food Chem. Toxicol.* **1986**, *24*, 1321–1324.
89. F. W. Semmler, *Ber.* **1890**, *23*, 1803–1810.
90. F. B. Power; A. H. Salway, *J. Chem. Soc.* **1907**, *91*, 2037–2058.
91. A. W. Archer, *J. Chromatogr. A* **1988**, *438*, 117–121.
92. D. Ehlers; J. Kirchhoff; D. Gerard; K.-W. Quirin, *Int. J. Food Sci. Tech.* **1998**, *33*, 215–223.
93. G. R. Mallavarapu; S. Ramesh, *J. Med. Aromat. Plant Sci.* **1998**, *20*, 746–748.
94. M. Jukić; O. Politeo; M. Miloš, *Croat. Chem. Acta* **2006**, *79*, 209–214.
95. M. Marotti; R. Piccaglia, *J. Essent. Oil Res.* **1992**, *4*, 569–576.
96. A. Orav; A. Raal; E. Arak, *Nat. Prod. Res. Part A* **2008**, *22*, 227–232.
97. R. Spitaler; E.-P. Ellmerer-Müller; C. Zidorn; H. Stuppner, *Sci. Pharm.* **2002**, *70*, 101–109.
98. E. P. Lichtenstein; T. T. Liang; K. R. Schulz; H. K. Schnoes; G. T. Carter, *J. Agric. Food Chem.* **1974**, *22*, 658–664.
99. A. J. MacLeod; C. H. Snyder; G. Subramanian, *Phytochemistry* **1985**, *24*, 2623–2627.
100. A. J. MacLeod; G. MacLeod; G. Subramanian, *Phytochemistry* **1988**, *27*, 373–375.
101. L. Tóth, *Planta Med.* **1967**, *15*, 157–172.
102. Y.-R. Naves, *Perfum. Essent. Oil Rec.* **1950**, *41*, 286–287.
103. R. J. Grayer; G. C. Kite; F. J. Goldstone; S. E. Bryan; A. Paton; E. Putievsky, *Phytochemistry* **1996**, *43*, 1033–1039.
104. M. Nitta; H. Kobayashi; M. Ohnishi-Kameyama; T. Nagamine; M. Yoshida, *Biochem. Syst. Ecol.* **2006**, *34*, 25–37.
105. N. Tabanca; A. W. Douglas; E. Bedir; F. E. Dayan; N. Kirimer; K. H. C. Baser; Z. Aytac; I. A. Khan; B. E. Scheffler, *Plant Genet. Resour.* **2005**, *3*, 149–169.
106. K. H. C. Baser; N. Tabanca; N. Kirimer; E. Bedir; I. A. Khan; D. E. Wedge, *Pure Appl. Chem.* **2007**, *79*, 539–556.
107. G. Özek; T. Özek; G. Iscan; K. H. C. Baser; A. Duran; E. Hamzaoglu, *J. Essent. Oil Res.* **2008**, *20*, 186–190.
108. O. Özek; T. Özek; G. Iscan; K. H. C. Baser; E. Hamzaoglu; A. Duran, *S. Afr. J. Bot.* **2007**, *73*, 563–569.
109. N. Tabanca; B. Demirci; K. H. C. Baser; E. Mincsovics; S. I. Khan; M. R. Jacob; D. E. Wedge, *J. Chromatog. B* **2007**, *850*, 221–229.
110. A. Tosun; M. Kürkçüoglu; K. H. C. Baser; H. Duman, *J. Essent. Oil Res.* **2007**, *19*, 153–154.
111. A. Smelcerovic; M. Spiteller; A. P. Ligon; Z. Smelcerovic; N. Raabe, *Biochem. Syst. Ecol.* **2007**, *35*, 99–113.
112. K. H. C. Baser; H. Duman; Z. Aytaç, *J. Essent. Oil Res.* **2000**, *12*, 475–477.
113. I. Bohn; K.-H. Kubeczka; W. Schultze, *Planta Med.* **1989**, *55*, 489–490.
114. V. Dev; C. S. Mathela; A. B. Melkani; N. M. Pope; N. S. Sturm; A. T. Bottini, *Phytochemistry* **1989**, *28*, 1531–1532.
115. S. A. Rensing; D. Lang; A. D. Zimmer; A. Terry; A. Salamov; H. Shapiro; T. Nishiyama; P.-F. Perroud; E. A. Lindquist; Y. Kamisugi; T. Tanahashi; K. Sakakibara; T. Fujita; K. Oishi; T. Shin-I; Y. Kuroki; A. Toyoda; Y. Suzuki; S.-I. Hashimoto; K. Yamaguchi; S. Sugano; Y. Kohara; A. Fujiyama; A. Anterola; S. Aoki; N. Ashton; W. B. Barbazuk; E. Barker; J. L. Bennetzen; R. Blankenship; S. H. Cho; S. K. Dutcher; M. Estelle; J. A. Fawcett; H. Gundlach; K. Hanada; A. Heyl; K. A. Hicks; J. Hughes; M. Lohr; K. Mayer; A. Melkozernov; T. Murata; D. R. Nelson; B. Pils; M. Prigge; B. Reiss; T. Renner; S. Rombauts; P. J. Rushton; A. Sanderfoot; G. Schween; S.-H. Shiu; K. Stueber; F. L. Theodoulou; H. Tu; Y. Van de Peer; P. J. Verrier; E. Waters; A. Wood; L. Yang; D. Cove; A. C. Cuming; M. Hasebe; S. Lucas; B. D. Mishler; R. Reski; I. V. Grigoriev; R. S. Quatrano; J. L. Boore, *Science* **2008**, *319*, 64–69.
116. M. V. Silber; H. Meimberg; J. Ebel, *Phytochemistry* **2008**, *69*, 2449–2456.
117. M. M. A. El-Gendy; U. W. Hawas; M. Jaspars, *J. Antibiot.* **2008**, *61*, 379–386.
118. M. Berner; D. Krug; C. Bihlmaier; A. Vente; R. Müller; A. Bechthold, *J. Bacteriol.* **2006**, *188*, 2666–2673.
119. F. Cullmann; K.-P. Adam; H. Becker, *Phytochemistry* **1993**, *34*, 831–834.

120. F. Cullmann; K.-P. Adam; J. Zapp; H. Becker, *Phytochemistry* **1996**, *41*, 611–615.
121. F. Cullmann; A. Schmidt; F. Schuld; M. L. Trennheuser; H. Becker, *Phytochemistry* **1999**, *52*, 1647–1650.
122. H. Tazaki; T. Hayashida; F. Ishikawa; D. Taguchi; T. Takasawa; K. Nabeta, *Tetrahedron Lett.* **1999**, *40*, 101–104.
123. H. Tazaki; K.-P. Adam; H. Becker, *Phytochemistry* **1995**, *40*, 1671–1675.
124. F. Cullmann; H. Becker, *Phytochemistry* **1999**, *52*, 1651–1656.
125. U. Martini; J. Zapp; H. Becker, *Phytochemistry* **1998**, *49*, 1139–1146.
126. J. M. Scher; J. Zapp; H. Becker, *Phytochemistry* **2003**, *62*, 769–777.
127. H. Tazaki, *Curr. Top. Phytochem.* **2002**, *5*, 141–151.
128. F. Nagashima; M. Toyota; Y. Asakawa, *Phytochemistry* **1990**, *29*, 2169–2174.
129. R. Takeda; J. Hasegawa; M. Shinozaki, *Tetrahedron Lett.* **1990**, *31*, 4159–4162.
130. R. Takeda; J. Hasegawa; K. Sinozaki, Phenolic compounds from Anthocerotae. In *Bryophytes. Their Chemistry and Chemical Taxonomy*; H. D. Zinsmeister, R. Mues, Eds.; Clarendon Press: Oxford, UK, 1990; pp 201–207.
131. R. C. Lin; A.-L. Skaltsounis; E. Seguin; F. Tillequin; M. Koch, *Planta Med.* **1994**, *60*, 168–170.
132. E. White; G. H. N. Towers, *Phytochemistry* **1967**, *6*, 663–667.
133. Z. Jin; Y. Matsumoto; T. Tange; T. Akiyama; M. Higuchi; T. Ishii; K. Iiyama, *J. Wood Sci.* **2005**, *51*, 424–426.
134. X.-k. Zheng; S.-p. Shi; Y.-f. Bi; W.-s. Feng; J.-f. Wang; J.-z. Niu, *Yaoxue Xuebao (Acta Pharmaceutica Sinica)* **2004**, *39*, 719–721.
135. X. K. Zheng; K. K. Li; Y. Z. Wang; W. S. Feng, *Chin. Chem. Lett.* **2008**, *19*, 79–81.
136. J.-K. Weng; X. Li; J. Stout; C. Chapple, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 7887–7892.
137. H. Wada; T. Kido; N. Tanaka; T. Murakami; Y. Saiki; C.-M. Chen, *Chem. Pharm. Bull.* **1992**, *40*, 2099–2101.
138. L. B. Davin; C.-Z. Wang; G. L. Helms; N. G. Lewis, *Phytochemistry* **2003**, *62*, 501–511.
139. T. Satake; T. Murakami; Y. Saiki; C.-M. Chen, *Chem. Pharm. Bull.* **1978**, *26*, 1619–1622.
140. S. Cao; M. M. Radwan; A. Norris; J. S. Miller; F. Ratovoson; A. Mamisoa; R. Andriantsiferana; V. E. Rasamison; S. Rakotonandrasana; D. G. I. Kingston, *J. Nat. Prod.* **2006**, *69*, 284–286.
141. N. G. Lewis; M. J. Kato; N. Lopes; L. B. Davin, Lignans: Diversity, Biosynthesis and Function. In *Chemistry of the Amazon: Biodiversity, Natural Products, and Environmental Issues*; R. S. Seidl, O. R. Gottlieb, M. A. C. Kaplan, Eds.; ACS Symposium Series: Washington, DC, USA, 1995; Vol. 588, pp 135–167.
142. M. A. Castro; M. Gordaliza; J. M. Miguel Del Corral; A. San Feliciano, *Phytochemistry* **1996**, *41*, 995–1011.
143. D. R. Gang; M. Fujita; L. B. Davin; N. G. Lewis, The ''Abnormal Lignins'': Mapping Heartwood Formation Through the Lignan Biosynthetic Pathway. In *Lignin and Lignan Biosynthesis*; N. G. Lewis, S. Sarkanen, Eds.; ACS Symposium Series: Washington, DC, USA, 1998; Vol. 697, pp 389–421.
144. B. Holmbom; C. Eckerman; P. Eklund; J. Hemming; L. Nisula; M. Reunanen; R. Sjöholm; A. Sundberg; K. Sundberg; S. Willför, *Phytochem. Rev.* **2003**, *2*, 331–340.
145. A. San Feliciano; M. Medarde; J. L. Lopez; P. Puebla; J. M. Miguel Del Corral; A. F. Barrero, *Phytochemistry* **1989**, *28*, 2863–2866.
146. J.-M. Fang; K.-C. Hsu; Y.-S. Cheng, *Phytochemistry* **1989**, *28*, 3553–3555.
147. M. Nose; M. A. Bernards; M. Furlan; J. Zajicek; T. L. Eberhardt; N. G. Lewis, *Phytochemistry* **1995**, *39*, 71–79.
148. L. H. Briggs; R. C. Cambie; J. L. Hoare, *Tetrahedron* **1959**, *7*, 262–269.
149. L. H. Briggs; R. C. Cambie; J. L. Hoare, *Tetrahedron Lett.* **1959**, *1*, 14–15.
150. A. San Feliciano; J. M. Miguel Del Corral; M. Gordaliza; A. Castro, *Phytochemistry* **1991**, *30*, 3483–3485.
151. J. A. F. Gardner; G. M. Barton; H. MacLean, *Can. J. Chem.* **1959**, *37*, 1703–1709.
152. J. A. F. Gardner; B. F. MacDonald; H. MacLean, *Can. J. Chem.* **1960**, *38*, 2387–2394.
153. J. A. F. Gardner; E. P. Swan; S. A. Sutherland; H. MacLean, *Can. J. Chem.* **1966**, *44*, 52–58.
154. R. L. Krahmer; R. W. Hemingway; W. E. Hillis, *Wood Sci.Technol.* **1970**, *4*, 122–139.
155. F. Kawamura; H. Ohashi; S. Kawai; F. Teratani; Y. Kai, *Mokuzai Gakkaishi* **1996**, *42*, 301–307.
156. C. I. Johansson; J. N. Saddler; R. P. Beatson, *Holzforschung* **2000**, *54*, 246–254.
157. H. Hovelstad; I. Leirset; K. Oyaas; A. Fiksdahl, *Molecules* **2006**, *11*, 103–114.
158. J.-M. Fang; C.-K. Lee; Y.-S. Cheng, *Phytochemistry* **1992**, *31*, 3659–3661.
159. P. K. Agrawal; S. K. Agarwal; R. P. Rastogi, *Phytochemistry* **1980**, *19*, 1260–1261.
160. W.-C. Su; J.-M. Fang; Y.-S. Cheng, *Phytochemistry* **1995**, *40*, 563–566.
161. C. Kraus; G. Spiteller, *Phytochemistry* **1997**, *44*, 59–67.
162. S. Suzuki; T. Umezawa, *J. Wood Sci.* **2007**, *53*, 273–284.
163. C. R. Enzell; B. R. Thomas, *Acta Chem. Scand.* **1965**, *19*, 913–919.
164. C. R. Enzell; Y. Hirose; B. R. Thomas, *Tetrahedron Lett.* **1967**, *8*, 793–798.
165. P. Henley-Smith; D. A. Whiting, *Phytochemistry* **1976**, *15*, 1285–1287.
166. K. Takahashi; M. Yasue; K. Ogiyama, *Phytochemistry* **1988**, *27*, 1550–1552.
167. K. Takahashi; K. Ogiyama, *Mokuzai Gakkaishi* **1986**, *32*, 457–461.
168. S. Suzuki; M. Yamamura; M. Shimada; T. Umezawa, *J. Chem. Soc. Chem. Commun.* **2004**, 2838–2839.
169. P. Daniels; H. Erdtman; K. Nishimura; T. Novin; P. Kierkegaard; A.-M. Pilotti, *J. Chem. Soc. Chem. Commun.* **1972**, 246–247.
170. A. Enoki; S. Takahama; K. Kitao, *Mokuzai Gakkaishi* **1977**, *23*, 579–586.
171. A. Enoki; S. Takahama; K. Kitao, *Mokuzai Gakkaishi* **1977**, *23*, 587–593.
172. H. Ohashi; S. Kawai; Y. Sakurai; M. Yasue, *Phytochemistry* **1992**, *31*, 1371–1373.
173. I. Iliya; Z. Ali; T. Tanaka; M. Iinuma; M. Furusawa; K.-i. Nakaya; J. Murata; D. Darnaedi; N. Matsuura; M. Ubukata, *Phytochemistry* **2003**, *62*, 601–606.
174. M. A. M. Nawwar; H. H. Barakat; J. Buddrus; M. Linscheid, *Phytochemistry* **1985**, *24*, 878–879.
175. S. V. Pullela; S. Takamatsu; S. I. Khan; I. A. Khan, *Planta Med.* **2005**, *71*, 789–791.
176. M. Tamada; K. Endo; H. Hikino, *Heterocycles* **1979**, *12*, 783–786.
177. M. Tamada; K. Endo; H. Hikino; C. Kabuto, *Tetrahedron Lett.* **1979**, *20*, 873–876.
178. C. Konno; M. Tamada; K. Endo; H. Hikino, *Heterocycles* **1980**, *14*, 295–298.
179. T. Kariyone; H. Kimura; I. Nakamura, *Yakugaku Zasshi* **1958**, *78*, 1152–1155.

180. E. Bedir; I. I. Tatli; R. A. Khan; J. Zhao; S. Takamatsu; L. A. Walker; P. Goldman; I. A. Khan, *J. Agric. Food Chem.* **2002**, *50*, 3150–3155.
181. R. S. Ward, *Nat. Prod. Rep.* **1997**, *14*, 43–74.
182. R. S. Ward, *Nat. Prod. Rep.* **1999**, *16*, 75–96.
183. R. D. Hartley; E. C. Jones, *Phytochemistry* **1976**, *15*, 1157–1160.
184. H. U. Markwalder; H. Neukom, *Phytochemistry* **1976**, *15*, 836–837.
185. C.-M. Chen; F.-Y. Jan; M.-T. Chen; T.-J. Lee, *Heterocycles* **1989**, *29*, 411–414.
186. A. R. Carroll; W. C. Taylor, *Aust. J. Chem.* **1991**, *44*, 1615–1626.
187. S. E. Drewes; M. M. Horn; B. M. Sehlapelo; N. Ramesar; J. S. Field; R. S. Shaw; P. Sandor, *Phytochemistry* **1995**, *38*, 1505–1508.
188. S. Nishibe; A. Sakushima; S. Kitagawa; B. Klimek; R. Benecke; H. Thieme, *Shoyakugaku Zasshi* **1988**, *42*, 324–328.
189. J. L. Hartwell; A. W. Schrecker, *J. Am. Chem. Soc.* **1951**, *73*, 2909–2916.
190. Y. Ikeya; H. Taguchi; I. Yosioka; H. Kobayashi, *Chem. Pharm. Bull.* **1979**, *27*, 1383–1394.
191. R. A. Davis; A. R. Carroll; S. Duffy; V. M. Avery; G. P. Guymer; P. I. Forster; R. J. Quinn, *J. Nat. Prod.* **2007**, *70*, 1118–1121.
192. E.-C. Wang; M.-H. Shih; M.-C. Liu; M.-T. Chen; G.-H. Lee, *Heterocycles* **1996**, *43*, 969–976.
193. B. Goezler; B. Kivcak; G. Arar; T. Goezler; M. Hesse, *Heterocycles* **1994**, *39*, 243–250.
194. M. Haruna; T. Koube; K. Ito; H. Murata, *Chem. Pharm. Bull.* **1982**, *30*, 1525–1527.
195. S. M. Kupchan; K. L. Stevens; E. A. Rohlfing; B. R. Sickles; A. T. Sneden; R. W. Miller; R. F. Bryan, *J. Org. Chem.* **1978**, *43*, 586–590.
196. M. N. Lopes; M. S. Da Silva; J. M. Barbosa; Z. S. Ferreira; M. Yoshida; O. R. Gottlieb, *Phytochemistry* **1986**, *25*, 2609–2612.
197. N. H. Anh; H. Ripperger; A. Porzel; T. V. Sung; G. Adam, *Phytochemistry* **1997**, *46*, 569–571.
198. R. Braz Filho; M. G. De Carvalho; O. R. Gottlieb; J. G. S. Maia; M. L. Da Silva, *Phytochemistry* **1981**, *20*, 2049–2050.
199. Z. Zhang; H. N. ElSohly; X.-C. Li; S. I. Khan; S. E. Broedel, Jr.; R. E. Raulli; R. L. Cihlar; C. Burandt; L. A. Walker, *J. Nat. Prod.* **2003**, *66*, 548–550.
200. H. Taguchi; Y. Ikeya, *Chem. Pharm. Bull*. **1977**, *25*, 264–366.
201. V. S. Parmar; S. C. Jain; K. S. Bisht; R. Jain; P. Taneja; A. Jha; O. D. Tyagi; A. K. Prasad; J. Wengel; C. E. Olsen; P. M. Boll, *Phytochemistry* **1997**, *46*, 597–673.
202. R. D. Hartley; C. W. Ford, Phenolic Constituents of Plant Cell Walls and Wall Biodegradability. In *Plant Cell Wall Polymers. Biogenesis and Biodegradation*; N. G. Lewis, M. G. Paice, Eds.; ACS Symposium Series: Washington, DC, USA, 1989; Vol. 399, 137–145.
203. C. W. Ford; R. D. Hartley, *J. Sci. Food Agric.* **1990**, *50*, 29–43.
204. S. Tachibana; K. Ohkubo; G. H. N. Towers, *Phytochemistry* **1992**, *31*, 81–83.
205. A. Suga; Y. Takaishi; S. Goto; T. Munakata; I. Yamauchi; K. Kogure, *Phytochemistry* **2003**, *64*, 991–996.
206. W.-Y. Tsui; G. D. Brown, *Phytochemistry* **1996**, *43*, 1413–1415.
207. S. Suzuki; M. Yamamura; T. Hattori; T. Nakatsubo; T. Umezawa, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 21008–21013.
208. E. Minami; M. Taki; S. Takaishi; Y. Iijima; S. Tsutsumi; T. Akiyama, *Chem. Pharm. Bull.* **2000**, *48*, 389–392.
209. I. Messana; J. D. Msonthi; Y. De Vicente; G. Multari; C. Galeffi, *Phytochemistry* **1989**, *28*, 2807–2809.
210. G. B. Marini-Bettolo; M. Nicoletti; I. Messana; C. Galeffi; J. D. Msonthi; W. A. Chapya, *Tetrahedron* **1985**, *41*, 665–670.
211. K. Terada; W. Kamisako, *Biol. Pharm. Bull.* **1999**, *22*, 561–566.
212. G. B. Marini Bettolo; M. Patamia; M. Nicoletti; C. Galeffi; I. Messana, *Tetrahedron* **1982**, *38*, 1683–1687.
213. W.-L. Chang; C.-H. Chen; S.-S. Lee, *J. Nat. Prod.* **1999**, *62*, 734–739.
214. K. M. Mohamed; M. A. Fouad; K. Matsunami; M. S. Kamel; H. Otsuka, *ARKIVOC* **2007**, 63–70.
215. B. Talapatra; T. Ray; S. K. Talapatra, *Indian J. Chem. Sect B* **1976**, *14B*, 613–614.
216. K. Kawazoe; A. Yutani; Y. Takaishi, *Phytochemistry* **1999**, *52*, 1657–1659.
217. M. Miyazawa; M. Hisama, *J. Agric. Food Chem.* **2003**, *51*, 6413–6422.
218. P. J. C. Benevides; P. Sartorelli; M. J. Kato, *Phytochemistry* **1999**, *52*, 339–343.
219. T. J. Mabry; D. R. DiFeo, Jr.; M. Sakakibara; C. F. Bohnstedt, Jr.; D. Seigler, The Natural Products Chemistry of *Larrea*. In *Creosote Bush – Biology and Chemistry of Larrea in New World Deserts*; T. J. Mabry, J. H. Hunziker, D. R. DiFeo, Jr., Eds.; Dowden, Hutchinson & Ross, Inc.: Stroudsburg, PA, USA, 1977; pp 115–134.
220. T. Umezawa; L. B. Davin; N. G. Lewis, *Biochem. Biophys. Res. Commun.* **1990**, *171*, 1008–1014.
221. T. Umezawa; L. B. Davin; E. Yamamoto; D. G. I. Kingston; N. G. Lewis, *J. Chem. Soc. Chem. Commun.* **1990**, 1405–1408.
222. T. Umezawa; L. B. Davin; N. G. Lewis, *J. Biol. Chem.* **1991**, *266*, 10210–10217.
223. T. Katayama; L. B. Davin; N. G. Lewis, *Phytochemistry* **1992**, *31*, 3875–3881.
224. T. Katayama; L. B. Davin; A. Chu; N. G. Lewis, *Phytochemistry* **1993**, *33*, 581–591.
225. Y. Jiao; L. B. Davin; N. G. Lewis, *Phytochemistry* **1998**, *49*, 387–394.
226. Z.-Q. Xia; M. A. Costa; J. Proctor; L. B. Davin; N. G. Lewis, *Phytochemistry* **2000**, *55*, 537–549.
227. M. Chiba; S. Hisada; S. Nishibe, *Shoyakugaku Zasshi* **1978**, *32*, 194–197.
228. K.-J. Kim, *Plant Syst. Evol.* **1999**, *218*, 113–123.
229. L. B. Davin; D. L. Bedgar; T. Katayama; N. G. Lewis, *Phytochemistry* **1992**, *31*, 3869–3874.
230. L. B. Davin; H.-B. Wang; A. L. Crowell; D. L. Bedgar; D. M. Martin; S. Sarkanen; N. G. Lewis, *Science* **1997**, *275*, 362–367.
231. S. C. Halls; L. B. Davin; D. M. Kramer; N. G. Lewis, *Biochemistry* **2004**, *43*, 2587–2595.
232. D. R. Gang; M. A. Costa; M. Fujita; A. T. Dinkova-Kostova; H.-B. Wang; V. Burlat; W. Martin; S. Sarkanen; L. B. Davin; N. G. Lewis, *Chem. Biol.* **1999**, *6*, 143–151.
233. M. K. Kim; J.-H. Jeon; M. Fujita; L. B. Davin; N. G. Lewis, *Plant Mol. Biol.* **2002**, *49*, 199–214.
234. S. C. Halls; N. G. Lewis, *Biochemistry* **2002**, *41*, 9455–9461.
235. D. Frishman; P. Argos, *Protein Eng.* **1996**, *9*, 133–142.
236. L. Zhuang; O. Seligmann; K. Jurcic; H. Wagner, *Planta Med.* **1982**, *45*, 172–176.
237. T. Nakatsubo; M. Mizutani; S. Suzuki; T. Hattori; T. Umezawa, *J. Biol. Chem.* **2008**, *283*, 15550–15557.

238. M. Kwon; V. Burlat; L. B. Davin; N. G. Lewis, Localization of Dirigent Protein Involved in Lignan Biosynthesis: Implications for Lignification at the Tissue and Subcellular Level. In *Plant Polyphenols 2: Chemistry, Biology, Pharmacology, Ecology*; G. G. Gross, R. W. Hemingway, T. Yoshida, Eds.; Kluwer Academic/Plenum Publishers: New York, NY, USA, 1999; pp 393–411.
239. V. Burlat; M. Kwon; L. B. Davin; N. G. Lewis, *Phytochemistry* **2001**, *57*, 883–897.
240. A. M. Patten; L. B. Davin; N. G. Lewis, *Phytochemistry* **2008**, *69*, 3032–3037.
241. L. B. Davin; N. G. Lewis, *Plant Physiol.* **2000**, *123*, 453–461.
242. M. K. Kim; J.-H. Jeon; L. B. Davin; N. G. Lewis, *Phytochemistry* **2002**, *61*, 311–322.
243. S. Ralph; J.-Y. Park; J. Bohlmann; S. D. Mansfield, *Plant Mol. Biol.* **2006**, *60*, 21–40.
244. S. G. Ralph; S. Jancsik; J. Bohlmann, *Phytochemistry* **2007**, *68*, 1975–1991.
245. S. Arteaga; A. Andrade-Cetto; R. Cárdenas, *J. Ethnopharmacol.* **2005**, *98*, 231–239.
246. C. Konno; Z.-Z. Lu; H.-Z. Xue; C. A. J. Erdelmeier; D. Meksuriyen; C.-T. Che; G. A. Cordell; D. D. Soejarto; D. P. Waller; H. H. S. Fong, *J. Nat. Prod.* **1990**, *53*, 396–406.
247. S. G. A. Moinuddin; S. Hishiyama; M.-H. Cho; L. B. Davin; N. G. Lewis, *Org. Biomol. Chem.* **2003**, *1*, 2307–2313.
248. P. Budowski; K. S. Markley, *Chem. Rev.* **1951**, *48*, 125–151.
249. M. Namiki, *Crit. Rev. Food Sci. Nutr.* **2007**, *47*, 651–673.
250. Y. Fukuda; T. Osawa; M. Namiki; T. Ozaki, *Agric. Biol. Chem.* **1985**, *49*, 301–306.
251. T. Osawa; M. Nagata; M. Namiki; Y. Fukuda, *Agric. Biol. Chem.* **1985**, *49*, 3351–3352.
252. Y. Fukuda; M. Nagata; T. Osawa; M. Namiki, *J. Amer. Oil Chem. Soc.* **1986**, *63*, 1027–1031.
253. A. De Candolle, *Origin of Cultivated Plants*; D. Appleton and Company: New York, NY, 1902.
254. H. Katsuzaki; S. Kawakishi; T. Osawa, *Phytochemistry* **1994**, *35*, 773–776.
255. S. N. Ryu; C.-T. Ho; T. Osawa, *J. Food Lipids* **1998**, *5*, 17–28.
256. A. A. Moazzami; R. E. Andersson; A. Kamal-Eldin, *J. Agric. Food Chem.* **2006**, *54*, 633–638.
257. W. Bauer; M. H. Zenk, *Phytochemistry* **1991**, *30*, 2953–2961.
258. E. Ono; M. Nakai; Y. Fukui; N. Tomimori; M. Fukuchi-Mizutani; M. Saito; H. Satake; T. Tanaka; M. Katsuta; T. Umezawa; Y. Tanaka, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10116–10121.
259. N. Ikezawa; K. Iwasa; F. Sato, *FEBS J.* **2007**, *274*, 1019–1035.
260. A. Noguchi; Y. Fukui; A. Iuchi-Okada; S. Kakutani; H. Satake; T. Iwashita; M. Nakao; T. Umezawa; E. Ono, *Plant J.* **2008**, *54*, 415–427.
261. P. A. Marchand; M. J. Kato; N. G. Lewis, *J. Nat. Prod.* **1997**, *60*, 1189–1192.
262. M. J. Kato; A. Chu; L. B. Davin; N. G. Lewis, *Phytochemistry* **1998**, *47*, 583–591.
263. A. T. Dinkova-Kostova; D. R. Gang; L. B. Davin; D. L. Bedgar; A. Chu; N. G. Lewis, *J. Biol. Chem.* **1996**, *271*, 29473–29482.
264. M. Fujita; D. R. Gang; L. B. Davin; N. G. Lewis, *J. Biol. Chem.* **1999**, *274*, 618–627.
265. D. R. Gang; H. Kasahara; Z.-Q. Xia; K. Vander Mijnsbrugge; G. Bauw; W. Boerjan; M. Van Montagu; L. B. Davin; N. G. Lewis, *J. Biol. Chem.* **1999**, *274*, 7516–7527.
266. T. Min; H. Kasahara; D. L. Bedgar; B. Youn; P. K. Lawrence; D. R. Gang; S. C. Halls; H. Park; J. L. Hilsenbeck; L. B. Davin; N. G. Lewis; C. Kang, *J. Biol. Chem.* **2003**, *278*, 50714–50723.
267. N. G. Lewis; L. B. Davin; A. T. Dinkova-Kostova; J. D. Ford; M. Fujita; D. R. Gang; S. Sarkanen. Recombinant Pinoresinol/Lariciresinol Reductase, Recombinant Dirigent Protein, and Methods of Use. U.S. Patent 6,210,942, 3 April, 2001.
268. N. G. Lewis; L. B. Davin; S.-J. Kim; D. G. Vassão; A. M. Patten; D. Eichinger. Genes encoding chavicol/eugenol synthase form the creosote bush (*Larrea tridentata).* PCT Application no. PCT/US2007/069911, filed 29 May, 2007.
269. A. Chu; A. Dinkova; L. B. Davin; D. L. Bedgar; N. G. Lewis, *J. Biol. Chem.* **1993**, *268*, 27026–27033.
270. K.-W. Kim; S. G. A. Moinuddin; L. B. Davin; N. G. Lewis **2010**, Manuscript in preparation.
271. M. Hickey, *100 Families of Flowering Plants*, 2nd ed.; University Press: Cambridge, UK, 1988.
272. A. Diederichsen; K. Richards, Cultivated Flax and the Genus *Linum* L. Taxonomy and Germplasm Conservation. In *Flax. The Genus Linum*; A. D. Muir, N. D. Westcott, Eds.; Taylor & Francis: New York, NY, USA, 2003; pp 22–54.
273. P. P. Kolodziejczyk; P. Fedec, Processing Flaxseed for Human Consumption. In *Flaxseed in Human Nutrition*; S. C. Cunnane, L. U. Thompson, Eds.; AOCS Press: Champaign, IL, USA, 1995; pp 261–280.
274. D. Zohary; M. Hopf, Oil and Fibre Crops. In *Domestication of Plants in the Old World. The Origin and Spread of Cultivated Plants in West Asia, Europe and the Nile Valley*; 2nd ed., D. Zohary, M. Hopf, Eds.; Clarendon Press: Oxford, UK, 1993; pp 118–133.
275. J. D. Ford; K.-S. Huang; H.-B. Wang; L. B. Davin; N. G. Lewis, *J. Nat. Prod.* **2001**, *64*, 1388–1397.
276. A. Kamal-Eldin; N. Peerlkamp; P. Johnsson; R. Andersson; R. E. Andersson; L. N. Lundgren; P. Åman, *Phytochemistry* **2001**, *58*, 587–590.
277. K. Struijs; J.-P. Vincken; R. Verhoef; W. H. M. van Oostveen-van Casteren; A. G. J. Voragen; H. Gruppen, *Phytochemistry* **2007**, *68*, 1227–1235.
278. K. Struijs; J.-P. Vincken; T. G. Doeswijk; A. G. J. Voragen; H. Gruppen, *Phytochemistry* **2009**, *70*, 262–269.
279. K. Struijs; J.-P. Vincken; R. Verhoef; A. G. J. Voragen; H. Gruppen, *Phytochemistry* **2008**, *69*, 1250–1260.
280. T. J. Schmidt; S. Hemmati; E. Fuss; A. W. Alfermann, *Phytochem. Anal.* **2006**, *17*, 299–311.
281. T. J. Schmidt; A. W. Alfermann; E. Fuss, *Rapid Commun. Mass Spectrom* **2008**, *22*, 3642–3650.
282. J. D. Ford. Cancer Chemopreventive Flax Seed Lignans: Delineating the Metabolic Pathway(s) to the SDG-HMG Ester-Linked Polymers. Ph.D. Thesis, Washington State University, WA, USA, 2001.
283. L. B. Davin; N. G. Lewis, *Phytochem. Rev.* **2003**, *2*, 257–288.
284. K. H. Teoh; J. D. Ford; M.-R. Kim; L. B. Davin; N. G. Lewis, Delineating the Metabolic Pathway(s) to Secoisolariciresinol Diglucoside Hydroxymethyl Glutarate Oligomers in Flaxseed (*Linum usitatissimum)*. In *Flaxseed in Human Nutrition*, 2nd ed; L. U. Thompson, S. C. Cunnane, Eds.; AOCS Press: Champaign, IL, USA, 2003; pp 41–62.
285. C. B. I. von Heimendahl; K. M. Schäfer; P. Eklund; R. Sjöholm; T. J. Schmidt; E. Fuss, *Phytochemistry* **2005**, *66*, 1254–1263.
286. Ü. Bayindir; A. W. Alfermann; E. Fuss, *Plant J.* **2008**, *55*, 810–820.
287. S. Hemmati; T. J. Schmidt; E. Fuss, *FEBS Lett.* **2007**, *581*, 603–610.
288. K. Federolf; A. W. Alfermann; E. Fuss, *Phytochemistry* **2007**, *68*, 1397–1406.

289. A. Mohagheghzadeh; T. J. Schmidt; Ü. Bayindir; E. Fuss; I. Mehregan; A. W. Alfermann, *Planta Med.* **2006**, *72*, 1165–1167.
290. T. J. Schmidt; S. Vößing; M. Klaes; S. Grimme, *Planta Med.* **2007**, *73*, 1574–1580.
291. M. Kwon; L. B. Davin; N. G. Lewis, *Phytochemistry* **2001**, *57*, 899–914.
292. C. Hano; I. Martin; O. Fliniaux; B. Legrand; L. Gutierrez; R. R. J. Arroo; F. Mesnard; F. Lamblin; E. Lainé, *Planta* **2006**, *224*, 1291–1301.
293. B. Madhusudhan; D. Wiesenborn; K. Tostenson; J. Schwarz; J. Gillespie, *Lebensm.-Wiss. u.-Technol.* **2000**, *33*, 268–275.
294. H. Jörnvall; J.-O. Höög; B. Persson, *FEBS Lett.* **1999**, *445*, 261–264.
295. Z.-Q. Xia; M. A. Costa; H. C. Pélissier; L. B. Davin; N. G. Lewis, *J. Biol. Chem.* **2001**, *276*, 12614–12623.
296. S. G. A. Moinuddin; B. Youn; D. L. Bedgar; M. A. Costa; G. L. Helms; C. Kang; L. B. Davin; N. G. Lewis, *Org. Biomol. Chem.* **2006**, *4*, 808–816.
297. B. Youn; S. G. A. Moinuddin; L. B. Davin; N. G. Lewis; C. Kang, *J. Biol. Chem.* **2005**, *280*, 12917–12926.
298. M.-H. Cho; S. G. A. Moinuddin; G. L. Helms; S. Hishiyama; D. Eichinger; L. B. Davin; N. G. Lewis, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 10641–10646.
299. M. Okigawa; T. Maeda; N. Kawano, *Tetrahedron* **1970**, *26*, 4301–4305.
300. A. Mohagheghzadeh; T. J. Schmidt; A. W. Alfermann, *J. Nat. Prod.* **2002**, *65*, 69–71.
301. H. R. Vardapetyan; A. B. Kirakosyan; A. A. Oganesyan; A. R. Penesyan; W. A. Alfermann, *Russ. J. Plant Physiol.* **2003**, *50*, 297–300.
302. A. Koulman; B. Konuklugil, *Biochem. Syst. Ecol.* **2004**, *32*, 91–93.
303. N. Vasilev; I. Ionkova, *Pharm. Biol.* **2005**, *43*, 509–511.
304. N. P. Vasilev; I. Ionkova, *Fitoterapia* **2005**, *76*, 50–53.
305. N. Vasilev; Elfahmi; R. Bos; O. Kayser; G. Momekov; S. Konstantinov; I. Ionkova, *J. Nat. Prod.* **2006**, *69*, 1014–1017.
306. B. Konuklugil; I. Ionkova; N. Vasilev; T. J. Schmidt; J. Windhövel; E. Fuss; A. W. Alfermann, *Nat. Prod. Res.* **2007**, *21*, 1–6.
307. K. Kranz; M. Petersen, *Phytochemistry* **2003**, *64*, 453–458.
308. G. A. Molog; U. Empt; S. Kuhlmann; W. van Uden; N. Pras; A. W. Alfermann; M. Petersen, *Planta* **2001**, *214*, 288–294.
309. J. Liu; R. D. Stipanovic; A. A. Bell; L. S. Puckhaber; C. W. Magill, *Phytochemistry* **2008**, *69*, 3038–3042.
310. P. Sartorelli; P. J. C. Benevides; R. M. Ellensohn; M. V. A. F. Rocha; P. R. H. Moreno; M. J. Kato, *Plant Sci.* **2001**, *161*, 1083–1088.
311. T. Katayama; Y. Kado, *J. Wood Sci.* **1998**, *44*, 244–246.
312. N. Lourith; T. Katayama; K. Ishikawa; T. Suzuki, *J. Wood Sci.* **2005**, *51*, 379–386.
313. N. Lourith; T. Katayama; T. Suzuki, *J. Wood Sci.* **2005**, *51*, 370–378.
314. R. B. Herbert; A. E. Kattah; E. Knagg, *Tetrahedron* **1990**, *46*, 7119–7138.
315. H. Kasahara; L. B. Davin; N. G. Lewis. Aryl Propenal Double Bond Reductase. U.S. Patent 6,703,229, 9 March, 2004.
316. H. Kasahara; Y. Jiao; D. L. Bedgar; S.-J. Kim; A. M. Patten; Z.-Q. Xia; L. B. Davin; N. G. Lewis, *Phytochemistry* **2006**, *67*, 1765–1780.
317. R. A. Savidge; H. Förster, *Phytochemistry* **2001**, *57*, 1095–1103.
318. B. Youn; S.-J. Kim; S. G. A. Moinuddin; C. Lee; D. L. Bedgar; A. R. Harper; L. B. Davin; N. G. Lewis; C. Kang, *J. Biol. Chem.* **2006**, *281*, 40076–40088.
319. J. Mano; Y. Torii; S.-i. Hayashi; K. Takimoto; K. Matsui; K. Nakamura; D. Inzé; E. Babiychuk; S. Kushnir; K. Asada, *Plant Cell Physiol.* **2002**, *43*, 1445–1455.
320. J. Mano; E. Belles-Boix; E. Babiychuk; D. Inzé; Y. Torii; E. Hiraoka; K. Takimoto; L. Slooten; K. Asada; S. Kushnir, *Plant Physiol.* **2005**, *139*, 1773–1783.
321. Y. Zhang; K. H. Teoh; D. W. Reed; L. Maes; A. Goossens; D. J. H. Olson; A. R. S. Ross; P. S. Covello, *J. Biol. Chem.* **2008**, *283*, 21501–21508.
322. K. L. Ringer; M. E. McConkey; E. M. Davis; G. W. Rushing; R. Croteau, *Arch. Biochem. Biophys.* **2003**, *418*, 80–92.
323. T. Hori; T. Yokomizo; H. Ago; M. Sugahara; G. Ueno; M. Yamamoto; T. Kumasaka; T. Shimizu; M. Miyano, *J. Biol. Chem.* **2004**, *279*, 22615–22623.
324. K. Vander Mijnsbrugge; H. Beeckman; R. De Rycke; M. Van Montagu; G. Engler; W. Boerjan, *Planta* **2000**, *211*, 502–509.
325. K. Vander Mijnsbrugge; H. Meyermans; M. Van Montagu; G. Bauw; W. Boerjan, *Planta* **2000**, *210*, 589–598.
326. X. Wang; X. He; J. Lin; H. Shao; Z. Chang; R. A. Dixon, *J. Mol. Biol.* **2006**, *358*, 1341–1352.
327. T. Akashi; S. Koshimizu; T. Aoki; S.-i. Ayabe, *FEBS Lett.* **2006**, *580*, 5666–5670.
328. K. Terada; C. Honda; K. Suwa; S. Takeyama; H. Oku; W. Kamisako, *Chem. Pharm. Bull.* **1995**, *43*, 564–566.
329. S. Suzuki; T. Umezawa; M. Shimada, *J. Chem. Soc. Perkin Trans. 1* **2001**, 3252–3257.
330. S. Suzuki; T. Nakatsubo; T. Umezawa; M. Shimada, *J. Chem. Soc. Chem. Commun.* **2002**, 1088–1089.
331. T. Imai; M. Nomura; Y. Matsushita; K. Fukushima, *J. Plant Physiol.* **2006**, *163*, 1221–1228.
332. T. Imai; M. Nomura; K. Fukushima, *J. Plant Physiol.* **2006**, *163*, 483–487.
333. T. Nagasaki; S. Yasuda; T. Imai, *Phytochemistry* **2002**, *60*, 461–466.
334. T. Nagasaki; S. Yasuda; T. Imai, *Phytochemistry* **2001**, *58*, 833–840.
335. A. D. Bayley; J. van Staden, *Plant Physiol. Biochem.* **1990**, *28*, 691–695.
336. A. D. Bayley; J. van Staden, *Plant Physiol. Biochem.* **1990**, *28*, 697–702.
337. A. D. Bayley; J. van Staden, *Plant Physiol. Biochem.* **1991**, *29*, 403–408.
338. K. Terada; C. Honda; S. Takeyama; K. Suwa; W. Kamisako, *Biol. Pharm. Bull.* **1995**, *18*, 1472–1475.
339. K. Terada; K. Suwa; S. Takeyama; C. Honda; W. Kamisako, *Biol. Pharm. Bull.* **1996**, *19*, 748–751.
340. A. J. Birch; M. Slaytor, *Chem. Ind.* **1956**, 1524.
341. K. Kaneko, *Chem. Pharm. Bull.* **1960**, *8*, 611–614.
342. K. Kaneko, *Chem. Pharm. Bull.* **1960**, *8*, 875–879.
343. K. Kaneko, *Chem. Pharm. Bull.* **1961**, *9*, 108–109.
344. K. Kaneko, *Chem. Pharm. Bull.* **1962**, *10*, 1085–1087.
345. R. Tressl; F. Drawert, *J. Agric. Food Chem.* **1973**, *21*, 560–565.
346. M. Klischies; J. Stöckigt; M. H. Zenk, *J. Chem. Soc. Chem. Commun.* **1975**, 879–880.

347. A. J. Birch, Biosynthetic Pathways. In *Chemical Plant Taxonomy*; T. Swain, Ed.; Academic Press: London, UK, 1963; pp 141–166.
348. T. A. Geissman; D. H. G. Crout, Compounds Derived from Shikimic Acid. In *Organic Chemistry of Secondary Plant Metabolism*; T. A. Geissman, D. H. G. Crout, Eds.; Freeman, Cooper & Company: San Francisco, CA, USA, 1969; pp 136–166.
349. L. Canonica; P. Manitto; D. Monti; A. M. Sanchez, *J. Chem. Soc. Chem. Commun.* **1971**, 1108–1109.
350. P. Manitto; D. Monti; P. Gramatica, *J. Chem. Soc. Perkin Trans. 1* **1974**, 1727–1731.
351. U. M. Senanayake; R. B. H. Wills; T. H. Lee, *Phytochemistry* **1977**, *16*, 2032–2033.
352. D. R. Gang; J. Wang; N. Dudareva; K. H. Nam; J. E. Simon; E. Lewinsohn; E. Pichersky, *Plant Physiol.* **2001**, *125*, 539–555.
353. J. C. Verdonk; M. A. Haring; A. J. van Tunen; R. C. Schuurink, *Plant Cell* **2005**, *17*, 1612–1624.
354. T. Koeduka; G. V. Louie; I. Orlova; C. M. Kish; M. Ibdah; C. G. Wilkerson; M. E. Bowman; T. J. Baiga; J. P. Noel; N. Dudareva; E. Pichersky, *Plant J.* **2008**, *54*, 362–374.
355. G. V. Louie; T. J. Baiga; M. E. Bowman; T. Koeduka; J. H. Taylor; S. M. Spassova; E. Pichersky; J. P. Noel, *PLoS ONE* **2007**, *2* (10), e993.
356. T. Koeduka; T. J. Baiga; J. P. Noel; E. Pichersky, *Plant Physiol.* **2009**, *149*, 384–394.
357. S.-J. Kim; M.-R. Kim; D. L. Bedgar; S. G. A. Moinuddin; C. L. Cardenas; L. B. Davin; C. Kang; N. G. Lewis, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1455–1460.
358. M. Jourdes; C. L. Cardenas; D. D. Laskar; S. G. A. Moinuddin; L. B. Davin; N. G. Lewis, *Phytochemistry* **2007**, *68*, 1932–1956.
359. D. M. Tieman; H. M. Loucas; J. Y. Kim; D. G. Clark; H. J. Klee, *Phytochemistry* **2007**, *68*, 2660–2669.
360. I. Damiani; K. Morreel; S. Danoun; G. Goeminne; N. Yahiaoui; C. Marque; J. Kopka; E. Messens; D. Goffner; W. Boerjan; A.-M. Boudet; S. Rochange, *Plant Mol. Biol.* **2005**, *59*, 753–769.
361. N. L. Paiva; R. Edwards; Y. Sun; G. Hrazdina; R. A. Dixon, *Plant Mol. Biol.* **1991**, *17*, 653–667.
362. N. L. Paiva; Y. Sun; R. A. Dixon; H. D. VanEtten; G. Hrazdina, *Arch. Biochem. Biophys.* **1994**, *312*, 501–510.
363. N. Oyama-Okubo; T. Ando; N. Watanabe; E. Marchesi; K. Uchida; M. Nakayama, *Biosci. Biotechnol. Biochem.* **2005**, *69*, 773–777.
364. J. Wang; N. Dudareva; S. Bhakta; R. A. Raguso; E. Pichersky, *Plant Physiol.* **1997**, *114*, 213–221.
365. J. Wang; E. Pichersky, *Arch. Biochem. Biophys.* **1998**, *349*, 153–160.
366. E. Lewinsohn; I. Ziv-Raz; N. Dudai; Y. Tadmor; E. Lastochkin; O. Larkov; D. Chaimovitsh; U. Ravid; E. Putievsky; E. Pichersky; Y. Shoham, *Plant Sci.* **2000**, *160*, 27–35.
367. D. R. Gang; N. Lavid; C. Zubieta; F. Chen; T. Beuerle; E. Lewinsohn; J. P. Noel; E. Pichersky, *Plant Cell* **2002**, *14*, 505–519.
368. M. Varbanova; I. Orlova; N. Dudareva; D. Reichhard; E. Pichersky. A cytochrome P450 enzyme from *Illicium parviflorum* capable of catalyzing the formation of a methylenedioxy bridge on eugenol to produce safrole. Phytochemical Society of North America Annual Meeting Abstract, St. Louis, MO, 2007. (http://www.psna-online.org/newsletters/PSNAN 461.web.pdf)
369. R. Martin; G. Schilling; J. Reichling, *Z. Naturforsch. C. Biosci.* **1988**, *43*, 328–336.
370. J. Reichling; R. Martin; U. Thron, *Z. Naturforsch. C. Biosci.* **1988**, *43*, 42–46.
371. J. Reichling; R. Martin, *Pharm. Ztg. Wiss.* **1991**, *4*, 225–231.
372. R. Dexter; A. Qualley; C. M. Kish; C. J. Ma; T. Koeduka; D. A. Nagegowda; N. Dudareva; E. Pichersky; D. Clark, *Plant J.* **2007**, *49*, 265–275.
373. B. St-Pierre; V. De Luca, Evolution of Acyltransferase Genes: Origin and Diversification of the BAHD Superfamily of Acyltransferases Involved in Secondary Metabolism. In *Recent Advances in Phytochemistry*, Elsevier: Oxford, UK, 2000; Vol. 34, Evolution of Metabolic Pathways, J. T. Romeo, R., Ibrahim, L. Varin, V. De Luca, Eds.; pp 285–315.
374. B. Harrison; D. R. Gang Characterization of coniferyl acetate acetyl transferase from sweet basil (Ocimum basilicum). In *The 19th Rocky Mountain Regional Meeting*, Tucson, AZ, 2006 (http://acs/confex.com/acs/rmrm06/tech program/P3334 HTM).
375. P. W. Sherman; G. A. Hash, *Evol. Human Behav.* **2001**, *22*, 147–163.
376. A. O. Gill; R. A. Holley, *Appl. Envir. Microbiol.* **2004**, *70*, 5750–5755.
377. K. Rhayour; T. Bouchiki; A. Tantaoui-Elaraki; K. Sendide; A. Remmal, *J. Essent. Oil Res.* **2003**, *15*, 356–362.
378. A. O. Gill; R. A. Holley, *Int. J. Food Microbiol.* **2006**, *108*, 1–9.
379. M. Greenberg; M. Dodds; M. Tian, *J. Agric. Food. Chem.* **2008**, *56*, 11151–11156.
380. S. E. Walsh; J.-Y. Maillard; A. D. Russell; C. E. Catrenich; D. L. Charbonneau; R. G. Bartolo, *J. Appl. Microbiol.* **2003**, *94*, 240–247.
381. N. A. Olasupo; D. J. Fitzgerald; A. Narbad; M. J. Gasson, *J. Food Prot.* **2004**, *67*, 596–600.
382. K. Voda; B. Boh; M. Vrtačnik; F. Pohleven, *Int. Biodet. Biodegr.* **2003**, *51*, 51–59.
383. U. Dev; C. Devakumar; J. Mohan; P. C. Agarwal, *J. Essent. Oil Res.* **2004**, *16*, 496–499.
384. S. Bennis; F. Chami; N. Chami; T. Bouchikhi; A. Remmal, *Lett. Appl. Microbiol.* **2004**, *38*, 454–458.
385. N. Chami; F. Chami; S. Bennis; J. Trouillas; A. Remmal, *Braz. J. Infect. Dis.* **2004**, *8*, 217–226.
386. E. W. Skinner, *The Science of Dental Materials*; W.B. Saunders Company: Philadelphia, PA, USA, 1940.
387. J. E. Weinberg; J. L. Rabinowitz; M. Zanger; A. R. Gennaro, *J. Dent. Res.* **1972**, *51*, 1055–1061.
388. J. Grush; D. L. G. Noakes; R. D. Moccia, *Zebrafish* **2004**, *1*, 46–53.
389. S. A. Guénette; P. Hélie; F. Beaudry; P. Vachon, *Vet. Anaesth. Analg.* **2007**, *34*, 164–170.
390. K. K. Sladky; C. R. Swanson; M. K. Stoskopf; M. R. Loomis; G. A. Lewbart, *Am. J. Vet. Res.* **2001**, *62*, 337–342.
391. S. A. Guénette; F. Béaudry; J. F. Marier; P. Vachon, *J. Vet. Pharmacol. Therap.* **2006**, *29*, 265–270.
392. S. A. Guénette; F. C. Uhland; P. Hélie; F. Beaudry; P. Vachon, *Aquaculture* **2007**, *266*, 262–265.
393. M. Hanusz, A Century of Kretek. In *Smoke: A Global History of Smoking*; S. L. Gilman, Z. Xun, Eds.; Reaktion Books: London, UK, 2004; pp 140–143.
394. M. H. Lee; K.-Y. Yeon; C.-K. Park; H.-Y. Li; Z. Fang; M. S. Kim; S.-Y. Choi; S. J. Lee; S. Lee; K. Park; J.-H. Lee; J. S. Kim; S. B. Oh, *J. Dent. Res.* **2005**, *84*, 848–851.
395. C.-K. Park; H. Y. Li; K.-Y. Yeon; S. J. Jung; S.-Y. Choi; S. J. Lee; S. Lee; K. Park; J. S. Kim; S. B. Oh, *J. Dent. Res.* **2006**, *85*, 900–904.
396. J. S. Cho; T. H. Kim; J.-M. Lim; J.-H. Song, *Brain Res.* **2008**, *1243*, 53–62.
397. L. M. Pessoa; S. M. Morais; C. M. L. Bevilaqua; J. H. S. Luciano, *Vet. Parasitol.* **2002**, *109*, 59–63.

398. L. F. L. Interaminense; J. H. Leal-Cardoso; P. J. C. Magalhaes; G. P. Duarte; S. Lahlou, *Planta Med.* **2005**, *71*, 376–378.
399. T. Ueda-Nakamura; R. R. Mendonça-Filho; J. A. Morgado-Díaz; P. K. Maza; B. P. D. Filho; D. A. G. Cortez; D. S. Alviano; M. D. S. S. Rosa; A. H. C. S. Lopes; C. S. Alviano; C. V. Nakamura, *Parasitol. Int.* **2006**, *55*, 99–105.
400. H. Xu; M. Delling; J. C. Jun; D. E. Clapham, *Nat. Neurosci.* **2006**, *9*, 628–635.
401. L. F. L. Interaminense; D. M. Jucá; P. J. C. Magalhães; J. H. Leal-Cardoso; G. P. Duarte; S. Lahlou, *Fundam. Clin. Pharmacol.* **2007**, *21*, 497–506.
402. H. Y. Li; C.-K. Park; S. J. Jung; S.-Y. Choi; S. J. Lee; K. Park; J. S. Kim; S. B. Oh, *J. Dent. Res.* **2007**, *86*, 898–902.
403. G. Chung; J. N. Rhee; S. J. Jung; J. S. Kim; S. B. Oh, *J. Dent. Res.* **2008**, *87*, 137–141.
404. G. Tao; Y. Irie; D.-J. Li; W. M. Keung, *Bioorg. Med. Chem.* **2005**, *13*, 4777–4788.
405. H. Raghavenra; B. T. Diwakr; B. R. Lokesh; K. A. Naidu, *Prostaglandins Leukot. Essent. Fatty Acids* **2006**, *74*, 23–27.
406. W. Li; R. Tsubouchi; S. Qiao; M. Haneda; K. Murakami; M. Yoshino, *Biomed. Res.* **2006**, *27*, 69–74.
407. B. B. Aggarwal; S. Shishodia, *Biochem. Pharm.* **2006**, *71*, 1397–1421.
408. A. S. L. Chan; H. Pang; E. C. H. Yip; Y. K. Tam; Y. H. Wong, *Planta Med.* **2005**, *71*, 634–639.
409. H. Yokota; H. Hashimoto; M. Motoya; A. Yuasa, *Biochem. Pharmacol.* **1988**, *37*, 799–802.
410. C.-B. Yoo; K.-T. Han; K.-S. Cho; J. Ha; H.-J. Park; J.-H. Nam; U.-H. Kil; K.-T. Lee, *Cancer Lett.* **2005**, *225*, 41–52.
411. B. S. Park; Y. S. Song; S. B. Yee; B. G. Lee; S. Y. Seo; Y. C. Park; J. M. Kim; H. M. Kim; Y. H. Yoo, *Apoptosis* **2005**, *10*, 193–200.
412. R. Ghosh; N. Nadiminty; J. E. Fitzpatrick; W. L. Alworth; T. J. Slaga; A. P. Kumar, *J. Biol. Chem.* **2005**, *280*, 5812–5819.
413. G. B. N. Chainy; S. K. Manna; M. M. Chaturvedi; B. B. Aggarwal, *Oncogene* **2000**, *19*, 2943–2950.
414. C. Y. Choi; K.-R. Park; J.-H. Lee; Y. J. Jeon; K.-H. Liu; S. Oh; D.-E. Kim; S. S. Yea, *Eur. J. Pharmacol.* **2007**, *576*, 151–159.
415. Z. S. Zhao; P. J. O'Brien, *Toxicol. Appl. Pharmacol.* **1996**, *140*, 411–421.
416. D.-W. Tang; K.-W. Chang; C.-W. Chi; T.-Y. Liu, *Toxicol. Lett.* **2004**, *152*, 235–243.
417. J. Zhang; G. Feng; Z.-Q. Ma; J.-T. Feng; X. Zhang, *Kunchong Xuebao* **2007**, *50*, 574–577.
418. F. Zhang; Q. Xu; S. Fu; X. Ma; H. Xiao; X. Liang, *Flav. Fragr. J.* **2005**, *20*, 318–320.
419. Z. Ma; J. He; Z. Jiang; Z. Wang; J. Feng; X. Zhang, Application of asaricin in control of plant pathogenic bacteria-induced plant diseases. China Patent Number 101027990, 2007.
420. C. H. Chuah; H. S. Yong; S. H. Goh, *Biochem. Syst. Ecol.* **1997**, *25*, 391–393.
421. E. Jang; D. McInnis; R. Vargas; R. Mau, Area-wide integrated pest management (IPM) of fruit flies in Hawaiian fruits and vegetables. http://www.ars.usda.gov/research/publications/publications.htm?SEQ_NO_115=167070
422. T. E. Shelly; E. Pahio; J. Edu, *Fla. Entomol.* **2004**, *87*, 481–486.
423. U. S. E. P. Agency. Pesticides: Regulating Pesticides. Methyl Eugenol (ME). http://www.epa.gov/oppbppd1/biopesticides/ingredients/factsheets/factsheet_203900.htm
424. D. M. Suckling; E. B. Jang; L. A. Carvalho; J. T. Nagata; E. L. Schneider; A. M. El-Sayed, *J. Chem. Ecol.* **2007**, *33*, 1494–1504.
425. R. L. Smith; T. B. Adams; J. Doull; V. J. Feron; J. I. Goodman; L. J. Marnett; P. S. Portoghese; W. J. Waddell; B. M. Wagner; A. E. Rogers; J. Caldwell; I. G. Sipes, *Food Chem. Toxicol.* **2002**, *40*, 851–870.
426. I. U. Fischer; G. E. von Unruh; H. J. Dengler, *Xenobiotica* **1990**, *20*, 209–222.
427. I. Gardner; H. Wakazono; P. Bergin; I. de Waziers; P. Beaune; J. G. Kenna; J. Caldwell, *Carcinogenesis* **1997**, *18*, 1775–1783.
428. S. M. F. Jeurissen; J. J. P. Bogaards; H. M. Awad; M. G. Boersma; W. Brand; Y. C. Fiamegos; T. A. van Beek; G. M. Alink; E. J. R. Sudhölter; N. H. P. Cnubben; I. M. C. M. Rietjens, *Chem. Res. Toxicol.* **2004**, *17*, 1245–1250.
429. I. M. C. M. Rietjens; M. G. Boersma; H. van der Woude; S. M. F. Jeurissen; M. E. Schutte; G. M. Alink, *Mutat. Res.* **2005**, *574*, 124–138.
430. S. M. F. Jeurissen; J. J. P. Bogaards; M. G. Boersma; J. P. F. ter Horst; H. M. Awad; Y. C. Fiamegos; T. A. van Beek; G. M. Alink; E. J. R. Sudhölter; N. H. P. Cnubben; I. M. C. M. Rietjens, *Chem. Res. Toxicol.* **2006**, *19*, 111–116.
431. S. M. F. Jeurissen; A. Punt; M. G. Boersma; J. J. P. Bogaards; Y. C. Fiamegos; B. Schilter; P. J. van Bladeren; N. H. P. Cnubben; I. M. C. M. Rietjens, *Chem. Res. Toxicol.* **2007**, *20*, 798–806.
432. H. Wakazono; I. Gardner; E. Eliasson; M. W. H. Coughtrie; J. G. Kenna; J. Caldwell, *Chem. Res. Toxicol.* **1998**, *11*, 863–872.
433. G.-D. Zhou; B. Moorthy; J. Bi; K. C. Donnelly; K. Randerath, *Envir. Mol. Mutag.* **2007**, *48*, 715–721.
434. M. H. G. Medeiros; P. Di Mascio; A. P. Pinto; R. R. Vargas; E. J. H. Bechara, *Free Radical Res.* **1996**, *25*, 5–12.
435. I. Gardner; P. Bergin; P. Stening; J. G. Kenna; J. Caldwell, *Chem. Res. Toxicol.* **1996**, *9*, 713–721.
436. R.-S. Tsai; P.-A. Carrupt; B. Testa; J. Caldwell, *Chem. Res. Toxicol.* **1994**, *7*, 73–76.
437. S. Yamada; Y. Murawaki; H. Kawasaki, *Biochem. Pharmacol.* **1993**, *46*, 1081–1085.
438. M. Nomura; Y. Ohtaki; T. Hida; T. Aizawa; H. Wakita; K.-I. Miyamoto, *Anticancer Res.* **1994**, *14*, 1967–1972.
439. S.-H. Kim; Y. S. Kim; S. S. Kang; K. Bae; T. M. Hung; S.-M. Lee, *J. Pharmacol. Sci.* **2008**, *106*, 225–233.
440. S. Fukumitsu; K. Aida; N. Ueno; S. Ozawa; Y. Takahashi; M. Kobori, *Br. J. Nutr.* **2008**, *100*, 669–676.
441. C. Canel; R. M. Moraes; F. E. Dayan; D. Ferreira, *Phytochemistry* **2000**, *54*, 115–120.
442. M. Gordaliza; P. A. García; J. M. Miguel del Corral; M. A. Castro; M. A. Gómez-Zurita, *Toxicon* **2004**, *44*, 441–459.
443. S. Apers; A. Vlietinck; L. Pieters, *Phytochem. Rev.* **2003**, *2*, 201–217.
444. V. Srivastava; A. S. Negi; J. K. Kumar; M. M. Gupta; S. P. S. Khanuja, *Bioorg. Med. Chem.* **2005**, *13*, 5892–5908.
445. A. Montecucco; G. Biamonti, *Cancer Lett.* **2007**, *252*, 9–18.
446. K. R. Hande, *Eur. J. Cancer* **1998**, *34*, 1514–1521.
447. M. A. Economou; S. Andersson; D. Vasilcanu; C. All-Ericsson; E. Menu; A. Girnita; L. Girnita; M. Axelson; S. Seregard; O. Larsson, *Invest. Ophthalmol. Vis. Sci.* **2008**, *49*, 2337–2342.
448. M. A. Economou; J. Wu; D. Vasilcanu; L. Rosengren; C. All-Ericsson; I. van der Ploeg; E. Menu; L. Girnita; M. Axelson; O. Larsson; S. Seregard; A. Kvanta, *Invest. Ophthalmol. Vis. Sci.* **2008**, *49*, 2620–2626.
449. J.-M. Barret; A. Kruczynski; C. Etiévant; B. T. Hill, *Cancer Chemother. Pharmacol.* **2002**, *49*, 479–486.
450. J. M. Sargent; A. W. Elgie; C. J. Williamson; B. T. Hill, *Anti-Cancer Drugs* **2003**, *14*, 467–473.
451. J. Kluza; R. Mazinghien; H. Irwin; J. A. Hartley; C. Bailly, *Anti-Cancer Drugs* **2006**, *17*, 155–164.
452. M. S. Butler, *Nat. Prod. Rep.* **2008**, *25*, 475–516.
453. C. Canel; F. E. Dayan; M. Ganzera; I. A. Khan; A. Rimando; C. L. Burandt, Jr.; R. M. Moraes, *Planta Med.* **2001**, *67*, 97–99.
454. E. Fuss, *Phytochem. Rev.* **2003**, *2*, 307–320.

455. R. Verpoorte; R. van der Heijden; H. J. G. ten Hoopen; J. Memelink, *Biotechnol. Lett.* **1999**, *21*, 467–479.
456. J. L. Billinsky; E. S. Krol, *J. Nat. Prod.* **2008**, *71*, 1612–1615.
457. J. L. Billinsky; M. R. Marcoux; E. S. Krol, *Chem. Res. Toxicol.* **2007**, *20*, 1352–1358.
458. J. D. Lambert; R. T. Dorr; B. N. Timmermann, *Pharm. Biol.* **2004**, *42*, 149–158.
459. Y. Cui; C. Lu; L. Liu; D. Sun; N. Yao; S. Tan; S. Bai; X. Ma, *Life Sci.* **2008**, *82*, 247–255.
460. T. Yoshida; T. Shiraishi; M. Horinaka; S. Nakata; T. Yasuda; A. E. Goda; M. Wakada; Y. Mizutani; T. Miki; A. Nishikawa; T. Sakai, *Cancer Sci.* **2007**, *98*, 1417–1423.
461. J. F. Youngren; K. Gable; C. Penaranda; B. A. Maddux; M. Zavodovskaya; M. Lobo; M. Campbell; J. Kerner; I. D. Goldfine, *Breast Cancer Res. Treat.* **2005**, *94*, 37–46.
462. J. E. Blecha; M. O. Anderson; J. M. Chow; C. C. Guevarra; C. Pender; C. Penaranda; M. Zavodovskaya; J. F. Youngren; C. E. Berkman, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 4026–4029.
463. G. E. Meyer; L. Chesler; D. Liu; K. Gable; B. A. Maddux; D. D. Goldenberg; J. F. Youngren; I. D. Goldfine; W. A. Weiss; K. K. Matthay; S. M. Rosenthal, *J. Cell. Biochem.* **2007**, *102*, 1529–1541.
464. M. Zavodovskaya; M. J. Campbell; B. A. Maddux; L. Shiry; G. Allan; L. Hodges; P. Kushner; J. A. Kerner; J. F. Youngren; I. D. Goldfine, *J. Cell. Biochem.* **2008**, *103*, 624–635.
465. M. West; M. Mhatre; A. Ceballos; R. A. Floyd; P. Grammas; S. P. Gabbita; L. Hamdheydari; T. Mai; S. Mou; Q. N. Pye; C. Stewart; S. West; K. S. Williamson; F. Zemlan; K. Hensley, *J. Neurochem.* **2004**, *91*, 133–143.
466. R. K. Mandraju; A. K. Kondapi, *Arch. Biochem. Biophys.* **2007**, *461*, 40–49.
467. B. Hausott; H. Greger; B. Marian, *J. Cancer Res. Clin. Oncol.* **2003**, *129*, 569–576.
468. D. L. Rowe; T. Ozbay; L. M. Bender; R. Nahta, *Mol. Cancer Ther.* **2008**, *7*, 1900–1908.
469. C. Plaza; M. Pavani; M. Faundez; J. D. Maya; A. Morello; M. I. Becker; A. De Ioannes; M. A. Cumsille; J. Ferreira, *In Vivo* **2008**, *22*, 353–361.
470. J. D. Lambert; R. O. Meyers; B. N. Timmermann; R. T. Dorr, *Cancer Lett.* **2001**, *171*, 47–56.
471. J. D. Heller; J. Kuo; T. C. Wu; W. M. Kast; R. C. C. Huang, *Cancer Res.* **2001**, *61*, 5499–5504.
472. R. A. Lopez; A. B. Goodman; M. Rhodes; J. A. L. Blomberg; J. Heller, *Anti-Cancer Drugs* **2007**, *18*, 933–939.
473. P. Smolewski, *IDrugs* **2008**, *11*, 204–214.
474. B. Raffaelli; A. Hoikkala; E. Leppälä; K. Wähälä, *J. Chromatog. B* **2002**, *777*, 29–43.
475. L.-Q. Wang, *J. Chromatog. B* **2002**, *777*, 289–309.
476. L. U. Thompson; P. Robb; M. Serraino; F. Cheung, *Nutr. Cancer* **1991**, *16*, 43–52.
477. H. Qu; R. L. Madl; D. J. Takemoto; R. C. Baybutt; W. Wang, *J. Nutr.* **2005**, *135*, 598–602.
478. G. Hallmans; J.-X. Zhang; E. Lundin; P. Stattin; A. Johansson; I. Johansson; K. Hultén; A. Winkvist; P. Lenner; P. Åman; H. Adlercreutz, *Proc. Nutr. Soc.* **2003**, *62*, 193–199.
479. D. Fuchs; R. Piller; J. Linseisen; H. Daniel; U. Wenzel, *Proteomics* **2007**, *7*, 3278–3288.
480. X. Lin; B. R. Switzer; W. Demark-Wahnefried, *Anticancer Res.* **2001**, *21*, 3995–3999.
481. N. Danbara; T. Yuri; M. Tsujita-Kyutoku; R. Tsukamoto; N. Uehara; A. Tsubura, *Anticancer Res.* **2005**, *25*, 2269–2276.
482. M. J. McCann; C. I. R. Gill; T. Linton; D. Berrar; H. McGlynn; I. R. Rowland, *Mol. Nutr. Food Res.* **2008**, *52*, 567–580.
483. C. L. Heald; M. R. Ritchie; C. Bolton-Smith; M. S. Morton; F. E. Alexander, *Br. J. Nutr.* **2007**, *98*, 388–396.
484. A. Kuijsten; I. C. W. Arts; P. C. H. Hollman; P. van't Veer; E. Kampman, *Cancer Epidemiol. Biomarkers Prev.* **2006**, *15*, 1132–1136.
485. A. Kuijsten; P. C. H. Hollman; H. C. Boshuizen; M. N. C. P. Buijsman; P. van 't Veer; F. J. Kok; I. C. W. Arts; H. B. Bueno-de-Mesquita, *Am. J. Epidemiol.* **2008**, *167*, 734–742.
486. S. Oikarinen; S.-M. Heinonen; T. Nurmi; H. Adlercreutz; M. Mutanen, *Eur. J. Nutr.* **2005**, *44*, 273–280.
487. A. Kuijsten; H. B. Bueno-de-Mesquita; J. M. A. Boer; I. C. W. Arts; F. J. Kok; P. van't Veer; P. C. H. Hollman, *Atherosclerosis* **2009**, *203*, 145–152.
488. B. Y. Hernandez; K. McDuffie; A. A. Franke; J. Killeen; M. T. Goodman, *Nutr. Cancer* **2004**, *49*, 109–124.
489. D. Miura; N. M. Saarinen; Y. Miura; R. Santti; K. Yagasaki, *Nutr. Cancer* **2007**, *58*, 49–59.
490. J. D. Brooks; L. U. Thompson, *J. Steroid Biochem. Mol. Biol.* **2005**, *94*, 461–467.
491. M. Cosentino; F. Marino; M. Ferrari; E. Rasini; R. Bombelli; A. Luini; M. Legnaro; M. G. Delle Canne; M. Luzzani; F. Crema; S. Paracchini; S. Lecchini, *Pharmacol. Res.* **2007**, *56*, 140–147.
492. P. Penttinen; J. Jaehrling; A. E. Damdimopoulos; J. Inzunza; J. G. Lemmen; P. van der Saag; K. Pettersson; G. Gauglitz; S. Mäkelä; I. Pongratz, *Endocrinology* **2007**, *148*, 4875–4886.
493. C. Carreau; G. Flouriot; C. Bennetau-Pelissero; M. Potier, *J. Steroid Biochem. Mol. Biol.* **2008**, *110*, 176–185.
494. J. Chen; L. U. Thompson, *Breast Cancer Res. Treat.* **2003**, *80*, 163–170.
495. R. Piller; J. Chang-Claude; J. Linseisen, *Eur. J. Cancer Prev.* **2006**, *15*, 225–232.
496. N. M. Saarinen; A. Wärri; M. Airio; A. Smeds; S. Mäkelä, *Mol. Nutr. Food Res.* **2007**, *51*, 857–866.
497. M. S. Touillaud; A. C. M. Thiébaut; A. Fournier; M. Niravong; M.-C. Boutron-Ruault; F. Clavel-Chapelon, *J. Natl. Cancer Inst.* **2007**, *99*, 475–486.
498. K. A. Power; L. U. Thompson, *Mol. Nutr. Food Res.* **2007**, *51*, 845–856.
499. A. Olsen; K. E. B. Knudsen; B. L. Thomsen; S. Loft; C. Stripp; K. Overvad; S. Møller; A. Tjønneland, *Cancer Epidemiol. Biomarkers Prev.* **2004**, *13*, 2084–2089.
500. G. Khan; P. Penttinen; A. Cabanes; A. Foxworth; A. Chezek; K. Mastropole; B. Yu; A. Smeds; T. Halttunen; C. Good; S. Mäkelä; L. Hilakivi-Clarke, *Reprod. Toxicol.* **2007**, *23*, 397–406.
501. S. Heinonen; T. Nurmi; K. Liukkonen; K. Poutanen; K. Wähälä; T. Deyama; S. Nishibe; H. Adlercreutz, *J. Agric. Food Chem.* **2001**, *49*, 3178–3186.
502. A. N. Begum; C. Nicolle; I. Mila; C. Lapierre; K. Nagano; K. Fukushima; S.-M. Heinonen; H. Adlercreutz; C. Rémésy; A. Scalbert, *J. Nutr.* **2004**, *134*, 120–127.
503. J. L. Peñalvo; S.-M. Heinonen; A.-M. Aura; H. Adlercreutz, *J. Nutr.* **2005**, *135*, 1056–1062.
504. L.-Q. Wang; M. R. Meselhy; Y. Li; G.-W. Qin; M. Hattori, *Chem. Pharm. Bull.* **2000**, *48*, 1606–1610.

505. E. Eeckhaut; K. Struijs; S. Possemiers; J.-P. Vincken; D. De Keukeleire; W. Verstraete, *J. Agric. Food Chem.* **2008**, *56*, 4806–4812.
506. L.-H. Xie; T. Akao; K. Hamasaki; T. Deyama; M. Hattori, *Chem. Pharm. Bull.* **2003**, *51*, 508–515.
507. T. Clavel; G. Henderson; W. Engst; J. Doré; M. Blaut, *FEMS Microbiol. Ecol.* **2006**, *55*, 471–478.
508. J.-S. Jin; N. Kakiuchi; M. Hattori, *Biol. Pharm. Bull.* **2007**, *30*, 2204–2206.
509. J.-S. Jin; Y.-F. Zhao; N. Nakamura; T. Akao; N. Kakiuchi; M. Hattori, *Biol. Pharm. Bull.* **2007**, *30*, 904–911.
510. J.-S. Jin; Y.-F. Zhao; N. Nakamura; T. Akao; N. Kakiuchi; B.-S. Min; M. Hattori, *Biol. Pharm. Bull.* **2007**, *30*, 2113–2119.
511. G. H. E. Jansen; I. C. W. Arts; M. W. F. Nielen; M. Müller; P. C. H. Hollman; J. Keijer, *Arch. Biochem. Biophys.* **2005**, *435*, 74–82.
512. A. Kuijsten; I. C. W. Arts; T. B. Vree; P. C. H. Hollman, *J. Nutr.* **2005**, *135*, 795–801.
513. C. Hu; Y. V. Yuan; D. D. Kitts, *Food Chem. Toxicol.* **2007**, *45*, 2219–2227.
514. W. Zhang; X. Wang; Y. Liu; H. Tian; B. Flickinger; M. W. Empie; S. Z. Sun, *Br. J. Nutr.* **2008**, *99*, 1301–1309.
515. W. Zhang; X. Wang; Y. Liu; H. Tian; B. Flickinger; M. W. Empie; S. Z. Sun, *J. Med. Food* **2008**, *11*, 207–214.
516. A. Pan; J. Sun; Y. Chen; X. Ye; H. Li; Z. Yu; Y. Wang; W. Gu; X. Zhang; X. Chen; W. Demark-Wahnefried; Y. Liu; X. Lin, *PLoS ONE* **2007**, *2*, e1148.
517. F. Chen; T. Wang; Y.-F. Wu; Y. Gu; X.-L. Xu; S. Zheng; X. Hu, *World J. Gastroenterol.* **2004**, *10*, 3459–3463.
518. S. Dikalov; T. Losik; J. L. Arbiser, *Biochem. Pharmacol.* **2008**, *76*, 589–596.
519. A. Garcia; Y. Zheng; C. Zhao; A. Toschi; J. Fan; N. Shraibman; H. A. Brown; D. Bar-Sagi; D. A. Foster; J. L. Arbiser, *Clin. Cancer Res.* **2008**, *14*, 4267–4274.
520. E.-R. Hahm; J. A. Arlotti; S. W. Marynowski; S. V. Singh, *Clin. Cancer Res.* **2008**, *14*, 1248–1257.
521. S.-K. Lee; H.-N. Kim; Y.-R. Kang; C. W. Lee; H.-M. Kim; D. C. Han; J. Shin; K. Bae; B.-M. Kwon, *Bioorg. Med. Chem.* **2008**, *16*, 8397–8402.
522. H. Liu; C. Zang; A. Emde; M. D. Planas-Silva; M. Rosche; A. Kühnl; C.-O. Schulz; E. Elstner; K. Possinger; J. Eucker, *Eur. J. Pharmacol.* **2008**, *591*, 43–51.
523. H. Luo; Q. Zhong; L.-j. Chen; X.-r. Qi; A.-f. Fu; H.-s. Yang; F. Yang; H.-g. Lin; Y.-q. Wei; X. Zhao, *J. Cancer Res. Clin. Oncol.* **2008**, *134*, 937–945.
524. S. M. Raja; S. Chen; P. Yue; T. M. Acker; B. Lefkove; J. L. Arbiser; F. R. Khuri; S.-Y. Sun, *Mol. Cancer Ther.* **2008**, *7*, 2212–2223.
525. C. F. Albrecht; E. J. Theron; P. B. Kruger, *S. Afr. Med. J.* **1995**, *85*, 853–860.
526. C. F. Albrecht; P. B. Kruger; B. J. Smit; M. Freestone; L. Gouws; R. Miller; P. P. van Jaarsveld, *S. Afr. Med. J.* **1995**, *85*, 861–865.
527. V. D. P. Nair; A. Dairam; A. Agbonon; J. T. Arnason; B. C. Foster; I. Kanfer, *J. Agric. Food Chem.* **2007**, *55*, 1707–1711.
528. E. J. Theron; C. F. Albrecht; P. B. Kruger; K. Jenkins; M. J. van der Merwe, *In Vitro Cell. Dev. Biol.* **1994**, *30A*, 115–119.
529. V. D. P. Nair; I. Kanfer, *Phytochem. Anal.* **2007**, *18*, 475–483.
530. P. Vinesi; M. Serafini; M. Nicoletti; L. Spanò; P. Betto, *J. Nat. Prod.* **1990**, *53*, 196–199.
531. J. R. Hwu; W. N. Tseng; J. Gnabre; P. Giza; R. C. C. Huang, *J. Med. Chem.* **1998**, *41*, 2994–3000.
532. R. C. C. Huang; Y. Li; P. E. Giza; J. N. Gnabre; I. S. Abd-Elazem; K. Y. King; J. R. Hwu, *Antivir. Res.* **2003**, *58*, 57–64.
533. R. Park; P. E. Giza; D. E. Mold; R. C. C. Huang, *Antivir. Res.* **2003**, *58*, 35–45.
534. J. Craigo; M. Callahan; R. C. C. Huang; A. L. DeLucia, *Antivir. Res.* **2000**, *47*, 19–28.
535. K. L. Allen; D. R. Tschantz; K. S. Awad; W. P. Lynch; A. L. DeLucia, *Mol. Carcinog.* **2007**, *46*, 564–575.
536. N. Khanna; R. Dalby; M. Tan; S. Arnold; J. Stern; N. Frazer, *Gyn. Oncol.* **2007**, *107*, 554–562.
537. N. Khanna; R. Dalby; A. Connor; A. Church; J. Stern; N. Frazer, *Sex. Transm. Dis.* **2008**, *35*, 577–582.
538. K. Yamashita, *J. Clin. Biochem. Nutr.* **2004**, *35*, 17–28.
539. S. Shimizu; K. Akimoto; Y. Shinmen; H. Kawashima; M. Sugano; H. Yamada, *Lipids* **1991**, *26*, 512–516.
540. T. Ide; L. Ashakumary; Y. Takahashi; M. Kushiro; N. Fukuda; M. Sugano, *Biochim. Biophys. Acta* **2001**, *1534*, 1–13.
541. T. Ide; D. D. Hong; P. Ranasinghe; Y. Takahashi; M. Kushiro; M. Sugano, *Biochim. Biophys. Acta* **2004**, *1682*, 80–91.
542. F. Hirata; K. Fujita; Y. Ishikura; K. Hosoda; T. Ishikawa; H. Nakamura, *Atherosclerosis* **1996**, *122*, 135–136.
543. C.-C. Lee; P.-R. Chen; S. Lin; S.-C. Tsai; B.-W. Wang; W.-W. Chen; C. E. Tsai; K.-G. Shyu, *J. Hypertens.* **2004**, *22*, 2329–2338.
544. T. Hayashi; R. H. Thomson, *Phytochemistry* **1975**, *14*, 1085–1087.
545. R. C. C. Martins; L. R. Latorre; P. Sartorelli; M. J. Kato, *Phytochemistry* **2000**, *55*, 843–846.
546. D. C. Chauret; C. B. Bernard; J. T. Arnason; T. Durst; H. G. Krishnamurty; P. Sanchez-Vindas; N. Moreno; L. S. Roman; L. Poveda, *J. Nat. Prod.* **1996**, *59*, 152–155.
547. R. C. C. Martins; J. H. G. Lago; S. Albuquerque; M. J. Kato, *Phytochemistry* **2003**, *64*, 667–670.
548. A. A. a Silva Filho; S. Albuquerque; M. L. A. Silva; M. N. Eberlin; D. M. Tomazela; J. K. Bastos, *J. Nat. Prod.* **2004**, *67*, 42–45.
549. P. S. Luize; T. Ueda-Nakamura; B. P. D. Filho; D. A. G. Cortez; C. V. Nakamura, *Biol. Pharm. Bull.* **2006**, *29*, 2126–2130.
550. E. Kadowaki; Y. Yoshida; T. Nitoda; N. Baba; S. Nakajima, *Biosci. Biotech. Biochem.* **2003**, *67*, 415–419.
551. S. D. Elakovich; K. L. Stevens, *J. Chem. Ecol.* **1985**, *11*, 27–33.
552. N. Sakurai; K. Shibata; S. Kamisaka, *Plant Cell Physiol.* **1974**, *15*, 709–716.
553. K. Shibata; T. Kubota; S. Kamisaka, *Plant Cell Physiol.* **1974**, *15*, 191–194.
554. N. Sakurai; K. Shibata; S. Kamisaka, *Plant Cell Physiol.* **1975**, *16*, 845–855.
555. K. Shibata; T. Kubota; S. Kamisaka, *Plant Cell Physiol.* **1975**, *16*, 871–877.
556. S. Kamisaka; K. Shibata, *Plant Cell Physiol.* **1977**, *18*, 1057–1066.
557. M. Cannac; V. Pasqualini; S. Greff; C. Fernandez; L. Ferrat, *Molecules* **2007**, *12*, 1614–1622.
558. A.-K. Borg-Karlson; G. Nordlander; A. Mudalige; H. Nordenhem; C. R. Unelius, *J. Chem. Ecol.* **2006**, *32*, 943–957.
559. W. A. Whistler, *Tropical Ornamentals: A Guide*; Timber Press: Portland, OR, USA, 2000.
560. J. K. Francis, *Guaiacum officinale* L. In *Tropical Tree Seed Manual*; J. A. Vozzo, Ed.; USDA Forest Service: Washington, DC, USA, 2002; pp 483–485.

Biographical Sketches

Daniel G. Vassão obtained his B.S. in chemistry from the University of São Paulo, Brazil, in 2002, and then completed his Ph.D. requirements in biochemistry under the supervision of Professor Lewis in 2008. His Ph.D. thesis involved establishing the biosynthetic routes leading to allyl/propenyl phenols, such as chavicol/eugenol in herbs and spices, and also to structurally related (neo) lignans. He has received several awards including the SMB Excellence in Graduate Research (2007), an Arthur M. and Kate Eisig-Tode Research Scholarship, the Charles Glen King Prize (2008), the John and Maggie McDougall Fellowship (2006/2007), and the Helen and Loyal H. Davis Fellowship (2004/2005).

Kye-Won Kim graduated in 2001 with a Ph.D. degree from Kyungpook National University (under the supervision of Professor Jong-Guk Kim). Her studies focused on the biosynthesis/molecular biology of capsaicinoid formation in *Capsicum annuum* and of factors involved in the metabolic regulation of capsaicinoids. She joined Professor Lewis' group as a postdoctoral fellow in 2001, and has worked largely on the functional characterization of genes/proteins involved in both lignan and lignin biosynthesis in plants. There is an emphasis in her work on dirigent proteins, which mediate phenoxy radical coupling, as well as how enantiospecificity of pinoresinol/lariciresinol reductase is achieved.

Laurence B. Davin received her B.Sc. (1983) and Ph.D. (1987) degrees in plant biochemistry/physiology from the Université Paul Sabatier (Toulouse, France). Dr. Davin next studied glucosinolate biogenesis with Ted Underhill in the National Research Council of Canada's Plant Biotechnology Institute as a postdoctoral fellow. Since 1989, she has turned her attention to lignan, lignin, and allyl/propenyl phenol biosynthesis at both the Virginia Polytechnic Institute and State University (Blacksburg, VA) and the Institute of Biological Chemistry (Washington State University). Dr. Davin is a member of the Editorial Board for Recent Advances in Phytochemistry.

Professor Norman G. Lewis completed his B.Sc. at the University of Strathclyde (Honors, chemistry, 1973) and his Ph.D. at the University of British Columbia (chemistry, 1977). He next pursued postdoctoral studies at Cambridge University (chemistry, 1978–80) with Professor Sir Alan R. Battersby. Before joining Washington State University in 1990, he held various research positions at ICI, the Pulp and Paper Research Institute of Canada, the National Research Council of Canada, and Virginia Tech. His research interests mainly involve lignan, lignin, and allyl/propenyl phenol biosynthesis. He serves on numerous scientific advisory committees, grant panels, and editorial boards, including as Regional Editor (*Phytochemistry*), Monitoring Editor (*Plant Physiology*), and as Executive Editor of *Advances in Plant Biochemistry and Molecular Biology.*

1.24 Plant Phenolics: Phenylpropanoids

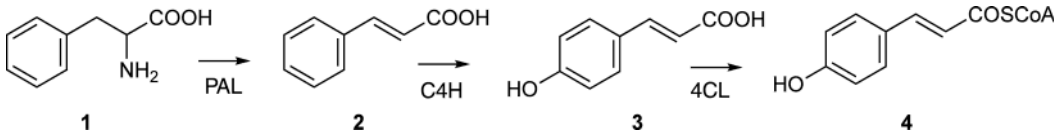
Shin-ichi Ayabe, Hiroshi Uchiyama, Toshio Aoki, and Tomoyoshi Akashi, Nihon University, Fujisawa, Japan

© 2010 Elsevier Ltd. All rights reserved.

1.24.1	**Introduction**	929
1.24.1.1	Brief Overview of Phenylpropanoid Research in the Past 10 Years	930
1.24.2	**New Enzyme Functions in Phenylpropanoid Metabolism**	931
1.24.2.1	CYP98A Catalyzing 3-Hydroxylation of 4-Coumaric Acid Esters	931
1.24.2.2	Phenylpropene Synthase and *O*-Methyltransferases Involved in Flavor and Scent Biosynthesis	931
1.24.2.3	Curcuminoid Synthase	933
1.24.2.4	Biosynthesis of Coumarins	934
1.24.2.4.1	Scopoletin biosynthesis in *Arabidopsis thaliana*: Identification of feruloyl-CoA 6′-hydroxylase as a 2-oxoglutarate-dependent dioxygenase	934
1.24.2.4.2	Furanocoumarin biosynthesis: Identification of psoralen synthase	935
1.24.2.4.3	Modification of coumarin skeleton	936
1.24.3	**Flavonoids**	937
1.24.3.1	Chalcones	937
1.24.3.1.1	6′-Deoxychalcone	938
1.24.3.1.2	Decoration of chalcones	939
1.24.3.1.3	Biosynthesis of xanthohumol	940
1.24.3.2	Aurones	940
1.24.3.3	Chalcone Isomerase and Flavanones	941
1.24.3.4	Flavones, Flavonols, Anthocyanins, and Proanthocyanidins	943
1.24.3.4.1	Construction of flavonoid skeletons	943
1.24.3.4.2	Decoration enzymes	948
1.24.3.5	Isoflavonoids	957
1.24.3.5.1	Biosynthesis of isoflavones in leguminous plants	958
1.24.3.5.2	Isoflavonoid phytoalexins	962
1.24.3.5.3	Transformation	966
1.24.4	**Functional Genomics and Evolution of Phenylpropanoid/Flavonoid Biosynthesis**	966
1.24.4.1	Metabolomics and Transcriptomics	966
1.24.4.2	Organization and Phylogenetic Relationship of Genes of Phenylpropanoid and (Iso)Flavonoid Pathway	967
1.24.4.3	Evolution of Phenylpropanoid/Flavonoid Biosynthesis	967
References		968

1.24.1 Introduction

This chapter describes the findings during the past 10 years on the biosynthesis of plant phenylpropanoids excluding lignans, lignin, and related monolignols, which will be reported in Chapter 1.23. Details on polyketide synthases including chalcone synthase (CHS), the first committed enzyme of flavonoid pathway, and related stilbene synthase will also be discussed in Chapter 1.06. The biosynthesis of phenylpropanoids starts with the deamination of L-phenylalanine (**1**) by L-phenylalanine ammonia-lyase (PAL) to yield *trans*-cinnamic acid (**2**), with very rare cases where tyrosine ammonia-lyase produces 4-coumaric acid (**3**). The subsequent actions on compound **2** by cinnamate 4-hydroxylase (C4H) and 4-coumaric acid-CoA ligase (4CL) yield 4-coumaroyl-CoA (**4**) (**Scheme 1**), from which the formal reducing processes on the carbonyl function and ring decorations by hydroxylation and *O*-methylation lead to monolignols, the direct precursors of lignin and lignans. The products of branching pathways



1 2 3 4

Scheme 1

of phenylpropanoids include a large family of flavonoids and isoflavonoids, stilbenes, coumarins, benzoic acid and benzaldehyde derivatives, phenylpropenes, diarylheptanoids, and their glycosides and other conjugates. The main focus in this chapter is on flavonoid and isoflavonoid biosynthesis, while new discoveries on the biosynthesis of coumarins, phenylpropenes, and diarylheptanoids are also outlined. The new enzymes for which genes and/or cDNAs are cloned and the biosynthetic pathways revealed by the discovery of the structural genes are the principal subjects, but some findings only based on biochemical characterization of the enzymes are also included.

1.24.1.1 Brief Overview of Phenylpropanoid Research in the Past 10 Years

The progresses of plant molecular biology and biochemistry have resulted in the identification of many genes and enzymes of the phenylpropanoid metabolism, and, more importantly, genome-oriented research that has grown explosively in the turn of the century has made available the structures of the genes that further expands into the studies on protein structure, evolutionary considerations of the pathway, and comparative ecophysiological applications. To find new enzyme functions, new methods based on genomic approaches combined with model plant resources have been successful; for example, a 2-oxoglutarate-dependent dioxygenase (2-ODD) involved in coumarin biosynthesis could be identified from the genome information of *Arabidopsis thaliana* and expression patterns in the MPSS database and further examining the mutant plants with suppressed candidate genes for the much reduced levels of the final product (see Section 1.24.2.4.1).[1] On the other hand, at the same time, the seemingly classical methodology in biosynthetic studies based on organic chemistry and biochemistry is recognized as increasingly more important, as the annotation of the gene functions depending solely on the amino acid sequence has often been inaccurate. Thus, the method starting with the purification of the enzyme protein gives very valuable information about the gene by which it is encoded; for example, the partial amino acid sequence of the peptide fragment of highly purified aureusidin synthase protein from 32 kg of snapdragon buds was combined with the subtractive hybridization approach to identify the new enzyme involved in aurone skeleton construction (see Section 1.24.3.2).[2] Also, although not a completely novel approach, combination of *in vitro* examination of the recombinant enzyme and fractionation of the cDNA library expressed in *Escherichia coli* or yeast has proven to be a powerful tool to characterize the enzyme for which the information on the protein is not available, such as 2-hydroxyisoflavanone dehydratase that yields isoflavones (see Section 1.24.4.5.1(ii)).[3] In the following sections, the new enzymes of the phenylpropanoid metabolism that have been characterized through these approaches are described.

Another remarkable progress regarding the enzymes of plant secondary metabolism in general is the X-ray crystallography-based determination of three-dimensional structures of the enzyme proteins. For the general phenylpropanoid pathway, the protein structure of the entry point enzyme, PAL from parsley (*Petroselinum crispum*) was determined.[4] An unusual prosthetic group (4-methylidene-imidazole-5-one, MIO) in common with histidine ammonia-lyase for the reaction was identified, and a critical motif in the protein structure has been defined (a compact review articles on PAL is available).[5] Although no protein structures of C4H and 4CL have been reported, the structures of many enzymes involved in phenylpropanoid metabolism like polyketide synthases, *S*-adenosyl-L-methionine (SAM)-dependent *O*-methyltransferases (OMTs), and reductases have been resolved, and the specific reaction mechanisms, substrate and product specificities, and the evolutionary mechanisms have been discussed.[6]

Homology modeling and site-directed mutagenesis have also been useful in elucidating the enzyme-catalyzed reaction specificities, especially for enzymes whose stereostructure are difficult to resolve through crystallography, such as cyochrome P-450s (P-450s). Higher plants are the richest source of P-450 molecules and the most abundant plant P-450 is C4H of the phenylpropanoid pathway, and homology modeling of this P-450 (CYP73A5) and other three P-450s of the phenylpropanoid pathway (coniferaldehyde 5-hydroxylase (CYP84A1), 4-coumaroyl shikimate 3-hydroxylase (CYP98A3), and flavonoid 3′,5′-hydroxylase (CYP75B1)) of *A. thaliana* on bacterial and mammal P-450s as template was performed, and the conserved active site architecture and the common

strategy to identify their substrate have been clarified.[7] Further studies on C4H structure and reaction mechanism have been carried out by site-directed mutagenesis,[8] and also nuclear magnetic resonance (NMR) was applied to locate the substrate analog in its active site.[9] Such a mutagenesis study was also performed with legume-specific P-450 that produces isoflavonoid skeleton (see Section 1.24.4.5.1(i)).[10] Glycosyltransferases of the phenylpropanoid and flavonoid metabolism are other examples for which modeling/mutagenesis yielded interesting findings on reaction specificity and evolution mechanism (see Section 1.24.4.5.1(v)(i)).[11]

Metabolic channeling as a mechanism of effective biosynthesis of specific biosynthetic pathway has been discussed for a long time, and a molecular biological–biochemical study using the yeast two-hybrid system methodology provided strong evidence of enzyme interaction in flavonoid metabolism in *A. thaliana*.[12] In planta analysis of multiple-transformed PAL and C4H genes further demonstrated the possible interactions of biosynthetic enzymes,[13,14] and several findings on channeling within phenylpropanoid/flavonoid metabolism have been reported.[15,16] However, reconstitution of the phenylpropanoid pathway in yeast revealed no evidence of channeling by multienzyme complex,[17] and there have been many debates on this matter; one reason is because of the little direct evidence of protein interactions. New techniques to detect weak and multiple protein interactions in plant secondary metabolism are highly expected. Also, there have been several interesting observations on subcellular localization of enzymes, such as nuclear,[18] plastidial,[19] and vacuolar[20] of flavonoid pathway.

Finally, what is not discussed here are: regulation of biosynthesis viewed from regulatory genes (transcription factors) and application-oriented researches such as microbial production of bioactive flavonoids or molecular breeding of ornamental plants with new traits, as the topics will be the subject in other volumes of this series.

For the recent advances in the research on phenylpropanoid and flavonoid metabolism, readers are referred to articles by Boudet and Petersen among others.[21,22]

1.24.2 New Enzyme Functions in Phenylpropanoid Metabolism

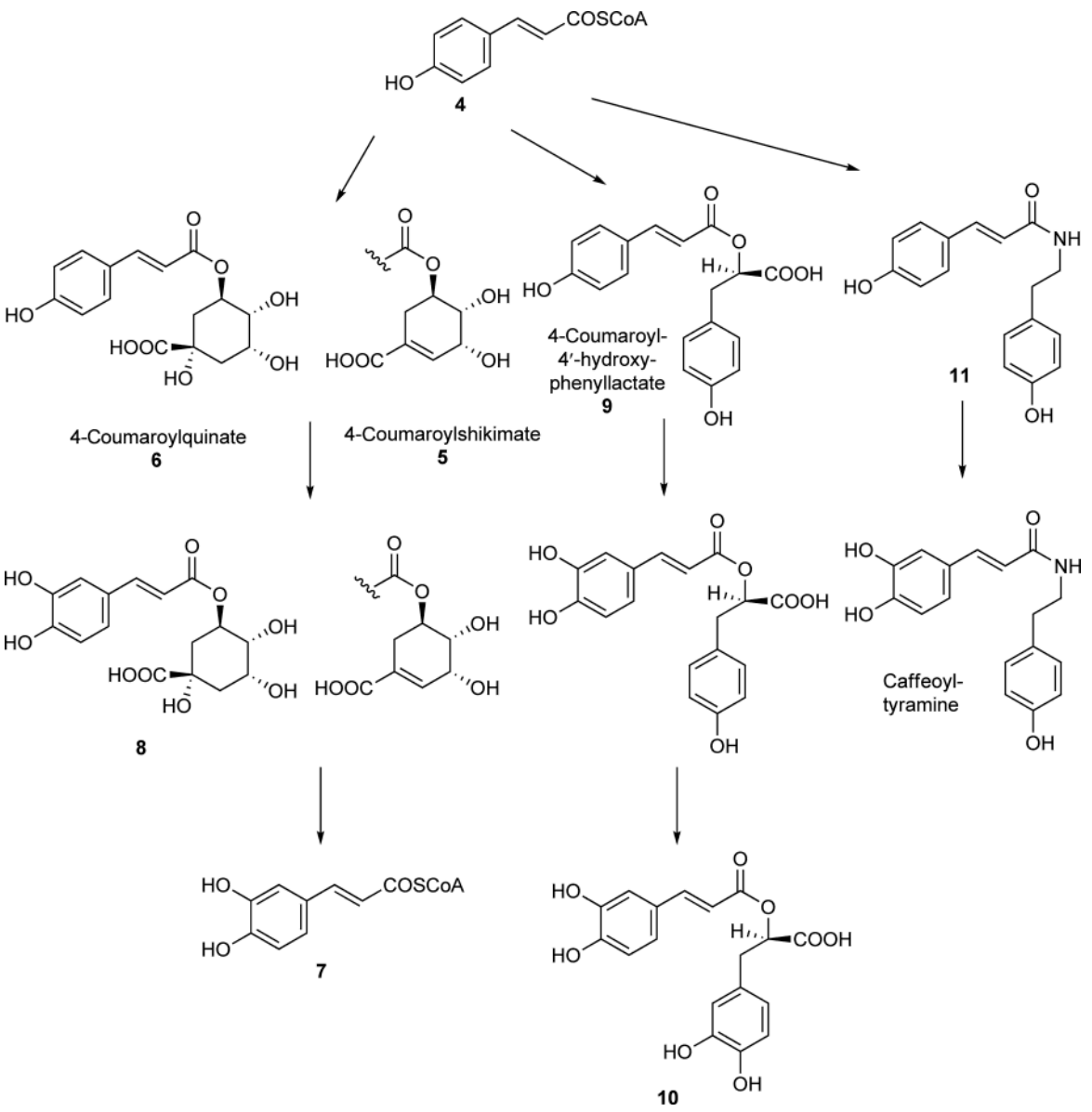
Monolignol pathway has been well understood in recent years. Because all the phenylpropanoid branching pathways originate in some of the precursors and intermediates of monolignol pathway, the new enzyme functions relevant to general and branching pathways within the plant phenylpropanoid metabolism are briefly discussed below, although there may be some overlap with other chapters.

1.24.2.1 CYP98A Catalyzing 3-Hydroxylation of 4-Coumaric Acid Esters

The *meta*-hydroxylation during the biosynthesis of phenylpropanoids with caffeic acid-type substituted aromatic ring has long been elusive. A P-450 of *A. thaliana*, CYP98A3, was for the first time shown to catalyze *meta*-hydroxylation of 5-*O*-shikimate (**5**) and 5-*O*-D-quinate (**6**) esters of 4-coumaric acid (**3**), but not of free acid (**Scheme 2**).[23,24,25] Together with the finding of acyltransferases producing shikimate/quinate esters of 4-coumarate from 4-coumaroyl-CoA (**4**) and reesterifying to yield caffeoyl-CoA (**7**), the missing link in the monolignol pathway was thus clarified, and the homologous P-450 enzymes have been detected from several plant species.[21,22,26] This finding also revealed the biosynthetic mechanism of chlorogenic acid (**8**) (caffeic acid quinate ester; **Scheme 2**). A P-450 enzyme of the same CYP98A subfamily has also been identified as 4-coumaroyl-4′-hydroxyphenyllactic acid (**9**) 3-hydroxylase (CYP98A6) of rosmarinic acid (**10**) biosynthetic pathway in *Lithospermum erythrorhizon* (the same activity was also detected in sweet basil (*Ocimum basilicum*) from which CYP98A13 has been characterized as 4-coumaroyl shikimate 3-hydroxylase).[27,28] Furthermore, wheat was found to contain several CYP98As, and they can *meta*-hydroxylate 4-coumaroyltyramine (**11**) (hydroxycinnamic acid amide), CYP98A12 being most effective, for the synthesis of suberin phenolic monomers.[26]

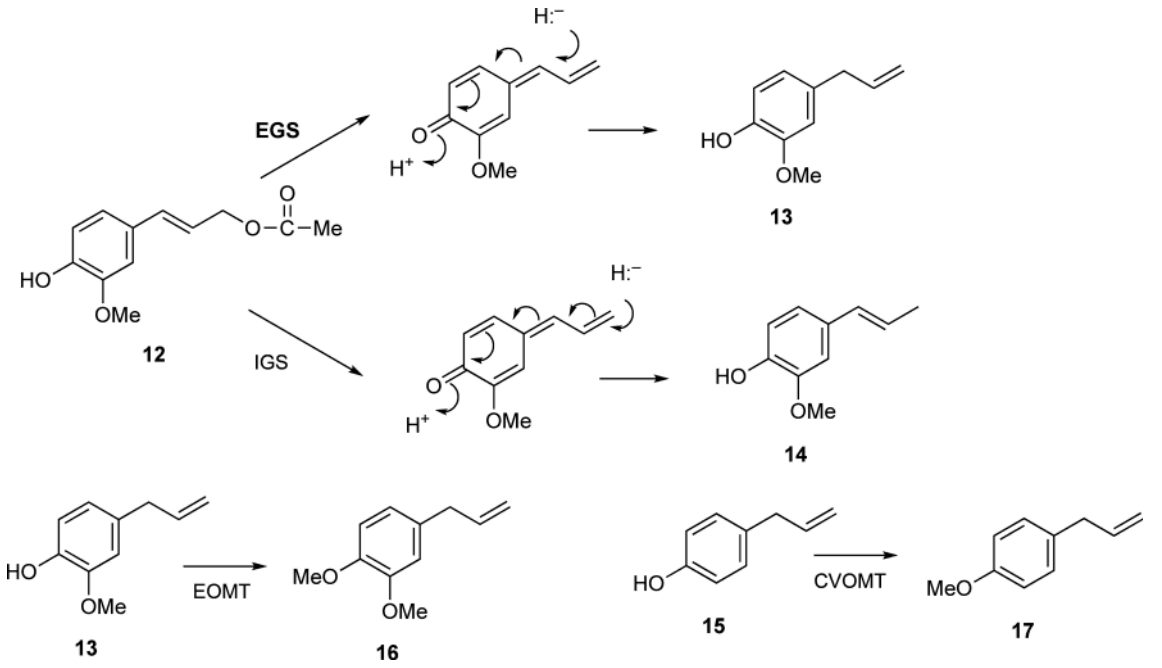
1.24.2.2 Phenylpropene Synthase and *O*-Methyltransferases Involved in Flavor and Scent Biosynthesis

Plant volatiles that act as defensive compounds as well as attractants to pollinators constitute of several classes of secondary metabolites such as terpenes and phenolics.[29,30] Phenylpropenes are derived from phenylpropanoid pathway, and the skeletons were shown to be produced from substituted cinnamyl acetate by specific classes of



Scheme 2

reductases (**Scheme 3**). In glandular trichomes of basil (*O. basilicum*), coniferyl acetate (**12**) is reduced to eugenol (**13**) by ObEGS1, while in petunia (*Petunia hybrida*) flowers, the same substrate is converted to isoeugenol (**14**) by PhIGS1.[31] Two proteins, CbIGS1 and CbEGS1, similar to ObEGS1, from *Clarkia breweri* flowers produce respective phenylpropenes, and substitution of only a single residue of these proteins critically affected the product specificity.[32] Furthermore, the third *C. breweri* reductase CbEGS2 and the newly characterized petunia PhEGS2 were shown to represent proteins of distinct lineages to the foregoing proteins. All these proteins are among the so-called PIP reductases (named after pinoresinol–lariciresinol reductase (PLR), isoflavone reductase (IFR), and phenylcoumaran benzylic ether reductase (PCBER))[32] within the more general short-chain dehydrogenase/reductase (SDR) or reductase–epimerase–dehydrogenase (RED) family,[33–35] and the enzymes among this family have further been characterized as leucoanthocyanidin reductase (LAR), anthocyanidin reductase (ANR), and pterocarpan reductase (PTR) in addition to the well-known dihydroflavonol 4-reductase (DFR) and vestitone reductase (VR) (see Sections 1.24.3.4.1(iii), 1.24.3.5.2(i)(b), and 1.24.3.5.2(i)(c)). In basil, **13** and chavicol

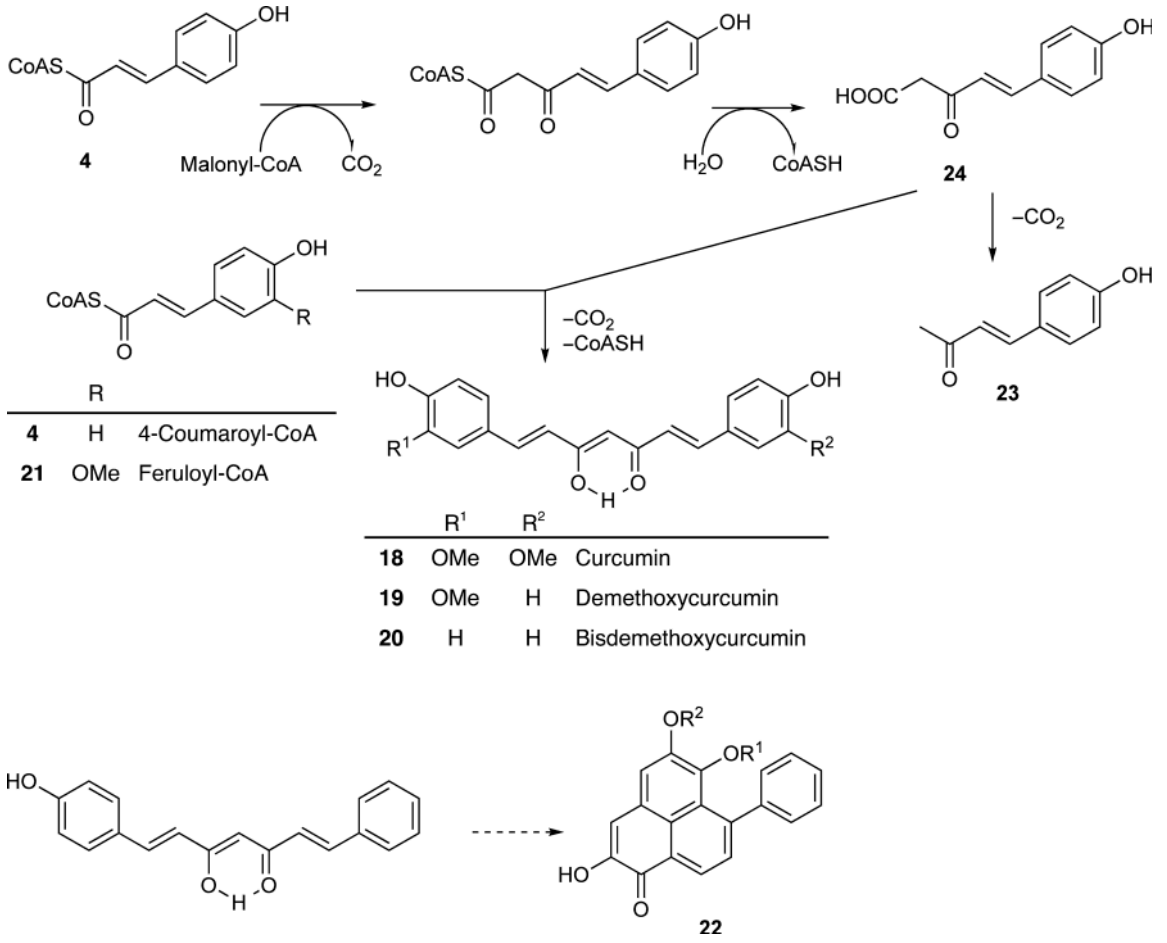


Scheme 3

(15) are methylated to yield methyleugenol (16) and methylchavicol (17) (Scheme 3), and respective OMTs (EOMT and CVOMT) were found to be very similar that the site-directed mutants in which one amino acid residue was converted into each other (CVOMT1 (F260S) and EOMT1 (S261F)) exhibited opposite substrate preferences to the respective native enzyme.[36] These OMTs are classified as the members of small molecule OMT (SMOMT) family, and phylogenetic analysis revealed that basil OMTs are clustered with orcinol OMTs (OOMTs) of rose producing a scent compound dimethoxytoluene and isoflavone 7-OMT (D7OMT) of alfalfa while another cluster contains caffeic acid OMTs (COMTs), *C. brewerі* isoeugenol OMT (IEMT), and rose phloroglucinol OMT (POMT).[29] The biosynthesis of *t*-anethole (7,8-double bond isomer of methylchavichol **17**) from coumaroyl acetate by a redutase (*t*-anol/isoeugenol synthase, AISl) of ObEGSl lineage and an *O*-methyltransferase (*t*-anol/isoeugenol OMT, AIMTl) in anise has also been reported.[37]

1.24.2.3 Curcuminoid Synthase

From turmeric (*Curcuma longa*) extracts, curcuminoid synthase activity producing curcumin (**18**), demethoxycurcumin (**19**), and bisdemethoxycurcumin (**20**) by the incubation of malonyl-CoA with feruloyl-CoA (21) and 4-coumaroyl-CoA (4) has been detected (Scheme 4).[38] A type III polyketide synthase (WtPKS1) cloned from root cultures of *Wachendorfia thyrsiflora* that contains curcuminoid-derived phenylphenalenones (22) and related compounds produced 4-hydroxybenzalacetone (23) from the starter compound **4** and a malonyl-CoA as the extension unit.[39] This reaction can be the first step in diarylheptanoid biosynthesis, and the observed production of pyrones may be the result of the absence of downstream processing proteins. Furthermore, a rice (*Oryza sativa*) polyketide synthase (os07g17010) synthesized compound **20** from two 4-coumaroyl-CoAs and one malonyl-CoA *in vitro* and was designated curcuminoid synthase (CUS).[40] This result is very surprising, because rice is not known to contain curcuminoids, and the reaction mechanism violates the traditional head-to-tail model of polyketide assembly; instead the diketide intermediate is hydrolyzed to a β-ketoacid (24) that subsequently serves as the second extender. The *in vivo* role of CUS in rice and the presence of CUS-like enzyme in Zingiberales that include turmeric and ginger, the famous producer of diarylheptanoids, are the next interesting questions, and, very recently, the involvement of two type III PKSs, diketide-CoA synthase (DCS) and curcumin synthase (CURS), in curcuminoid synthesis in turmeric (*C. longa*) was demonstrated.[41]



Scheme 4

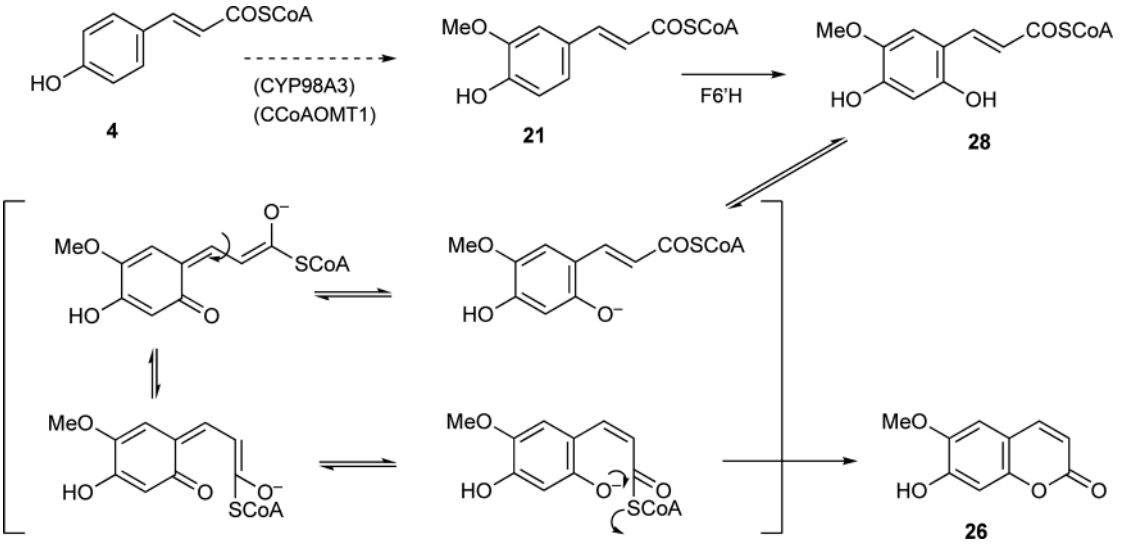
1.24.2.4 Biosynthesis of Coumarins

Detailed description on coumarin biosynthesis can be found in a review by Matern *et al.*[42] Earlier studies have shown that elicited cell cultures of Apiaceae (Umbelliferae) plants, *Ammi majus* and *P. crispum*, accumulate linear furanocoumarins such as bergapten (**25**) as phytoalexins. Recent enzymatic and molecular biological studies have revealed several specific enymes that decorate the coumarin skeleton, produce the furanocoumarin skeleton, and are responsible for the production of the direct precursor of coumarin skeleton formation. Remarkable findings are derived from the studies with *A. majus*, *Nicotiana tabacum*, and, interestingly, a monocot, rye, and, reasonably, *A. thaliana.*

1.24.2.4.1 Scopoletin biosynthesis in Arabidopsis thaliana*: Identification of feruloyl-CoA 6′-hydroxylase as a 2-oxoglutarate-dependent dioxygenase*

Schematically, the coumarin skeleton is formed through *ortho*-hydroxylation of the hydroxylated-cinnamate derivative, *trans–cis* isomerization of the C_3 part, and final lactonization. None of these processes had been characterized at the molecular level.

Arabidopsis thaliana was found to accumulate scopoletin (**26**) and its β-D-glucopyranoside, scopolin (**27**), mainly in the root part ($\sim$1200 nmol gFw^{-1}).[43] The shoot part contained lesser amount ($\sim$180-fold smaller) of **27**, but treatment with an auxin, 2,4-dichlorophenoxyacetic acid, induced the production of the coumarins. The *A. thaliana* mutant line carrying a T-DNA insertion in the CYP98A3 gene produced a much reduced amount of coumarins ($\sim$3% of that in the wild-type root) indicating that 3′-hydroxylation of the 4-coumarate unit within



Scheme 5

the phenylpropanoid pathway by this P-450 is critically important in coumarin biosynthesis (**Scheme 5**). Similarly, the homozygous mutant of caffeoyl-CoA *O*-methyltransferase 1 (*CCoAOMT1*) contained considerably lower levels of **26** (30%) and **27** (15%) compared to wild-type roots. Therefore, this enzyme and the enzyme product, feruloyl-CoA (**21**), should be involved in the biosynthesis of compound **26** in *A. thaliana*.

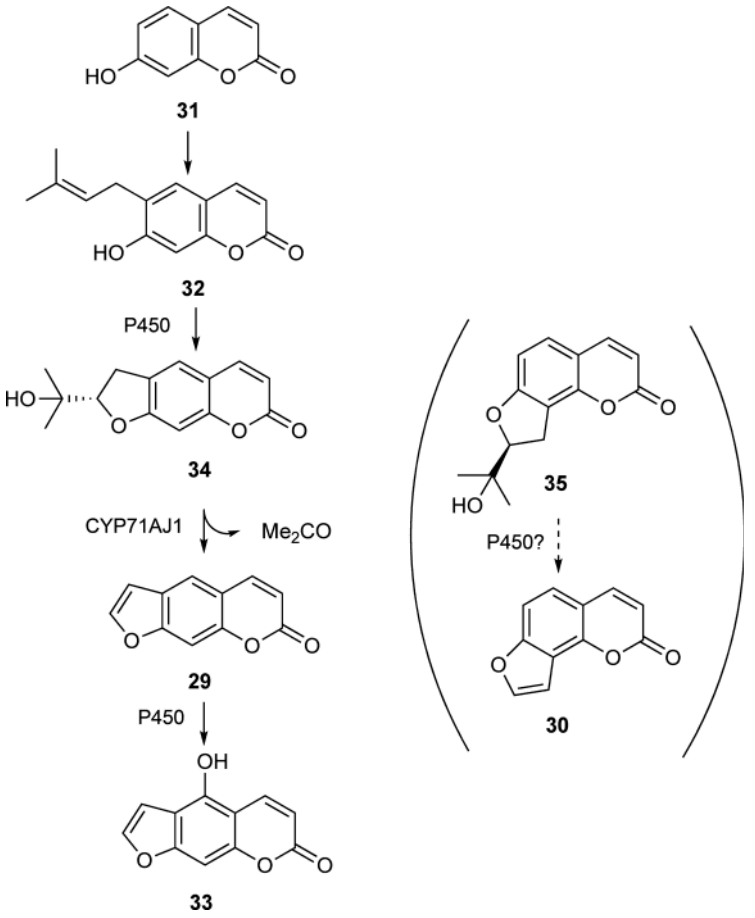
Based on these findings, candidate genes encoding oxygenases that catalyze *ortho*-hydroxylation of feruloyl-CoA (**21**) were selected, and a 2-ODD gene (At3g13610) was found to be predominantly expressed in roots and its mutant plant contained only very low levels of the coumarins in roots.[1] Indeed, the recombinant enzymes transcribed from the gene *F6′H1* and its homolog *F6′H2* exhibited feruloyl-CoA 6′-hydroxylase activity yielding 6′-hydroxyferuloyl-CoA (**28**) (**Scheme 5**). The substrate specificity of these enzymes was similar: **21** served as the predominant substrate and 4-coumaroyl-CoA (**4**) yielded a trace amount of the product, but cinnamoyl-CoA and caffeoyl-CoA (**7**) and free acids did not act as the substrate. Very interestingly, the enzyme reaction mixture under dark contained a significant level of scopoletin (**26**), which should be produced through nonenzymatic enolization of CoA thioester, *trans–cis* isomerization of the enolate anion, and lactone ring formation with $CoAS^-$ as the leaving group.

The mechanism of scopoletin biosynthesis in *A. thaliana* is thus now well understood, awaiting a more detailed characterization of geometrical isomerization and lactonization processes. Whether this 2-ODD-dependent *ortho*-hydroxylation of phenylpropanoid CoA-thioester is the general mechanism for the construction of coumarin skeleton in other plants also needs further study, but undoubtedly the finding reported above is a breakthrough in the coumarin biosynthesis mechanism.

1.24.2.4.2 Furanocoumarin biosynthesis: Identification of psoralen synthase

Linear and angular furanocoumarins such as psoralen (**29**) and angelicin (**30**) are biosynthesized by initial prenylation at 6- or 8-position of umbelliferone (**31**), and the respective products, demethylsuberosin (6-dimethylallylumbelliferone, **32**) and osthenol (8-dimethylallylumbelliferone), are subjected to P-450-catalyzed oxidations (**Scheme 6**). Linear series has been the target of the research, and separate P-450s, (+)-marmesin synthase, psoralen synthase, and psoralen 5-monooxygenase have been assumed to produce bergaptol (**33**), which is finally *O*-methylated to yield bergapten (**25**). *Ammi majus* cell cultures have been the excellent material to study the biosynthesis of linear furanocoumarins, as the biosynthetic enzymes are rapidly induced in the elicited cells. Recently, a novel P-450 psoralen synthase was identified at the molecular level.[44]

Total RNA from cells treated with elicitor for 3–5 h was used for the template of differential display RT-PCR amplification of P-450 sequences, and *CYP71AJ1* was cloned. The successful heterologous expression of this P-450 needed modification of the N-terminal 37 amino acid residues and the use of specific yeast strain WAT11, and the recombinant CYP71AJ1mut produced psoralen (**29**) from marmesin (**34**) requiring

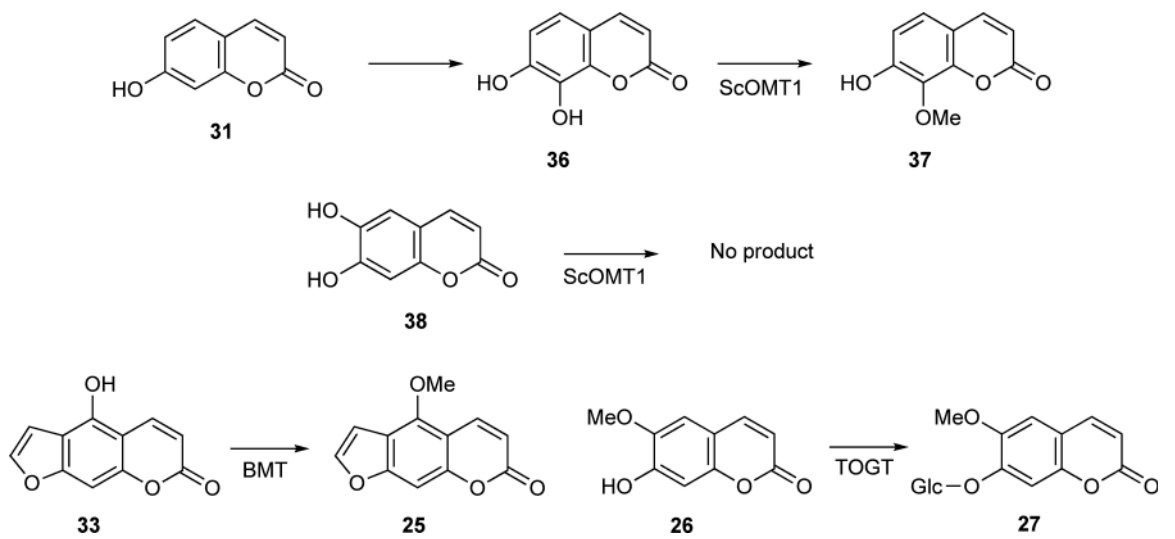


Scheme 6

(nicotinamide-adenine dinucleotide phosphate (NADPH) reduced form) as the cofactor. The yeast harboring the wild-type *CYP71AJ1* also yielded **29** when the precursor was fed to the culture. Thus, CYP71AJ1 was identified as psoralen synthase, which first abstracts the *sym*-hydrogen at C-3′ to the isopropyloxy group and, after the carbon chain cleavage, releases acetone as a by-product. The enzyme is specific to (+)-marmesin (**34**) with a minor activity to 5-hydroxymarmesin, and its transcript accumulated to the maximum 4 h after elicitation followed by the maximum **29** level at 9–10 h. The intermediate of angular furanocoumarin, (+)-columbianetin (**35**), competitively inhibited psoralen synthase activity, and the modeling study on the active site structure suggested that a limited number of mutations may have transformed psoralen synthase to angelicin synthase.

1.24.2.4.3 Modification of coumarin skeleton

1.24.2.4.3(i) Daphnetin 8-O-methyltransferase Daphnetin (7,8-dihydroxycoumarin, **36**) is a coumarin that could be synthesized by 8-hydroxylation of umbelliferone (**31**), the simple coumarin originated from 4-coumarate-type substituted phenylpropanoid (**Scheme 7**). During cold acclimation of rye plants, several genes are associated with increased photosystem II excitation pressure, and among these genes, *ScOMT1* exhibited homology to plant OMT genes. The biochemical characterization of recombinant ScOMT1 displayed a novel enzyme activity, yielding 7-hydroxy-8-methoxycoumarin (**37**).[45] Its substrate specificity was limited to daphnetin, not employing esculetin (6,7-dihydroxycoumarin, **38**) as a substrate, and the regiospecificity of methyltransfer reaction was also only against 8-hydroxyl group. As daphnetin is known as a protein kinase inhibitor, the upregulation of this OMT in response to PSII excitation pressure and low temperature is considered to be involved in low-temperature signaling.



Scheme 7

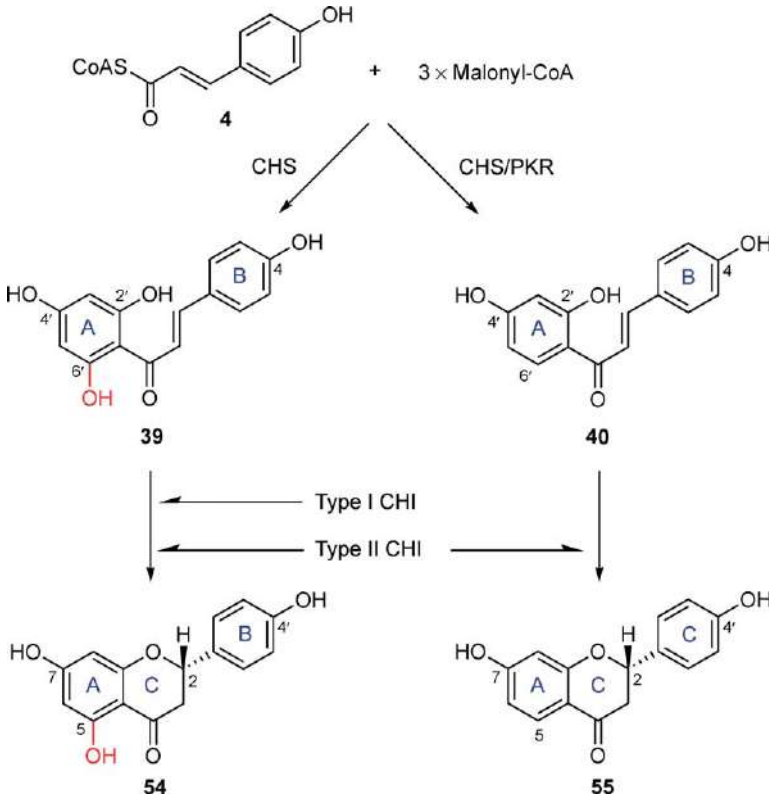
1.24.2.4.3(ii) O-Methyltransferases involved in furanocoumarin biosynthesis The 5- or 8-hydroxylated furanocoumarin (bergaptol (**33**) or xanthotoxol) are *O*-methylated to yield bergapten (**25**) or xanthotoxin, and the OMTs involved in the reactions have been characterized in *P. crispum* and *Ruta graveolens* cultured cells. Bergaptol 5-*O*-methyltransferase (BMT) cDNA was cloned from dark-grown *Ammi majus* cells treated with a crude fungal elicitor.[46] BMT was shown to be an elicitor-inducible labile enzyme highly specific to bergaptol in contrast to constitutively expressed caffeate OMT (COMT) with broader specificity to caffeate derivatives including esculetin (**38**) and daphnetin (**36**).

1.24.2.4.3(iii) Coumarin glycosyltransferases and plant defense in tobacco Tobacco plants have traditionally been used as the experimental system for pathogen resistance and phenylpropanoid metabolism. Uridine diphosphate (UDP)-glucose: phenylpropanoid glucosyltransferase (TOGT) of tobacco produces scopolin (**27**) from scopoletin (**26**). Downregulation of TOGT by antisense methodology resulted in a significant reduction of **27** and, unexpectedly, also **26**, associated with weakened resistance to infection with tobacco mosaic virus (TMV).[47] This result is attributed to a direct antiviral effect and reactive oxygen intermediate (ROI) buffering of the coumarins. Phenylpropanoid glucosyltransferase-overexpressor tobacco plants showed an increase in the bright blue fluorescence surrounding the hypersensitive response (HR) necrotic lesions under UV light upon TMV infection and suggested an accelerated HR, although no clear difference in viral content in each lesion was detected,[48] whereas increased resistance of TOGT overexpressing tobacco to infection with potato virus Y (PVY) has been reported.[49] cDNAs for new stress-related glucosyltransferases with a broad substrate specificity including **26** as a substrate and another glucosyltransferase that reacts on 7-hydroxyl of flavonol and 3-hydroxyl of coumarin skeleton have been cloned from cultured tobacco cells.[50,51]

1.24.3 Flavonoids

1.24.3.1 Chalcones

Chalcone is a yellow pigment, and its skeleton (1,3-diphenyl-2-propen-1-one) is the initial intermediate structure used in the biosynthesis of all flavonoids. Chalcones, most typically 2′,4′,6′,4-tetrahydroxychalcone (THC or naringenin chalcone, **39**), are synthesized *in vivo* by CHS from one molecule of phenylpropanoid-derived CoA-esters, 4-coumaroyl-CoA (**4**) being the most common substrate, and three molecules of malonyl-CoA (**Scheme 8**).[52] Sequential condensation of three acetate units from malonyl-CoA and subsequent



Scheme 8

cyclization form the ring A of chalcones and other flavonoids, and the ring B is originated from phenylpropanoid-derived CoA-esters. CHS, which is categorized as type III polyketide synthase, is a ubiquitous enzyme in higher plants[53] and its biochemistry is discussed in detail in Chapter 1.04.

1.24.3.1.1 6′-Deoxychalcone

6′-Deoxychalcone such as 2′,4′,4-trihydroxychalcone (isoliquiritigenin, **40**) and its derived 5-deoxyflavonoids are very common compounds in leguminous plants; however, they are fairly rare in the other plant species. 6′-Deoxychalcone is *in vivo* synthesized by coaction of CHS and NADPH-dependent chalcone polyketide reductase (PKR) (or chalcone reductase, CHR) in leguminous plants (**Scheme 8**).[54] Reaction mechanism of PKR has been examined using artificial intermediates involved in the chalcone formation as substrates for PKR[55] and by the analysis of the crystal structure of PKR and computer-assisted docking of possible substrates.[56] Sequence analysis indicates that PKR is a member of aldo/keto reductase (AKR) superfamily enzymes.[56,57]

Nonleguminous plants (petunia and tobacco) transformed with the *PKR* gene from leguminous plants (alfalfa, licorice, *Pueraria*, and *Lotus*) synthesized deoxychalcones, isoliquiritigenin (**40**) and butein glucosides, and changed their flower color.[58–60] On the other hand, RNA interference (RNAi) silencing of *PKR* gene or isoflavonoid synthase (IFS) gene in soybean roots strongly reduced the production of 5-deoxyisoflavonoids, and led to a breakdown of Rps-mediated resistance to race 1 of *Phytophthora sojae* (see Section 1.24.3.5.3).[61] Furthermore, silencing of *PKR* or *IFS* led to the suppression of hypersensitive cell death associated with race-specific resistance to *P. sojae*. Thus, it is suggested that the 5-deoxyisoflavonoids play a critical role in the race-specific resistance and establishment of hypersensitive cell death.

Although 5-deoxyflavonoids have occasionally been reported from nonleguminous plants, such as *Crinum bulbispermum* (Amaryllidaceae)[62] and *Calea platylepis* (Asteraceae),[63] the metabolic pathway of deoxyflavonoids in these plants is still unclear.

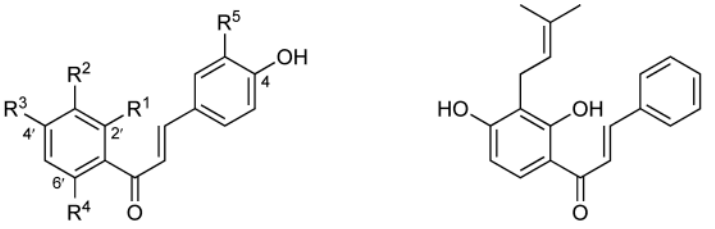
1.24.3.1.2 Decoration of chalcones

Chalcone itself is unstable and readily converted into flavanones in plant cells, but various modified chalcones that are more stable than the precursor occur in many plant species. They can play key roles in the ecophysiolosical system as pigments, phytoalexins, and symbiotic signals, and recently they are also attracting attention as bioactive compounds having cytotoxic and chemoprotective properties.[64,65] More than 900 chalcone derivatives occurring naturally have been reported in the literature to the end of 2003.[66]

1.24.3.1.2(i) Methylation of chalcones Chalcone *O*-methyltransferase (ChOMT) methylates the 2′-hydroxyl of isoliquiritigenin (**40**) to form 4,4′-dihydroxy-2′-methoxychalcone (**41**) (**Scheme 9**). The latter compound in alfalfa (*Medicago sativa*) of Fabaceae acts as the transcriptional activator of *nod* genes of *Sinorhizobium meliloti*.[67] The X-ray crystal structure of ChOMT was revealed and the reaction mechanism was elucidated.[68] Phylogenetic analysis of plant OMTs shows that the ChOMT was grouped with hydroxycinnaic acid OMTs rather than flavonoid OMTs.[69] An OMT (GeLMT/ChOMT) of licorice (*Glycyrrhiza echinata*), highly similar to and clustered, in phylogenetic trees, with alfalfa ChOMT, has been found to methylate 2′-hydroxyl of a dibenzoylmethane, licodione, as well as of **40**, and should be involved in the biosynthesis of a retrochalcone, echinatin.[70]

1.24.3.1.2(ii) Hydroxylation of chalcones The enzymatic 3-hydroxylation of isoliquiritigenin (**40**) to form butein (**42**) was demonstrated with microsomal preparations from yellow petals of *Dahlia variabilis* (Asteraceae).[71] The enzyme was tentatively addressed as chalcone 3-hydroxylase.[52] However, the substance of the enzyme has not yet been elucidated either by biochemical or molecular biological approaches.[72]

1.24.3.1.2(iii) Glycosylation of chalcones Chalcone glucosides are among the major pigments to create yellow coloration in flowers. Butein 4′-glucoside is accumulated in the flower of composite genera, such as *Dahlia* and *Coreopsis*, and its *in vitro* biosynthesis has been demonstrated.[73] Although the glucosyltransferase involved in the 4′-glucosylation of 6′-deoxychalcone has not been isolated, the gene of 6′-hydroxychalcone 4′-*O*-glucosyltransferase (*Am4′CGT*) has been cloned from snapdragon (*Antirrhinum majus*) of Scrophulariaceae and its function analyzed.[74] Recombinant Am4′CGT catalyzes glucosylation of THC (**39**) and 2′,4′,6′,3,4-pentahydroxychalcone (PHC, **43**) to form THC 4′-*O*-glucoside (**44**) and PHC



47

	R^1	R^2	R^3	R^4	R^5	
39	OH	H	OH	OH	H	Naringenin chalcone or THC
40	OH	H	OH	H	H	Isoliquiritigenin
41	OMe	H	OH	H	H	4,4′-Dihydroxy-2′-methoxychalcone
42	OH	H	OH	H	OH	Butein
43	OH	H	OH	OH	OH	Eriodictyol chalcone or PHC
44	OH	H	*O*-Glc	OH	H	THC 4′-*O*-glucoside
45	OH	H	*O*-Glc	OH	OH	PHC 4′-*O*-glucoside
46	*O*-Glc	H	OH	OH	H	Isosalipurposide
47	OH	Prenyl	OH	H	OH	4,2′,4′-Trihydroxy-3′-prenylchalcone

Scheme 9

4′-*O*-glucoside (**45**), respectively. On the other hand, yellow flowers of carnation (*Dianthus caryophyllus*) synthesize THC 2′-*O*-glucoside, and two cDNAs encoding chalcone 2′-*O*-glucosyltransferase were cloned.[75] The two enzymes, DicGT4 and DicGT5, exhibit broad substrate specificity and transfer a glucose from UDP-glucose to the hydroxyl group of a chalcone (2′-OH) as well as flavonol (3-OH and 7-OH), flavanone (7-OH), flavone (7-OH), and anthocyanidin (3-OH and 7-OH) molecules. Interestingly, both the chalcone isomerase (*CHI*) and *DFR* genes were disrupted by the transposable element *dTdic1s* in the carnation cultivars bearing yellow flowers,[76] although the transcripts of the *CHI* gene were detected by the more sensitive method of RT-PCR than that of Northern blot analysis.[77] In the case of barley (*Hordeum vulgare*) *anthocyanin-less310* mutant lacking a functional *CHI*, isosalipurposide (THC 2′-*O*-glucoside, **46**) was accumulated instead of saponarin (apigenin 6-C-glucosyl-7-*O*-glucoside) during primary leaf development.[78,79]

A multifunctional (iso)flavonoid glycosyltransferase, MtUGT85H2, has been characterized from *Medicago truncatula*.[80] MtUGT85H2 showed activity toward flavonols, isoflavones, and the chalcone isoliquiritigenin (**40**). The enzyme produced a single product with **40**; however, the position of glycosylation was not determined. The catalytic efficiency for **40** was 4.9-fold lower than for a flavonol (kaempferol: for *M. truncatula* GTs, see also Section 1.24.3.5.1(v)(a)).

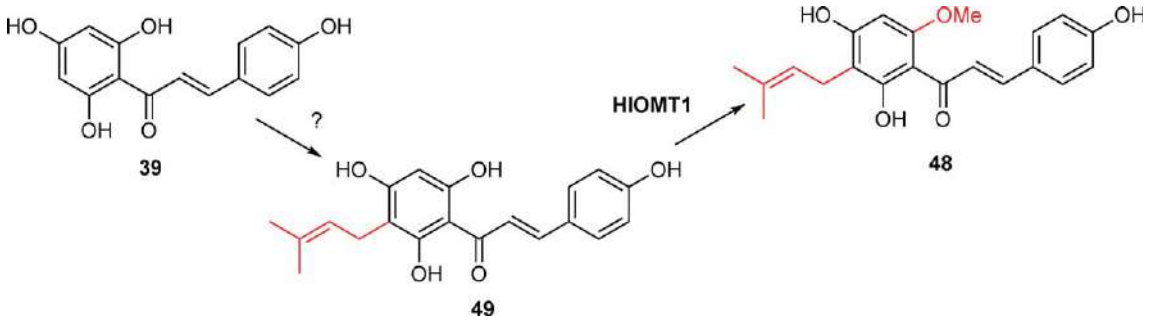
1.24.3.1.2(iv) Prenylation of chalcones Prenyltransferase activity to be able to prenylate the 2′,4′-dihydroxychalcone yielding isocordoin (**47**) and isoflavone genistein (**117**) was shown from the microsomal fractions of cell cultures of black mulberry (*Morus nigra*).[81] The activity required divalent cations, particularly Mg^{2+}, and the apparent K_m value for 2′,4′-dihydroxychalcone was 142 μmol l^{-1}. Recently, gene cloning of flavonoid-specific prenyltransferase found in *Sophora flavenscens* (Fabaceae) was performed, but the prenyltransferase responsible for the prenylation of the flavanone naringenin at the 8-position was not able to accept chalcones as substrates (see Section 1.24.3.4.2(i)).[19]

1.24.3.1.3 Biosynthesis of xanthohumol

Xanthohumol (**48**) (3′-(3,3-dimethylallyl)-2′,4′,4-trihydroxy-6′-methoxychalcone) is a prenylated chalcone that occurs only in the hop plant (*Humulus lupulus*), especially in the female inflorescences. Xanthohumol (**48**) and its related prenylflavonoids are important for the ingredient of beer and have received much attention as cancer chemopreventive agents.[82] Xanthohumol (**48**) originates from naringenin chalcone (**39**), but subsequent biosynthetic pathway involved in the prenylation and methylation of **39** is not completely understood. In hop cones, the presence of desmethylxanthohumol (**49**) and the absence of 6′-*O*-methylated naringenin chalcone suggest that prenylation occurs first.[82] Although the prenyltransferase related to the biosynthesis of **48** has not been detected, an *O*-methyltransferase (HlOMT1) that methylates **49** to form **48** has been identified by the expressed sequence tag (EST) analysis from hop glandular trichomes.[83] In a molecular phylogenetic tree, *HlOMT1* clustered with OMT enzymes involved in alkaloid biosynthesis and not with leguminous 2′-*O*-methyltransferases (alfalfa ChOMT and licorice ChOMT/LMT) that methylate the same position of the A ring. High and specific expression of the HlOMT1 in the lupulin glands, as evidenced by both RT-PCR and EST counts, was correlated with the high accumulation of **48**. The HlOMT1 reaction required divalent cations and showed a moderate affinity for **49** ($K_m = 18$ μmol l^{-1}) and a low affinity for SAM ($K_m = 286$ μmol l^{-1}). These features of HlOMT1 clarify the order of reactions in the xanthohumol pathway: **39** is prenylated and subsequently methylated to form **48** (**Scheme 10**). HlOMT2, another OMT of hop with a broad substrate specificity, methylated **48** to produce 4-*O*-methylxanthohumol, which is not known from hop.[83]

1.24.3.2 Aurones

Aurones give brighter yellow coloration in flowers than chalcones and are found in a limited number of plant genera such as *Antirrhinum* (Scrophulariaceae), *Cosmos bipinnatus* (Asteraceae), and *Limonium* (Plumbaginaceae). In addition to the flowering plants, aurones are also found in gymnosperms,[84] bryophytes,[85] and brown algae (4′-chloroaurone).[86] Aurones are generated in plants by the oxidative cyclization of 2′-hydroxychalcones. From the yellow flowers of snapdragon that accumulate the aurone (aureusidin (3′,4′,4,6-tetrahydroxyaurone, **50**) and bracteatin (3′,4′,5′,4,6-pentahydroxyaurone, **51**)) glucosides (**52**, **53**), aureusidin synthase (AmAS1), which is a homolog of plant polyphenol oxidase (PPO), has been identified and characterized.[2,87] The AmAS1 can accept THC



Scheme 10

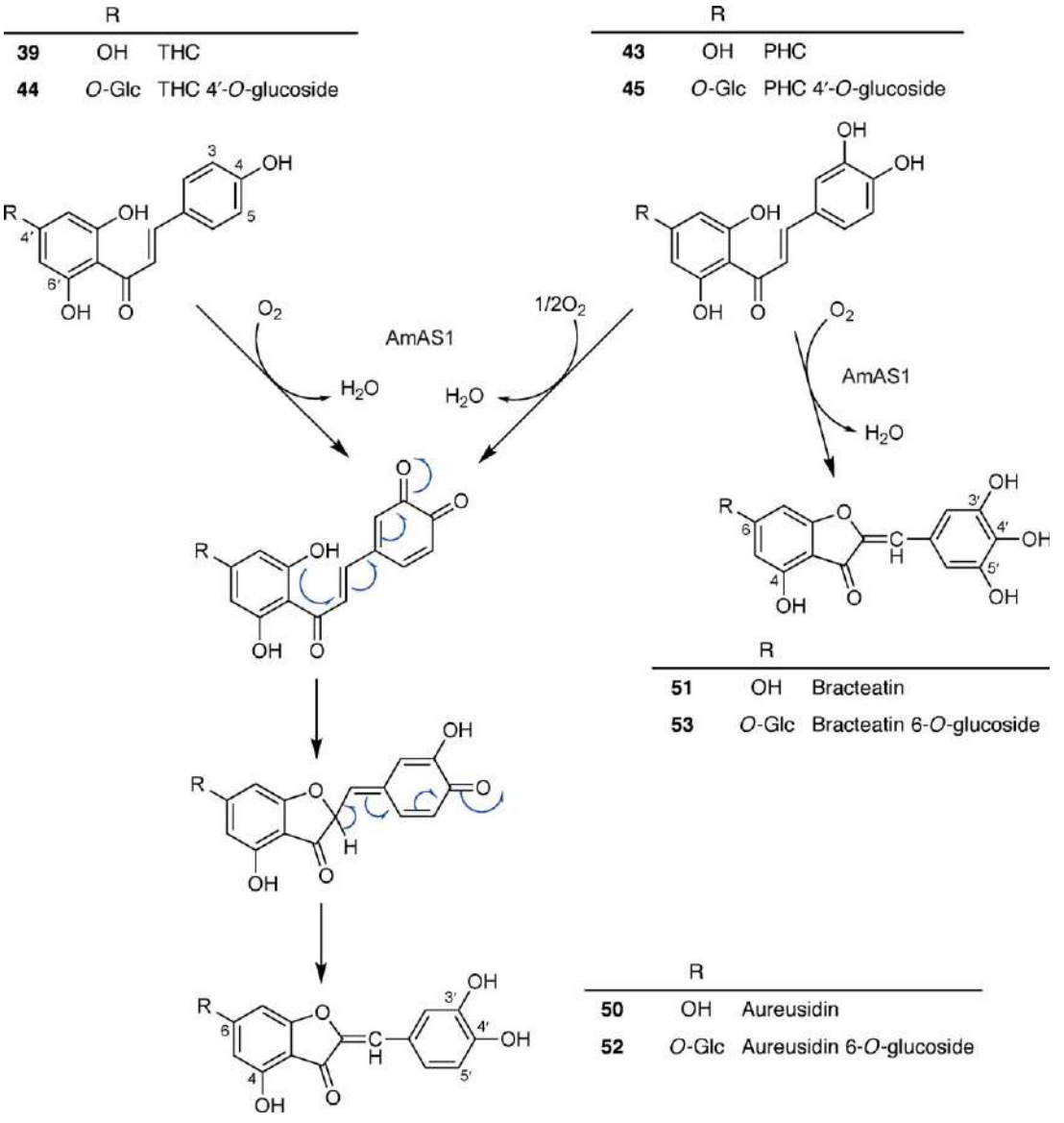
(**39**) and PHC (**43**) and, furthermore, 4′-glucosides of THC and PHC as substrates (**Scheme 11**). AmAS1 has some unusual features compared with known plant PPOs as follows: the enzyme is a copper-containing glycoprotein, cannot accept general substrates for plant PPOs (e.g., tyrosine and 3,4-dihydroxyphenylalanine), and does not share the N-terminal signal sequence needed for the transition into plastids. Thus, the AmAS1 is regarded to be an aurone biosynthesis-specific PPO. Interestingly, **50** is produced from either **39** or **43**, and the oxidative cyclization of **39** must be preceded by 3-oxygenation of the B ring as the first step of the enzyme catalysis (**Scheme 11**).[87] As expected, **51** is also synthesized from **43** together with **50** in the ratio of aureusidin:bracteatin = 6:1.

Although the AmAS1 is considered as a key enzyme for aurone biosynthesis from chalcones in snapdragon, transgenic plants of *Torenia hybrida* overexpressing *AmAS1* gene did not accumulate aurone glucosides in their flowers.[74] Therefore, the abovementioned chalcone 4′-*O*-glucosyltransferase (Am4′CGT) was examined in snapdragon, and then coexpression of both *Am4′CGT* and *AmAS1* genes in *Torenia* flowers sufficiently synthesized aureusidin 6-*O*-glucoside. Moreover, in the results of the analysis of subcellular localization of both Am4′CGT and AmAS1 by transient expression of chimeric proteins, the Am4′CGT-GFP and AmAS1(N_{1-60})-mRFP chimeric proteins localized separately in the cytoplasm and vacuole, respectively. The vacuole targeting signal of the PPO was identified to be located within the 53-amino acid residue N-terminal sequence.[20] These results indicate that chalcone 4′-*O*-glucosides synthesized by Am4′CGT in the cytoplasm were transported into the vacuole and therein transformed into aurone 6-*O*-glucosides by AmAS1. On the other hand, as aurones are found in a limited number of unrelated species, different types of enzymes could be involved in aurone biosynthesis.[88,89]

1.24.3.3 Chalcone Isomerase and Flavanones

Although chalcones bearing a hydroxyl at C-2′, especially those further possessing a 6′-hydroxy substitution, are spontaneously converted into a racemic mixture of the respective (2*S*)- and (2*R*)-flavanones easily, they are stereospecifically isomerized into (2*S*)-flavanones more rapidly by chalcone isomerase (CHI; formally called chalcone–flavanone isomerase, CFI) than spontaneous conversion in planta (see **Scheme 8**).[52,54] Only (2*S*)-flavanones can be utilized as intermediates of the subsequent flavonoid biosynthesis. CHIs found in nonleguminous plants accept only 6′-hydroxychalcone (naringenin chalcone, **39**) as substrate to form 5-hydroxyflavanone ((2*S*)-naringenin, **54**), while CHIs found in leguminous plants accept both **39** and 6′-deoxychalcone (isoliquiritigenin, **40**) as substrates to yield **54** and 5-deoxyflavanone ((2*S*)-liquiritigenin, **55**), respectively.

Three isozymes of CHIs were found in a leguminous plant, *G. echinata*, and their substrate specificities with partially purified proteins were examined.[90] As a result, two of them catalyzed the isomerization of both naringenin chalcone (**39**) and isoliquiritigenin (**40**), while the third one acted only on **39**. Thereafter, three cDNA clones (*LjCHI1–LjCHI3*) of *CHI* genes were isolated from a model legume, *Lotus japonicus*, and properties of the CHIs were investigated using recombinant proteins.[91] Among the three CHIs, LjCHI1 and LjCHI3 coding for type II CHI isomerized both **39** (LjCHI1: $K_m = 3\ \mu mol\ l^{-1}$ for naringenin chalcone) and **40** (LjCHI1: $K_m = 4\ \mu mol\ l^{-1}$ for isoliquiritigenin), while LjCHI2 coding for type I CHI was active only on **39** ($K_m = 3\ \mu mol\ l^{-1}$ for naringenin chalcone) (**Scheme 8**). Nucleotide and deduced amino acid sequences of the type II CHIs were highly homologous to the leguminous CHIs reported previously, whereas those of the type I CHI showed high homology with CHIs of nonleguminous plants and were clustered with nonleguminous CHIs



Scheme 11

in a phylogenetic tree. Thus, it has been revealed that the type I CHI is ubiquitous in angiosperms, while the type II CHI is specific in the family Fabaceae. Moreover, genome sequence analysis revealed that type I and II *CHI* genes form a tandem cluster within 15 kb, and type II *CHI* gene could have evolved from ancestral type I *CHI* gene by a local gene duplication and subsequent accumulation of mutations. This neofunctionalization of *CHI* genes in Fabaceae may be driven from the selective pressure of the ecological and physiological significant role of 5-deoxy(iso)flavonoids, which is originated from the products of the type II CHI activity. This report on *CHI* is one of the pioneer researches tackled with a viewpoint of genomic and evolutionary analyses, although gene duplication has been shown for several gene families involved in plant secondary metabolism.[92] The same situation has been confirmed in soybean CHIs (GmCHI1A and GmCHI1B2 of type II and GmCHI2 of type I), and CHI-like proteins with low sequence identity (GmCHI3 and GmCHI4, both of which do not metabolize the chalcones) have also been found.[93]

To date, functional plant *CHI* genes have been reported from over 25 species only in the angiosperms. Expression pattern among different organs or varieties and during organ development has been investigated in

some species.[94–96] *CHI* genes are regulated together with other flavonoid genes, such as chalcone synthase, flavanone 3-hydroxylase, and dihydroflavonol reductase, by the Myb-type transciption factors in *Arabidopsis*,[97] Japanese morning glory,[98] and maize and sorghum.[99] Transgenic plants overexpressing *CHI* gene began to produce a large amount of flavonoids versus their control plants. A flavonol glycoside rutin was mainly accumulated in the transgenic tomato[100] (transformed with petunia *CHI*) and tobacco[101] (transformed with *Saussura medusa CHI*) plants, and a flavone apigenin in the transgenic hairy roots[102] of *Saussura involucrata* (transformed with *S. medusa CHI*). In contrast, suppression of *CHI* gene by the method of RNAi reduced flavonoid contents and changed flavonoid components resulting in color alteration of flowers and pollens in the transgenic tobacco plants.[103] *CHI* suppression by the antisense method with *S. medusa CHI* sequence also reduced flavonoid levels in the transgenic tobacco plants.[101] These results indicate that CHI plays a significant role in the cyclization reaction of chalcones to form flavanones *in planta*.

Crystal structure of alfalfa CHI protein in complex with (2*S*)-naringenin (**54**) revealed plant-specific three-dimensional structure, an open-faced β-sandwich fold, which resembles an up-down bouquet.[104] The location of **54** in the structure defined the active sites, and the (2*R*)-naringenin (**56**) did not bind in the active sites. Reaction mechanism of CHI was suggested by the structural, mutational, and functional analysis,[104–106] and theoretical study of CHI catalysis has been continuing.[107,108] Although the residues concerning the substrate preference were inferred from sequence comparison of CHIs,[104] the residues that distinguish the difference between type I and type II CHIs are still unclear.

CHI has once been considered as one of the plant-specific genes, based on the apparent lack of similarity to any nonplant sequences or structures;[104] however, CHI activity has also been found in bacteria and fungi recently. The purified enzyme from cell-free extracts of *Eubacterium ramulus* displayed CHI activity and accepted naringenin chalcone (**39**) ($K_m = 43\ \mu mol\ l^{-1}$), isoliquiritigenin (**40**) ($K_m = 14.3\ \mu mol\ l^{-1}$), and butein (**42**) ($K_m = 7.5\ \mu mol\ l^{-1}$) as substrates.[109] This CHI activity was considerably inhibited by *N*-bromosuccinimide, and flavonoids naringenin and phloretin. On the other hand, using the methods of sequence comparison and fold recognition, homologous genes of the plant *CHI* were identified in completely sequenced fungi, slime molds, and gammaproteobacteria.[110] The residues involved in the catalytic function inferred from alfalfa CHI are conserved among the species. In contrast, microbial species that have *CHI*-like genes tend to lack the orthologs of *CHS*. Besides, the sequences of microbial *CHI*s show very low identity (<30%) with plant *CHI*-like genes found in soybean[93] and tomato[111] as well as plant *CHI*s.

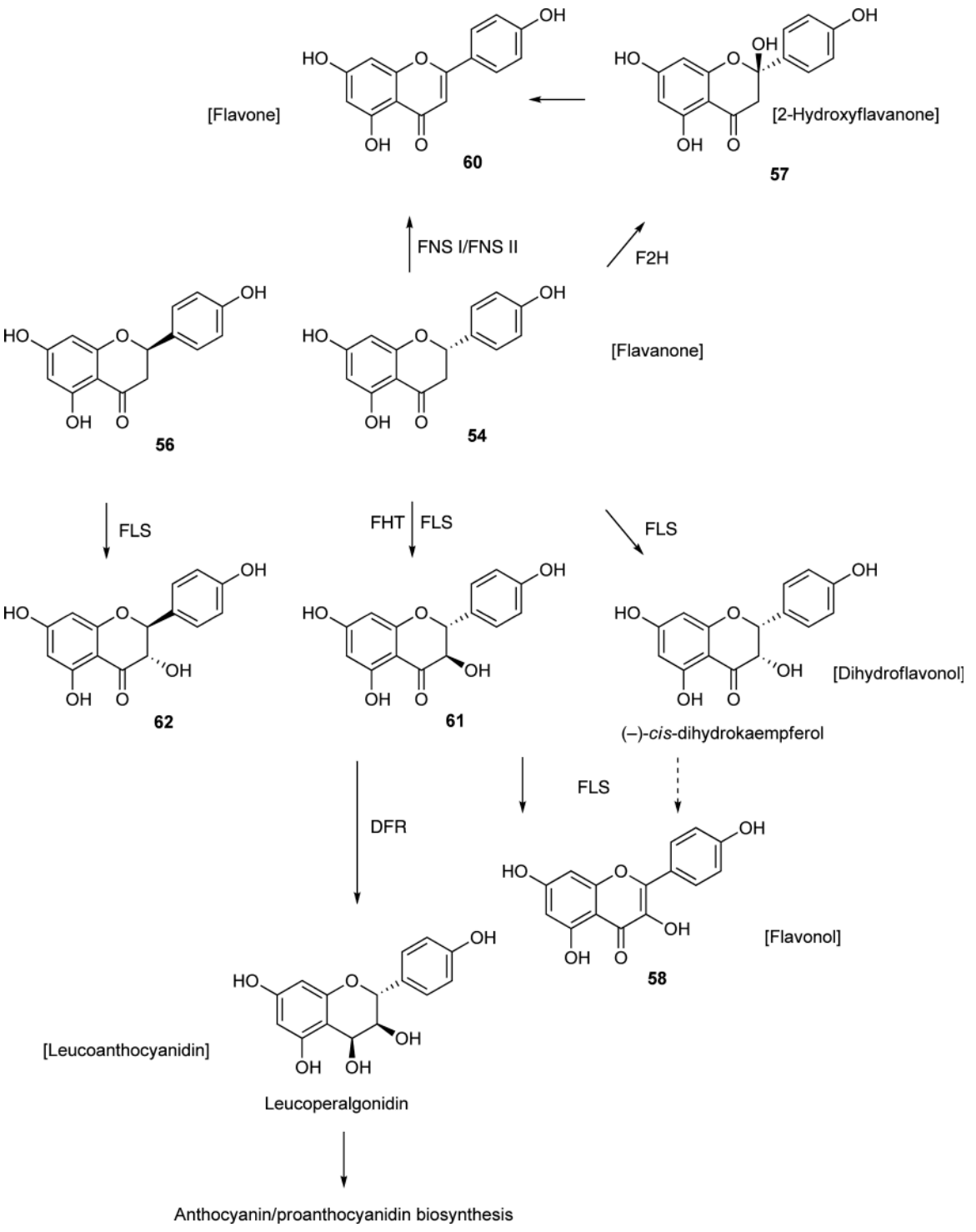
1.24.3.4 Flavones, Flavonols, Anthocyanins, and Proanthocyanidins

1.24.3.4.1 Construction of flavonoid skeletons

All the natural flavonoids having the 2-arylchromone-based skeletons are biosynthesized from a flavanone, most typically (2*S*)-naringenin (**54**). Different skeletons are constructed by several steps of oxidation and/or reduction. The past decade has seen the molecular identification, and detailed characterization in some cases, of the oxidases and reductases involved in the formation of the flavonoid skeletons, including flavone, dihydroflavonol, flavonol, anthocyanidin, and flavan-3-ol.

1.24.3.4.1(i) Cytochrome P-450s responsible for flavone formation (flavanone 2-hydroxylase and flavone synthase II) The biosynthesis of flavones from flavanones is catalyzed by different types of oxidases depending on the plant species (**Scheme 12**). Flavone formation in a wide range of plants is thought to be catalyzed by cytochrome P-450-dependent monooxygenase enzyme system designated flavone synthase II (FNS II).[112] On the other hand, a soluble 2-oxoglutarate-dependent dioxygenase (2-ODD), flavone synthase I (FNS I), has been confined to Apiaceae, and FNS I was identified from parsley, *P. crispum*.[113,114] The high sequence similarity of FNS I to another 2-ODD in the flavonoid pathway, flavanone 3-hydroxylase (FHT), has aroused the interest in the molecular evolution of 2-ODDs in Apiaceae, as discussed below.

A cytochrome P-450 CYP93B1 of a leguminous plant *G. echinata* was shown to be flavanone 2-hydroxylase (F2H).[115] F2H catalyzes the formation of 2-hydroxynaringenin (**57**) and licodione, which is a specific constituent of the limited lineage of the subfamily Papilionoideae of Fabaceae. Whereas 2-hydroxyflavanones are spontaneously dehydrated in the acidic conditions, they are rather stable in the neutral. Thus, the involvement of a dehydratase in the flavone formation can be postulated, but 2-hydroxyflavanone dehydratase has not been found



Scheme 12

so far. On the other hand, CYP93B2 of *Gerbera* hybrid,[116] CYP93B3 of snapdragon (*Antirrhinum majus*), and CYP93B4 of *Torenia* hybrid[117] were shown to directly yield flavones from flavanones, representing FNS II activity, presumably through dehydrogenation of C-2 and C-3 resulting in the formation of C-2–C-3 double bond instead of the dehydration of 2-hydroxyflavanone.[116] FNS II has been found in a number of plants.[112,118,119]

Recently, CYP93B10 and CYP93B11 of barrel medic (*M. truncatula*) were characterized to be F2H. The RNAi experiments provided the genetic evidence that flavones are important for the root nodule formation in *Medicago–Sinorhizobium meliloti* symbiosis.[120] It may be likely that F2H is responsible for the flavone formation in some species, at least, of leguminous plants. Soybean was reported to have the FNS II activity,[112] but an F2H-like sequence, or other CYP93B-like, is missing in the data of the whole genome sequencing program of soybean. An unknown P-450 other than CYP93B subfamily may catalyze the flavone formation in soybean.

1.24.3.4.1(ii) 2-Oxoglutarate-dependent dioxygenases involved in the formation of flavones, flavonols, and anthocyanidins

1.24.3.4.1(ii)(a) Flavanone 3-hydroxylase Although the molecular identification was previously accomplished, several genetic and biochemical findings on FHT have been reported in this decade (**Scheme 12**). Genes of *A. thaliana* essential for the dark brown seed coloration are named the *TRANSPARENT TESTA* (*TT*) genes, for the phenotypes of the knockout mutants, which have been serving as useful tools for studies on flavonoid biosynthesis.[121] Genetic analysis of a mutant collection of *A. thaliana*, generated by the insertion mutagenesis of maize transposable elements *En-1*, revealed that *TT6* encodes FHT.[122]

The estimation of the molecular mass of FHT has suggested monomeric or dimeric enzyme composition, depending on the analytical methods. Recently, thorough examination of the molecular composition of petunia FHT by size exclusion chromatography, protein cross-linking experiments, and equilibrium sedimentation analysis revealed that monomeric polypeptide comprises the functional FHT. At the same time, the tendency of the FHT to aggregate to oligomeric complexes in solution was established.[123] These results may support the idea of the association of FHT in soluble or membrane-associated protein complexes, which is consistent with the assembly of several flavonoid enzymes of *A. thaliana* confirmed by two-hybrid assays.[12]

1.24.3.4.1(ii)(b) Flavonol synthase Flavonols, such as kaempferol (**58**) and quercetin (**59**), are produced by the desaturation of dihydroflavonols, catalyzed by flavonol synthase (FLS; **Scheme 12**). Although FLS was previously known as a member of the 2-ODD family, its detailed characteristics, such as optimum temperature and pH, kinetic parameters, and essential amino acid residues, were uncovered recently.[124] The formation of dihydroflavonols was thought to be catalyzed by FHT, but functional characterization of FLS from *A. thaliana* and *Citrus unshiu* revealed that FLS also has the activity of the 3-hydroxylation of flavanones to produce dihydroflavonols, that is, similar activity to FHT.[125,126] Whereas all the FHT so far studied recognize only (2*S*)-naringenin (**54**) as the substrates, FLS was shown to possess the hydroxylation activity toward both **54** and (2*R*)-naringenin (**56**). However, the stereospecificity of the hydroxylation and desaturation activities is different among the plant species. The FLS from *A. thaliana* has the activity of both 3α- and 3β-hydroxylation of **54** and 3β-hydroxylation of **56**, as well as weak activity of desaturation to produce a flavone, apigenin (**60**).[125] On the other hand, *C. unshiu* FLS catalyzes 3β-hydroxylation of (2*S*)-naringenin (**54**) and 3α-hydroxylation of (2*R*)-naringenin (**56**) to produce (+) (**61**) and (−) *trans*-dihydrokaempferol (**62**), respectively.[126]

Recently, FLS was shown to be responsible for the flower color of soybean, possibly because of the copigmentation effect of flavonols.[127] Soybean with the dominant allele *Wm* has purple flower color, while the recessive *wm* allele yields magenta flower color. The genetic and biochemical analysis revealed that the *Wm* gene encodes FLS and that the polypeptide encoded by the *wm* allele lacks the FLS activity. A single-base deletion in the *wm* allele causes a frameshift resulting in a truncated polypeptide lacking the dioxygenase domains.[127]

An *En*-induced FLS knockout mutant of *A. thaliana* showed abolished quercetin (**59**) accumulation but wild-type level of kaempferol (**58**), which suggested the existence of another FLS.[122] Recently, five FLS-like genes, which were predicted in the genome of *A. thaliana*, were functionally characterized, but only one gene that had been already known showed the FLS activity.[128] These paralogous genes differently expressed in organs, and their protein products showed interaction with other flavonoid enzymes in yeast two-hybrid experiments. It was thus suggested that the FLS paralogues are not simple pseudogenes but play some regulatory roles in the flavonoid biosynthesis by enzyme–enzyme interactions.[128]

1.24.3.4.1(ii)(c) Anthocyanidin synthase (synonym of leucoanthocyanidin dioxygenase) Anthocyanins constitute a major flavonoid class in higher plants and are responsible for the brick red, red, and blue colors of flowers, and attract pollinators. The anthocyanin biosynthesis is one of the most extensively studied

pathways of natural products, and a group of 2-ODD-related proteins were postulated to be the enzyme designated anthocyanidin synthase (ANS) responsible for the conversion of leucoanthocyanidins, such as leucopelargonidin (**63**), leucocyanidin (**64**), and leucodelphinidin (**65**), into anthocyanidins, such as pelargonidin (**66**), cyanidin (**67**), and delphinidin (**68**), respectively (**Scheme 13**). However, no biochemical information of ANS had been obtained at the beginning of this decade.

Biochemical characterization of the heterologously expressed ANS of *Perilla frutescens*, whose reddish-purple form highly accumulates anthocyanins in the whole plant, showed the direct evidence that ANS is a 2-ODD.[129] The first stable colored metabolite detected in plants was anthocyanin 3-glucosides, in which biosynthesis ANS and UDP-glucose: flavonoid 3-*O*-glucosyltransferase (3-GT) is involved. *In vitro* formation of cyanidin 3-*O*-glucoside (**69**) via cyanidin (**67**) was demonstrated by incubation with recombinant ANS and 3-GT, showing no requirement of any other enzymes such as dehydratase but necessity of acidic condition after the enzymatic reactions.[130] Meanwhile, it was revealed that ANS does not simply convert leucoanthocyanidins to anthocyanidins but also to flavonols via dihydroflavonols.[131] The initial reaction of ANS was shown to be oxidation at C-3 of leucoanthocyanidins and flavanones.[132] The C-2 stereochemistry of flavanone substrates has a significant effect on the product selectivity.[133] The mechanistic investigation using density functional theory method indicated that the C-3 and C-4 positions of leucoanthocyanidins can be oxidized but that C-4 oxidation is preferable, suggesting that the C-3 oxidation, a key step to the anthocyanidin biosynthesis, can be a side reaction of ANS.[134] Assays with recombinant ANS from *Gerbera* hybrid using catechin (**70**) and epicatechin (**71**) isomers as model substrates demonstrated that ANS recognizes the substrate with a 3β-hydroxy group and catalyzes 3α-hydroxylation to form 3-geminal diol.[135]

Anthocyanidin synthase knockout mutants of *A. thaliana* were independently isolated by screening of carbon-ion-irradiated and T-DNA-tagged mutants and designated *transparent testa* (*tt*) *18* and *tannin deficient seed* (*tds*) *4*, respectively.[136,137] The phenotypic analysis of the mutants shows that ANS plays a part in proanthocyanidin (condensed tannin) pathway, as well as in anthocyanin formation.[137] The occurrence of the *ANS* genes in Caryophyllales plants that do not produce anthocyanins and their expression in developing seeds, which accumulate proanthocyanidins, are well consistent with the role of ANS in the proanthocyanidin pathway suggested by the phenotype of the knockout mutants.[138]

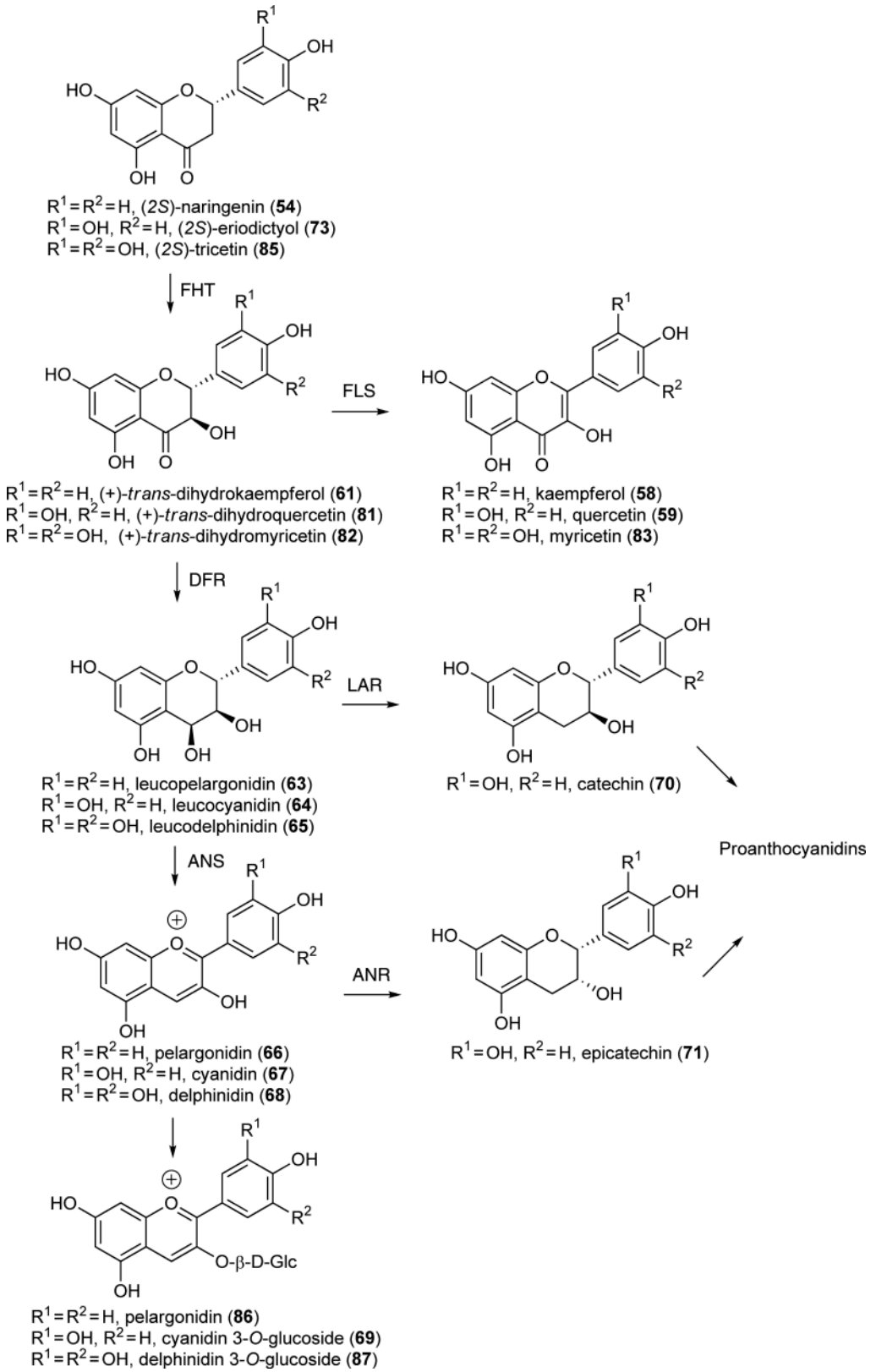
1.24.3.4.1(ii)(d) Molecular evolution of the related 2-oxoglutarate-dependent dioxygenases The 2-ODDs involved in flavonoid biosynthesis, such as FHT, FLS, ANS, and FNS I in Apiaceae, share considerable sequence identities and overlapping substrate and product selectivity, as described above. The mechanisms and molecular evolution of these 2-ODDs have recently been investigated.

cDNAs for FHT and FLS were cloned from parsley, and the recombinant protein products were compared to the recombinant FNS I from the same plant. The similar substrate specificities of FNS I and FHT suggested their close relationship, in spite of the equivalent desaturation reactions catalyzed by FNS I and FLS.[139]

The reaction mechanisms of ANS, FLS, and FHT were studied in detail under the atmosphere containing $^{18}O_2$ and/or $^{18}OH_2$.[140] The results obtained in the study and previous ones together suggest that the reaction of the 2-ODDs, ANS, FLS, FHT, and FNS, starts with the oxidation at C-3 of the substrates, although the possibility of C-2 or C-4 oxidation is not excluded in some case. The results also support the idea that ANS/FLS and FHT/FNS are the pairs of common substrate specificity and stereoselectivity. ANS/FLS and FHT/FNS pairs may be categorized into α- and β-face-selective oxygenases, respectively.[140]

cDNAs for FNS I and FHT were identified from several Apiaceae species, and their sequences were used for the phylogenetic analysis of 2-ODDs of flavonoid pathways.[141] In the phylogenetic tree, there are three clades, such as ANS, FLS, and FHT/FNS, and ANS/FLS and FHT/FNS shows relatively distant relationship. Apiaceae FHT and FNS I form a monophyletic group, suggesting that the *FNS I* gene was generated by a gene duplication of the *FHT* gene in an ancestral Apiaceae followed by functional diversification.[140]

1.24.3.4.1(iii) Reductases Recent progresses concerning the reductase family proteins in the flavonoid pathway include elucidation of the genomic organization of a gene family of a key biosynthetic enzyme in the anthocyanin biosynthesis and molecular identification of two types of (epi)catechin-forming enzymes.


R1 = R2 = H, (2S)-naringenin (54)
R1 = OH, R2 = H, (2S)-eriodictyol (73)
R1 = R2 = OH, (2S)-tricetin (85)
FHT
R1 = R2 = H, (+)-trans-dihydrokaempferol (61)
R1 = OH, R2 = H, (+)-trans-dihydroquercetin (81)
R1 = R2 = OH, (+)-trans-dihydromyricetin (82)
FLS
R1 = R2 = H, kaempferol (58)
R1 = OH, R2 = H, quercetin (59)
R1 = R2 = OH, myricetin (83)
DFR
R1 = R2 = H, leucopelargonidin (63)
R1 = OH, R2 = H, leucocyanidin (64)
R1 = R2 = OH, leucodelphinidin (65)
LAR
R1 = OH, R2 = H, catechin (70)
ANS
R1 = R2 = H, pelargonidin (66)
R1 = OH, R2 = H, cyanidin (67)
R1 = R2 = OH, delphinidin (68)
ANR
R1 = OH, R2 = H, epicatechin (71)
Proanthocyanidins
O-β-D-Glc
R1 = R2 = H, pelargonidin (86)
R1 = OH, R2 = H, cyanidin 3-O-glucoside (69)
R1 = R2 = OH, delphinidin 3-O-glucoside (87)

Scheme 13

1.24.3.4.1(iii)(a) Genomic organization of the genes encoding dihydroflavonol 4-reductase Dihydroflavonol 4-reductase (DFR) is a pivotal enzyme of the flavonoid pathway leading to common anthocyanins and proanthocyanidins. This enzyme yields flavan 3,4-diols by the reduction of corresponding dihydroflavonols, which are also intermediates of flavonol biosynthesis through the FLS reaction (**Scheme 13**). The *DFR* genes of many plants form multigene families caused by gene duplication in the evolutional processes of higher plants, and in many cases the putative duplicated genes are expressed.[142–145] A comprehensive analysis of a model legume *L. japonicus* revealed that a total of five DFR genes, all of which are differently expressed in organs, form a cluster within a 38-kb region: three genes encode active DFRs, one yields unusual mRNAs due to a mutation at a splicing accepter site, and one encodes an inactive enzyme in spite of no apparent structural aberration. The three active DFRs have distinct preference of substrates with different hydroxylation pattern of the B ring. Phylogenetic analysis suggests that the gene duplication in part occurred prior to speciation of *L. japonicus* and other species.[146] The adaptive significance of the duplicated paralogous DFR in the evolution of higher plants is intriguing.

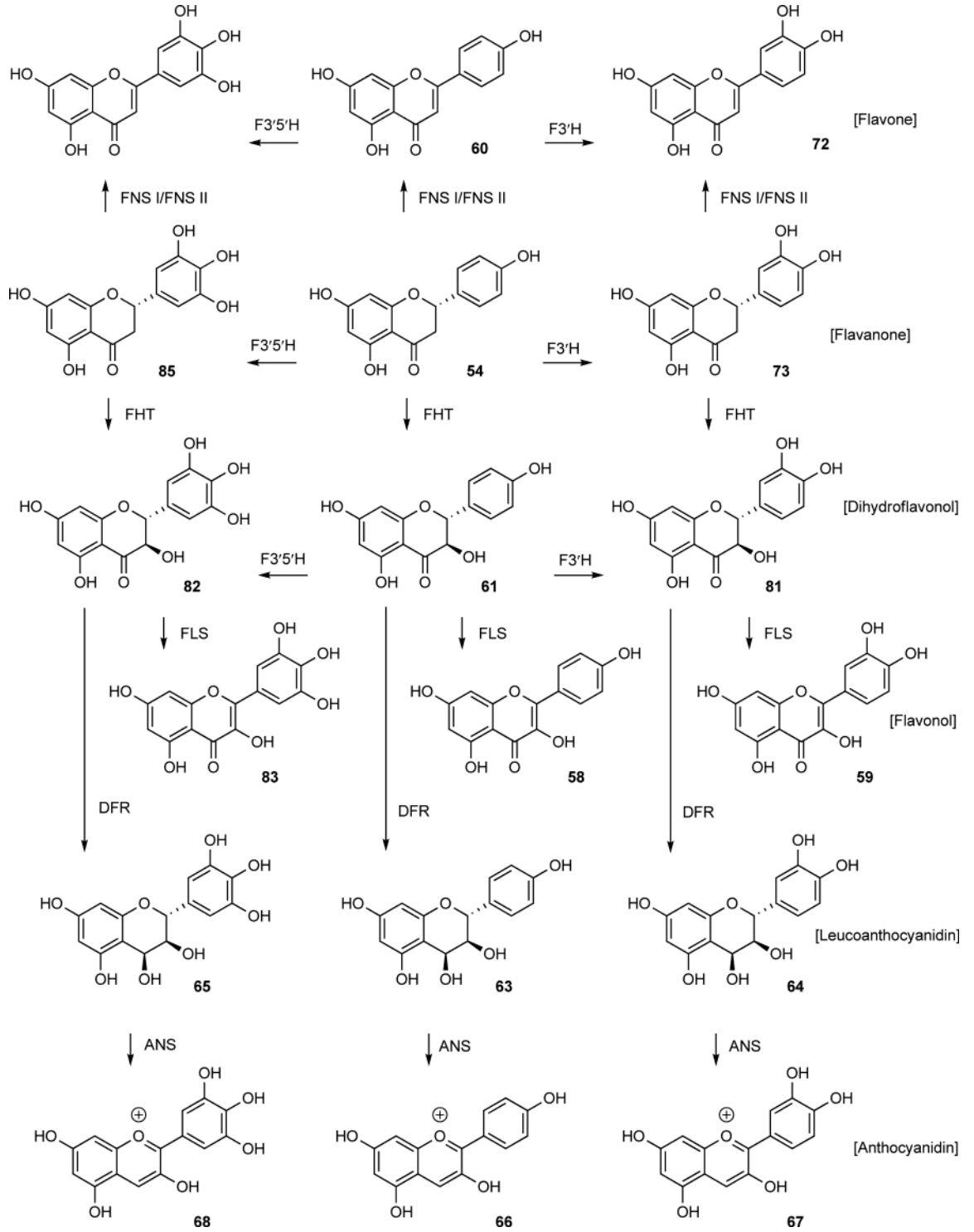
1.24.3.4.1(iii)(b) Reductases in flavan3-ol biosynthesis, anthocyanidin reductase, and leucoanthocyanidin reductase Proanthocyanidins or condensed tannins are flavonoid oligomers that are widely distributed in plants. They strongly interact with proteins and have beneficial effects on animal and human health.[147] Flavan-3-ols, such as catechin (**70**) and epicatechin (**71**), are components of proanthocyanidins and serve as flavor and astringency in tea and wine. To date, two types of reductases have been reported to be involved in flavan-3-ol formation (**Scheme 13**). The *BANYULS* (*BAN*) gene of *A. thaliana*, whose knockout mutation causes the precocious accumulation of red anthocyanins and loss of proanthocyanidins in the seed coat, was known to encode a reductase.[148] Recently, the *BAN* gene and its orthologous gene of a model legume, *M. truncatula*, were shown to encode anthocyanidin reductase (ANR).[149,150] ANR converts anthocyanidins to their corresponding 2,3-*cis*-flavan-3-ols. The other route is the reduction of flavan-3,4-diols, such as leucocyanidin (**64**), to produce 2,3-*trans*-flavan-3-ols, such as catechin (**70**). The enzyme catalyzing this reaction, LAR, was first identified from *Desmodium uncinatum*.[151] Although activity has been detected in many plant species, there is no obvious orthologous gene in the genome of *A. thaliana*. The significance of the two partially redundant pathways has not been clarified until now. Genes for LAR and ANS were identified from *M. truncatula*, and their functional characterization and tissue-specific expression were reported.[152] The transcript level of the *LAR* gene in developing seeds did not parallel the accumulation of proanthocyanidins, and transgenic tobacco plants overexpressing the *LAR* gene did not bring about the increase in catechin or proanthocyanidin content. The function of LAR in *Medicago* remained to be addressed.[152]

1.24.3.4.2 Decoration enzymes

1.24.3.4.2(i) Prenyltransferases A leguminous plant *Sophora flavescens* produces specific prenylated flavonoids that possess a lavandulyl group, an irregular monoterpenoid unit, at C-6 or C-8 position of 2′-hydroxyflavanone. The lavandulyl group is thought to be formed by two successive prenylation of a flavanone. The cDNA encoding naringenin 8-prenyltransferase, responsible for the first prenylation, was recently identified[19] and are expected to serve as a useful tool to identify other flavonoid prenyltransferases (see Section 1.24.3.5.2(ii)).

1.24.3.4.2(ii) Hydroxylases

1.24.3.4.2(ii)(a) Flavonoid 3′-hydroxylase and flavonoid 3′,5′-hydroxylase The hydroxylation pattern of the B ring is one of the significant origins of flavonoid diversity (**Scheme 14**). Flavonoid 3′-hydroxylase (F3′H) is essential for the *ortho*-dihydroxy (catechol) structure of the B ring of many classes of flavonoids, including cyanidin (**67**; anthocyanidin), luteolin (**72**; flavone), quercetin (**59**; flavonol), and (epi)catechin (**70**, **71**; flavan-3-ol). F3′H, as well as flavonoid 3′,5′-hydroxylase (F3′5′H), has been of interest particularly because the B ring hydroxylation catalyzed by the enzymes is a significant determinant of petal color. Although the activities of both F3′H and F3′5′H were previously detected and known as microsomal P-450s, molecular identification of F3′H was preceded by that of F3′5′H. The first cloned cDNA for F3′H corresponds to the *Ht1* locus of *Petunia* ×*hybrida* controlling the color of the limb and tube of the corolla, and the heterologously expressed recombinant F3′H was confirmed to convert naringenin (**54**) to eriodictyol (**73**).[153] F3′H of *A. thaliana* was reported to be encoded by the *TT7* gene, suggesting that F3′H is responsible for the brown color of the seed coat, as well as anthocyanin accumulation.[154]


F3′5′H
F3′H
FNS I/FNS II
FHT
FLS
DFR
ANS
[Flavone]
[Flavanone]
[Dihydroflavonol]
[Flavonol]
[Leucoanthocyanidin]
[Anthocyanidin]
60
72
85
54
73
82
61
81
83
58
59
65
63
64
68
66
67

Scheme 14

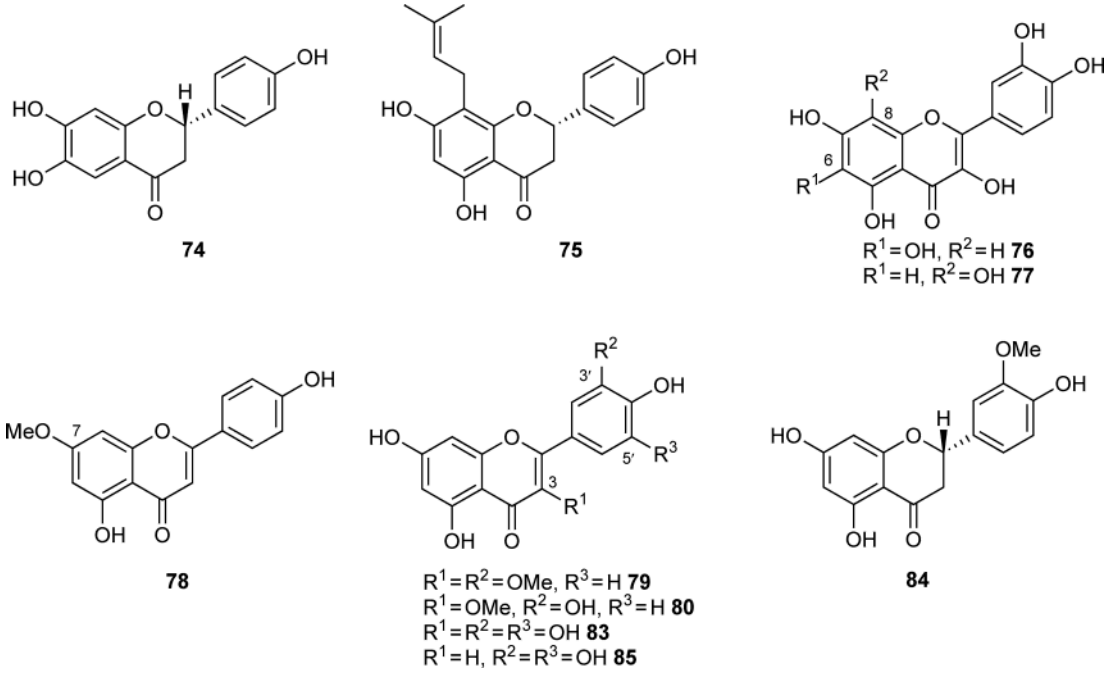
Perilla frutescens is well known to include two distinct forms on the anthocyanin accumulation in the whole plant, reddish-purple and green ones, and served as sources for cloning of several biosynthetic genes of flavonoids.[118,129,155,156] F3′H was also identified from *P. frutescens*, and the recombinant F3′H catalyzed *in vitro* the 3′-hydroxylation of naringenin (**54**), dihydrokaempferol (**61**), and apigenin (**60**). The F3′H gene was revealed to be regulated coordinately with other anthocyanin biosynthetic genes.[118]

The pleiotropic *T* locus of soybean controls pubescence and seed coat color, chilling tolerance, and structural integrity of the seed coats. The long-standing presumption that the *T* locus should encode F3′H, which was suggested by chromatographic experiments using seeds of several genotypes, was confirmed by genetic analyses. Furthermore, the recessive alleles were shown to have one-base deletion resulting in the truncated F3′H protein that lacks the heme-binding domain essential for P-450. The molecular identification of soybean F3′H will allow further dissection of the pleiotropic effects associated with the *T* locus, such as chilling tolerance and structural integrity of the seed coat as well as pigmentation.[157,158]

The sequence identity between F3′H and F3′5′H, both of which determine the B ring hydroxylation pattern of flavonoids, is generally low, implying their divergence early in the evolution of higher plants. The phylogenetic analysis showed that most of the F3′5′H form a distinct clade but that some Asteraceae F3′5′H are not included in the F3′5′H clade but in the distinct angiospermic F3′H clade. It was suggested that Asteraceae F3′5′H were generated by gene duplication of an F3′H gene in an ancestral Asteraceae, independently of other plants' F3′5′H.[159] Domain swapping experiments and site-directed mutagenesis based on the cDNAs for *Osteospermum* hybrid F3′5′H and *Gerbera* hybrid F3′H showed that relatively few amino acid residues were required for the functional exchange from F3′H to F3′5′H activity.[160]

1.24.3.4.2(ii)(b) Other hydroxylases of flavonoid skeleton Soybean produces isoflavonoid constituents possessing 6,7-dihydroxy substitution patterns on ring A. A novel P-450 (CYP71D9) was shown to be flavonoid 6-hydroxylase (F6H), which effectively hydroxylates flavanones (**Scheme 15**). The resultant 6,7,4′-trihydroxyflavanone(**74**) is then used as a substrate of the enzyme catalyzing the 1,2-aryl migration of the B ring to form the isoflavonoid skeleton (see Section 1.24.3.5.1(iv)).[161]

The 2′-hydroxylase involved in the biosynthesis of lavandulylated flavanones in *S. flavescens*, 8-dimethylallylnaringenin 2′-hydroxylase, was detected in cell suspension cultures and was suggested to be



Scheme 15

a P-450 by the cofactor requirement and inhibitory effects of P-450 inhibitors. The enzyme hydroxylated only 8-dimethylallylnaringenin (**75**) but does not hydroxylate the other flavonoids tested, such as naringenin (**54**) and 6-dimethylallylnaringenin. These results suggest that 2′-hydroxylation catalyzed by the P-450 occurs between two prenylations.[162] The naringenin 8-prenyltransferase, responsible for the first prenylation, was shown to be localized to plastid,[19] whereas most of the P-450s are associated with the endoplasmic reticulum. The reaction site of the second prenylation is intriguing.

The most common flavonoids have hydroxy groups in position 5 and 7 of the aromatic A ring. Yellow flavonoids such as quercetagenin (6-hydroxyquercetin) (**76**), gossypetin (8-hydroxyquercetin) (**77**), and their derivatives show additional hydroxy groups in position 6 and/or 8. The methylated and/or glucosylated derivatives of such flavonols with extra hydroxyls on A ring occur in Asteraceae, Malvaceae, and Fabaceae. An enzyme catalyzing 8-hydroxylation of flavonols and flavones was recently characterized. The enzyme was localized in the microsomal fraction and uses NADPH and FAD as cofactors, but inhibition experiments with carbon monoxide/blue light and antibodies specific to P-450 reductase did not show the involvement of a classical P-450.[163]

1.24.3.4.2(iii) O-Methyltransferases The basic flavonoid structure is composed of two benzene rings (A and B rings) connected with a C-3 chain, which forms the additional C ring. To our best knowledge, enzymatic *O*-methylations of A, B, and C rings are catalyzed by distinct classes of OMTs. Among the OMTs mentioned below, flavonoid 7-*O*-methyltransferase and flavonol 3-*O*-methyltransferase are involved in A ring and C ring methylation, respectively, and the others are OMTs of B ring methylation (**Scheme 15**). In this decade, the diversity of OMTs has been uncovered, but probably only in part.

A pathogen-induced OMT previously isolated from barley (*H. vulgare*) was shown to be flavonoid 7-*O*-methyltransferase. The enzyme has the highest activity toward apigenin (**60**) to produce genkwanin (7-*O*-methylapigenin) (**78**). Immunological analyses showed that the OMT is absent in normal leaves but accumulated strongly in response to the attack by the pathogenic fungus *Blumeria graminis*. This OMT is suggested to be involved in the production of a methylated flavonoid phytoalexin.[164]

Serratula tinctoria (Asteraceae) accumulates 3,3′-di-*O*-methylquercetin (**79**) and 3-*O*-methylquercetin (**80**), as major and minor constituents, respectively. The flavonol 3-*O*-methyltransferase was partially purified and its characteristics were investigated. The OMT exhibited specificity for the methylation of position 3 of quercetin (**59**), although it also utilized kaempferol (**58**), myricetin (**83**), and some monomethylflavonols as substrates.[165]

It has been believed that distinct OMTs are involved in the *O*-methylation of phenylpropanoids and flavonoids. However, two cDNAs isolated from *Chrysosplenium americanum* were shown to encode OMTs for the methylation of both flavonoid and phenylpropanoid compounds: they catalyze 3′-*O*-methylation of quercetin (**59**) and luteolin (**72**), and also 3/5-*O*-methylation of caffeic and 5-hydroxyferulic acids.[166]

A cDNA clone previously isolated from *A. thaliana* annotated as encoding a caffeic acid/5-hydroxyferulic acid OMT was identified to encode a novel flavonol 3′-*O*-methyltransferase. The OMT used quercetin (**59**) as the preferred substrate but no phenylpropanoid substrate tested, in spite of 80% identity of amino acid sequence to the aspen bispecific lignin OMT.[167] An OMT isolated from rice (*Oryza sativa*) that showed a 73% identity to caffeic acid OMTs from maize and wheat, was reported to transfer a methyl group specifically to the 3′-hydroxy group of eriodictyol (**73**), luteolin (**72**), quercetin (**59**), and dihydroquercetin (**81**).[168]

Catharanthus roseus produces flavonoids with a simple methylation pattern. A total of five cDNAs encoding OMTs have been analyzed. CrOMT2 represented a new type of OMT that catalyzes two sequential methylations at the 3′- and 5′-positions of the B ring of dihydromyricetin (**82**) and myricetin (**83**). Since the resulting 3′, 5′ dimethoxy pattern is characteristic for the flavonols and anthocyanins of this plant, CrOMT2 is postulated to be involved in their biosyntheeis.[169] CrOMT6 was found to be another active OMT of novel specificity.[170] It catalyzes 4′-*O*-methylation of a flavanone 3′-*O*-methyleriodictyol (**84**) and the corresponding flavones and flavonols. In order to address the question whether methylation influences channeling flavonoid biosynthesis, the activities of four recombinant 2-ODDs, that is, FHT, FNS I, FLS, and ANS, toward 3′-*O*-methyleriodictyol (**84**) were investigated. The results suggested that B ring 3′-methylation has little inhibitory effect on flavonoid 2-ODDs.[170]

An OMT that catalyzes a sequential *O*-methylation of flavonoids was identified from wheat, *Triticum aestivum*. It uses several flavones, flavonols, dihydroxyflavonols, and 5-hydroxyferulic acid as substrates, and the most preferred is tricetin (3′,4′,5,5′,7-pentahydroxyflavone) (**85**). The sequential order of tricetin methylation is presumed to proceed via its 3′-mono, 3′,5′-di, and 3′,4′,5′-trimethyl ether derivatives.[171]

1.24.3.4.2(iv) Glycosyltransferases Many thousands of plant chemicals are glycosylated. They are often postulated to be transportable storage compounds or detoxified products because glycosylation results enhanced water solubility and lower chemical reactivity. The enzymes responsible for flavonoid glycosylation so far identified utilize UDP-glycosides, such as UDP-glucose, UDP-galactose, and UDP-rhamnose, as glycosyl donors. UDP-glycosyltransferases (UGTs), which are contained in the family 1 of the 54 families of the glycosyltransferase superfamily, have been identified in plants, animals, fungi, and bacteria.[172]

Glycosyltransferases generally have a common motif, termed UDP-glycosyltransferase-specific conserved region (UDPGT), which is particularly well conserved in the glycosyltransferases involved in the plant secondary metabolism. The UDPGT motif has been thought to correspond to a UDP-binding site, and one amino acid residue (His or Gln), positioned at the end of the UDPGT motif, was found to be highly conserved among the plant glucosyltransferases and galactosyltransferases. A single point mutation replacing His with Gln of UDP-galactose: anthocyanin 3-*O*-galactosyltransferase isolated from Japanese asparagus (*Aralia cordata*) altered the glycosyl donor specificity; the enzyme acquired glucosyltransferase activity in addition to the original galactosyltransferase activity.[173]

O-Glycosylated flavonoids are widely distributed to plants, and the major glycosyltransferases so far identified were *O*-glycosyltransferases. However, several species accumulate flavone-*C*-glycosides. Recently, enzymes that catalyze UDP-glucose-dependent *C*-glucosylation of 2-hydroxyflavanone (**57**) were identified from rice and wheat.[174]

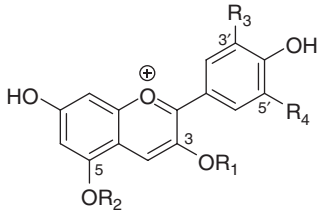
1.24.3.4.2(iv)(a) Glucosyltransferases After the first molecular identification of a flavonoid glucosyltransferase by transposon tagging of the maize *Bronze-1* locus in 1984, cDNAs encoding glucosyltransferases have been isolated from a number of plant species based on the sequence information. However, glucosyltransferase had not been subjected to a detailed characterization using purified proteins, and thus a clear picture of the glucosyltransferase properties was not available, in spite of reports on characterization of transcript levels and on the ectopic overexpression and the knocking down of the glucosyltransferase genes. Remarkable progresses in the studies on glucosyltransferase in this decade should be due to careful characterization using heterologous expression systems with yeast and *E. coli* cells.[175] Extensive *in vitro* characterization of cDNAs, isolated using several strategies, have provided not only a better understanding of the well-known UDP-glucose: flavonoid 3-*O*-glucosyltransferase[175] but also opportunities for the identification of various new glucosyltransferases as described below (**Scheme 16**).

A cDNA encoding UDP-glucose: anthocyanin 5-*O*-glucosyltransferase (5-GT) was for the first time identified by mRNA differential display using two forms of *P. frutescens*, that is, reddish-purple and green leaves ones.[155] The cDNA exhibits a low sequence homology with other glucosyltransferases including UDP-glucose: anthocyanin 3-*O*-glucosyltransferase. The cDNA was then used as the probe to isolate an orthologous cDNA clone from *Verbena* × *hybrida*.[155] The recombinant enzyme was extracted from yeast cells expressing the cDNA and shown to convert pelargonidin, cyanidin, and delphinidin 3-*O*-glucosides (**86**, **69**, **87**) into the corresponding anthocyanidin 3,5-di-*O*-glucosides (**88**–**90**). Its biochemical properties were similar to those previously reported for 5-GT.

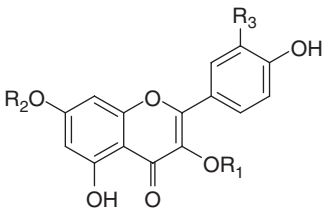
The cDNA encoding 5-GT, together with that encoding UDP-glucose: flavonoid 3-*O*-glucosyltransferase, was also identified from *P. hybrida* by hybridization screening of a cDNA library with heterologous probes, followed by the biochemical characterization of the protein products heterologously expressed in yeast cells. The 5-GT from *P. hybrida* exhibited a strict substrate specificity toward anthocyanidin 3-acylrutinoside, whereas the 5-GT from *P. frutescens* and *V. hybrida* showed broad substrate specificities toward several anthocyanidin 3-*O*-glucoside.[176]

The petals of *Gentiana triflora* var. *japonica* predominantly contain an unusual blue and stable anthocyanin, gentiodelphin, which is a derivative of 3,5,3′-tri-*O*-glucosyl delphinidin (**91**). In the biosynthesis of gentiodelphin, glucosylation and subsequent acylation of the 3′-hydroxy group of the B ring of delphinidin takes place, which is believed to be particularly important for the stability and the blue color. UDP-glucose: anthocyanin 3′-*O*-glucosyltransferase was purified to homogeneity from the blue petals of *G. triflora*, and the corresponding cDNA was isolated based on the internal amino acid sequences of the purified protein. Characterization of the enzymatic properties using the recombinant enzyme showed strict substrate specificity: it specifically uses delphinidin-type anthocyanins containing glucose groups at C-3 and C-5 positions as glucosyl acceptors.[177]

Flower pigments of rose (*Rosa hybrida*) are mainly composed of the structurally simple cyanidin 3,5-di-*O*-glucoside (**89**), in spite of the wide variety of flower colors. The glucosyltransferase responsible for the cyanidin 3,5-di-*O*-glucoside (**89**) formation was characterized by biochemical assays using both the enzyme



	R^1	R^2	R^3	R^4
86	glucose	H	H	H
69	glucose	H	OH	H
87	glucose	H	OH	OH
88	glucose	glucose	H	H
89	glucose	glucose	OH	H
90	glucose	glucose	OH	OH
91	glucose	glucose	*O*-glucose	OH
92	H	glucose	OH	H
94	6″-*O*-malonylglucose	H	OH	OH
95	6″-*O*-malonylglucose	H	*O*-glucose	H
96	6″-*O*-malonylglucose	H	*O*-glucose	*O*-glucose
104	glucose	caffeoylglucose	OH	OH
105	caffeoylglucose	H	OH	H
106	caffeoylglucose	glucose	OH	H
107	caffeoylglucose	dimalonylglucose	H	H
108	*p*-coumaroylglucose	glucose	OH	H
109	caffeoylglucose	glucose	H	H
110	caffeoylglucose	malonylglucose	H	H
111	malonylglucose	H	H	H
112	malonylglucose	H	OH	H
113	rutinose (rhamnosylglucose)	H	OH	H



	R^1	R^2	R^3
93	glucose	H	OH
97	rhamnose	rhamnose	H
98	rhamnose	rhamnose	OH
99	rhamnose	glucose	OH
100	glucose	H	H
101	robinobiose (rhamnosylgalactose)	rhamnose	H
102	galactose	H	H
103	galactose	H	OH
114	H	glucose	H

Scheme 16

preparation of the flower extract and the recombinant proteins produced by *E. coli* expressing a cDNA clone. In the assays, glycosylation occurs first at the 5-hydroxy and then at the 3-hydroxy groups of cyanidin. It uses either cyanidin (**67**) or cyanidin 5-*O*-glucoside (**92**) as a glucosyl acceptor, but not cyanidin 3-*O*-glucoside (**69**). Since cyanidin (**67**) and cyanidin 5-*O*-glucoside (**92**) are unstable without the additional glycosylation of 3-hydroxy group, the first stable anthocyanin in rose is cyanidin 3,5-di-*O*-glucoside (**89**).[178]

cDNAs encoding UDP-glucose: flavonoid 7-*O*-glucosyltransferase of different substrate specificity and regioselectivity have been cloned from onion (*Allium cepa*) and red beet (*Beta vulgaris*). Two flavonoid glucosyltransferases were identified from *A. cepa*. One glucosyltransferase (UGT73G1) used a wide range of flavonoid substrates, and the regioselectivity was indicated for the C-3, C-7, and C-4′ of the flavans to yield flavonoid mono- and diglucosides. The other *A. cepa* glucosyltransferase (UGT73J1) used only a flavonol monoglucoside isoquercitrin (**93**) and an isoflavone aglycone genistein (see Section 1.24.3.5) as the glucosyl acceptors with regioselectivity for the C-7 position.[179] The

two glucosyltransferases from *B. vulgaris* use a variety of flavonols, flavones, flavanones, and coumarins as substrates. One of the *B. vulgaris* glucosyltransferases (UGT73A4) showed a preference for the C-4′ and C-7 positions, while the other (UGT71F1) preferentially glucosylated the C-3 and C-7 positions.[180]

Clitoria ternatea (Fabaceae) accumulates polyacylated anthocyanins termed ternatins. In the biosynthesis of ternatins, delphinidin 3-*O*-(6″-*O*-malonyl)-*β*-glucoside (**94**) is first converted to delphinidin 3-*O*-(6″-*O*-malonyl)-*β*-glucoside-3′-*O*-*β*-glucoside (**95**), and it is then converted to delphinidin 3-*O*-(6″-*O*-malonyl)-*β*-glucoside-3′,5′-di-*O*-*β*-glucoside (ternatin C-5) (**96**). The enzyme responsible for the two steps of B ring glucosylation, UDP-glucose: anthocyanin 3′,5′-*O*-glucosyltransferase, was purifed from the petals of *C. ternatea*. The biochemical characterization revealed its regioselective glucosylation activities of 3′ or 5′ OH groups of anthocyanins.[181]

(−)-Epicatechin (**71**) is a precursor of proanthocyanidins. An epicatechin-specific glucosyltransferase expressed in the seed coat of *M. truncatula* was identified by a transcript profiling approach.[182]

***1.24.3.4.2(iv)(b) Glycosyltransferases from* A. thaliana** Since *A. thaliana* is the first exploited model plant, various research resources for forward and reverse molecular genetics and genomics are available. *A. thaliana* accumulates many flavonol glycosides, which constitute one of the most prominent plant products. Rhamnosyltransferases were for the first time identified from this plant using functional genomic approaches, together with several types of glucosyltransferases.

Candidate genes for glycosyltransferase were selected based on the sequence homology to other known flavonoid glycosyltransferases, and the T-DNA insertional knockout mutants that lack kaempferol and quercetin 3,7-di-*O*-rhamnosides (**97, 98**) and quercetin-3-*O*-rhamnoside-7-*O*-glucoside (**99**) were screened. The protein products of the genes isolated were biochemically identified using the recombinant proteins expressed in *E. coli*. As a result, UGT78D1 and UGT73C6 were shown to be UDP-rhamnose: flavonol-3-*O*-rhamnosyltransferase and UDP-glucose: flavonol-3-*O*-glycoside-7-*O*-glucosyltransferase, respectively.[183]

An activation-tagged mutant of bright purple phenotype termed *pap1-D*, in which the *PAP1* gene encoding an MYB transcription factor is overexpressed and consequently a series of anthocyanin biosynthetic genes are upregulated,[184] serves as a powerful tool for the studies on anthocyanin biosynthesis in *A. thaliana*. The transcriptomic and metabolomic analyses suggested two putative UGTs, At5g17050 and At4g14090, and their functions were confirmed to be UDP-glucose: flavonoid 3-*O*-glucosyltransferase and UDP-glucose: anthocyanin 5-*O*-glucosyltransferase, respectively, by *in vitro* assays using the recombinant enzymes.[185]

A rhamnosyltransferase essential for the flavonoid composition in *A. thaliana* was identified by transcriptome coexpression analysis and reverse genetics.[186] Transcriptome coexpression analysis on the ATTED-II public database revealed that the expression pattern of a UGT gene, *UGT89C1*, is highly correlated with known flavonoid biosynthetic genes. Two T-DNA insertional *ugt89c1* mutants exhibited no C-7 rhamnosylated flavonols, whose metabolite deficiency was complemented by stable transformation with the genomic fragment containing intact *UGT89C1*. The recombinant UGT89C1 transferred rhamnosyl moiety at the 3-*O* position of kaempferol 3-*O*-glucoside (**100**). This enzyme recognized 3-*O*-glycosylated flavonols and UDP-rhamnose as substrates but did not recognize flavonol aglycones, 3-*O*-glycosylated anthocyanins or other UDP-sugars. These results show that UGT89C1 is a flavonol 7-*O*-rhamnosyltransferase.[186]

Transcriptome coexpression analysis also revealed another novel flavonoid glycosyltransferase, that is, a flavonol 3-*O*-arabinosyltransferase.[187] The expression pattern of the *UGT78D3* gene was correlated with that of *MYB111*, which encodes a flavonoid-related transcription factor. The identity of UGT78D3 as an arabinosyltransferase was established by the analyses of *ugt78d3* knockout mutant and the recombinant protein expressed in *E. coli*.

1.24.3.4.2(iv)(c) Galactosyltransferases Black gram, *Vigna mungo*, accumulates many kinds of flavonol and anthocyanin glycosides. The young leaves (first simple leaves) of *V. mungo* accumulates kaempferol 3-*O*-robinobioside-7-*O*-rhamnoside (robinin) (**101**), which possesses galactose at the 3-*O* position, as the predominant flavonoid, and they exhibited high UDP-galactose: flavonoid 3-*O*-galactosyltransferase (UF3GaT) activity. The cDNA encoding UF3GaT was for the first time isolated by hybridization screening of a cDNA library constructed from this material using a cDNA for UDP-glucose: flavonoid 3-*O*-glucosyltransferase (UF3GT) of *Antirrhinum majus* as a probe. The deduced amino acid sequence showed 42% identity with that of *A. majus* (UF3GT). The recombinant enzyme exhibited high UF3GaT but low UF3GT activities, and the kinetic parameters, optimal pH, and substrate specificity were similar to those of the native UF3GaT partially purified from *V. mungo*.[188]

Flavonols have been known to be required for pollen germination in maize and petunia. As a step to understand the mechanism of flavonol-induced pollen germination, a UDP-galactose: flavonol 3-*O*-galactosyltransferase of petunia, which catalyzes the production of kaempferol 3-*O*-galactoside (**102**) and quercetin 3-*O*-galactoside (**103**), was purified from petunia pollen. It is exclusively expressed in the male gametophyte and postulated to control the formation of a pollen-specific class of glycosylated flavonol. The corresponding cDNA was cloned based on the amino acid sequences of three peptide fragments of the purified enzyme. The recombinant enzyme uses only UDP-galactose and flavonols as the substrates, and it also catalyzes the reverse reaction, the generation of flavonols from UDP and flavonol 3-*O*-galactosides (**102**, **103**).[189] The pollen-specific galactosyltransferase of petunia is distinguished by the highly restricted tissue-specific expression and substrate usage from other flavonoid glycosyltransferases.[190] The promoter region of the encoding gene includes large blocks of chloroplast and mitochondrial DNA, which was also observed in a orthologous gene of *N. tabacum*. In spite of *cis* elements in the promoter region, which may be activated by Myb and bHLH transcription factors, the experimental evidence did not support its regulation by the Myb factor known to regulate anthocyanin accumulation in petunia anthers. These unique features of the gene structure and enzyme activity were confirmed to be conserved throughout the Solanaceae.[190]

1.24.3.4.2(iv)(d) Sugar-sugar glycosyltransferases The glycosyltransferases that catalyze glycosylation of existing sugar moieties of flavonoid glycosides are important for structural diversity of flavonoids because they yield a large number of combinations of sugars and their binding positions. Several sugar-sugar glycosyltransferases have been identified in this decade.[191–194]

1.24.3.4.2(v) Anthocyanin/flavonoid acyltransferases Acylation of anthocyanins is known to have significant effects on the stability and color of anthocyanins. Anthocyanin acyltransferase (AAT) catalyzes the transfer of aromatic or aliphatic carboxylic acid moieties to the glycosyl groups of anthocyanins or flavonoids (**Scheme 16**). Engineering the genes encoding AATs is expected to be useful to modify flower color.

The AATs that have been well studied so far use CoA esters as acyl donors. Although the acyl-acceptor specificity of this kind of AATs differs greatly with the enzyme, comparing amino acid sequences of AATs revealed that they belong to the benzylalcohol acetyltransferase, anthocyanin-*O*-hydroxy-cinnamoyltransferase, anthranilate-*N*-hydroxy-cinnamoyl/benzoyltransferase, deacetylvindoline acetyltransferase (BAHD) enzyme superfamily, or versatile acyltransferase (VAT) family, which share highly conserved sequences such as HXXXD and DFGWG, as motives 1 and 3, respectively. Moreover, the anthocyanin-specific members of the BAHD family were shown to share an additional well-conserved YFGNC sequence as the motif 2.[195] On the other hand, some acyl transferases that share these motives were shown to use flavonoids other than anthocyanins as the acyl acceptors.

The other type of AAT uses acylglucoses as acyl donors. In a crude extract of carnation (*Dianthus caryophyllus* L.), the activity of 1-*O*-malylglucose: pelargonidin 3-*O*-glucose-6″-*O*-malyltransferase was detected.[196]

1.24.3.4.2(v)(a) Hydroxycinnamoyl-CoA: anthocyanidin 5-O-glucoside-6‴-O-hydroxycinnamoyltransferase The first identified cDNA clone encoding an acyltransferase involved in anthocyanin biosynthesis was that encoding hydroxycinnamoyl-CoA: anthocyanidin 5-*O*-glucoside-6‴-*O*-hydroxycinnamoyltransferase (anthocyanin 5-aromatic acyltransferase), which catalyses the acylation of delphinidin 3,5-diglucoside (**90**) to form delphinidin 3-glucoside-5-caffeoylglucoside (**104**). This enzyme was purified to homogeneity from *Gentiana triflora*. The corresponding cDNA was isolated based on the internal amino acid sequence of the purified enzyme and was heterologously expressed in *E. coli* and yeast for the biochemical identification. The enzymatic characteristics of the recombinant protein were consistent with those of the native enzyme. The deduced amino acid sequence contains consensus motifs that are conserved among the putative acyl-CoA-mediated AAT.[197]

1.24.3.4.2(v)(b) Hydroxycinnamoyl-CoA: anthocyanidin 3-O-glucoside-6″-O-hydroxycinnamoyltransferase A cDNA encoding hydroxycinnamoyl-CoA: anthocyanin 3-*O*-glucoside-6″-*O*-hydroxycinnamoyltransferase was isolated from *P. frutescens*. The biochemical characterization was carried out using the recombinant enzyme expressed in *E. coli* and yeast. It recognizes the putative native substrates, cyanidin 3-*O*-glucoside (**69**) and cyanidin 3,5-di-*O*-glucoside (**89**) to form cyanidin 3-caffeoylglucoside (**105**) and cyanidin

3-caffeoylglucoside-5-glucoside (**106**), as well as other anthocyanins. The inhibitory experiments suggest that histidine and cysteine residues are essential for their catalytic activity.[198]

***1.24.3.4.2(v)(c) Malonyl-CoA: anthocyanidin 5-O-glucoside malonyltransferases from* Salvia splendens** The first purified and characterized anthocyanin malonyltransferase was malonyl-CoA: anthocyanin 5-*O*-glucoside-6‴-*O*-malonyltransferase (Ss5MaT1) purified from scarlet sage, *Salvia splendens*, which produces salvianin (**107**). A cDNA corresponding to the enzyme was isolated on the basis of the partial amino acid sequences of the purified enzyme. The native and recombinant enzymes catalyze the regiospecific transfer of the malonyl group from malonyl-CoA to the 6‴-hydroxy group of the 5-glucosyl moiety of anthocyanins. They effectively acted on shisonine (cyanidin 3-*O*-(6″-*O*-*p*-coumaroyl)-glucoside-5-*O*-glucoside (**108**)) and bisdemalonylsalvianin (pelargonidin 3-(6″-*O*-caffeylglucoside)-5-glucoside (**109**)); however, anthocyanins without the aromatic acyl group at the 3-glucoside moiety did not serve as substrates. The importance of the aromatic acyl group at the 3-glucoside for substrate recognition was suggested by the observation that the activity was competitively inhibited by *p*-coumaric acid, which mimics an aromatic acyl group.[199]

To elucidate the reaction mechanism and key amino acid residues of the BAHD family plant acyltransferase, steady-state kinetic analyses and site-directed mutagenesis of Ss5MaT1 were carried out. As a result, the most likely was a ternary complex mechanism in which both substrates and the enzyme form a complex before catalysis occurs, as in the case of chloramphenicol *O*-acetyltransferase and histone acetyltransferase. The alanine scanning mutagenesis of the polar or ionized amino acid residues invariant among the plant acyltransferases revealed two essential residues, His167 and Asp390.[200]

Another anthocyanin 5-*O*-glucoside malonyltransferase of *S. splendens*, malonyl-CoA: anthocyanin 5-*O*-glucoside 4‴-*O*-malonyltransferase (Ss5MaT2), which produces salvianin (**107**) from monodemalonylsalvianin (pelargonidin 3-(6″-*O*-caffeylglucoside)-5-(6‴-*O*-malonylglucoside) (**110**)) was purified, followed by the molecular cloning of the corresponding cDNA. The deduced amino acid sequence of Ss5MaT2 contained motives 1 and 3, but lacked motif 2, which is conserved in AATs. Moreover, Ss5MaT2 showed only 22–24% identity to other AATs including Ss5MaT1 (24%); it was most identical (31%) to acetyl-CoA: benzylalcohol acetyltransferase (BEAT) of *C. breweri*. In the phylogenetic tree of the BAHD family, Ss5MaT2 and BEAT, together with other related proteins, are located in a distinct clade from that of AATs including Ss5MaT1. However, Ss5MaT2 showed no BEAT or BEAT-related activity.[201]

1.24.3.4.2(v)(d) Malonyl-CoA: anthocyanin 3-O-glucoside-6″-O-malonyltransferase and their related enzymes Anthocyanins in flowers of Asteraceae plants are known to have malonyl group(s) at the 3-glucosyl moiety. A cDNA encoding malonyl-CoA: anthocyanin 3-*O*-glucoside-6″-*O*-malonyltransferase was cloned from *Dahlia variabilis* based on the amino acid sequences specifically conserved among AATs of the BAHD family. The recombinant enzyme catalyzed the regiospecific transfer of the malonyl group from malonyl-CoA to pelargonidin 3-*O*-glucoside (**86**) to produce pelargonidin 3-*O*-6″-*O*-malonylglucoside (**111**). The enzymatic characteristics of the recombinant protein were consistent with those of the native enzyme activity in the extract of dahlia flowers. Cyanidin 3-*O*-6″-*O*-malonylglucoside (**112**) was shown to be more stable both in the pHs corresponding to those observed in the petal extract and against β-glucosidase digestion than cyanidin 3-*O*-glucoside (**69**), suggesting that malonylation should serve as a strategy for pigment stabilization in the flowers.[202]

Homology-based cDNA cloning using the conserved YFGNC sequence (motif 2) specific to AATs yielded cDNAs encoding flavonol acyltransferase from *Verbena* × *hybrida* and *Lamium purpureum*. The recombinant enzymes catalyzed the regiospecific transfer of the malonyl group from malonyl-CoA to the 6″-hydroxy group of quercetin 3-*O*-glucoside (**93**) with high specificities for both the acyl donor and acceptor. Thus, they were designated malonyl-CoA: flavonol 3-*O*-glucoside-6″-*O*-malonyltransferase. It was shown that members of the BAHD (VAT) family sharing the motif 2 are not restricted to AATs but should include acyltransferases on flavonoids other than anthocyanins.[203]

***1.24.3.4.2(v)(e) Molecular genetic and functional genomic approaches to anthocyanin acyltransferases from* A. thaliana** The model plant *A. thaliana* produces cyanidin derivatives modified by glycosylation and by addition of malonyl, *p*-coumaroyl, and sinapoyl, and thus the plant was expected to be a useful source for the identification of several AATs using molecular genetic and functional genomic approaches, particularly in the

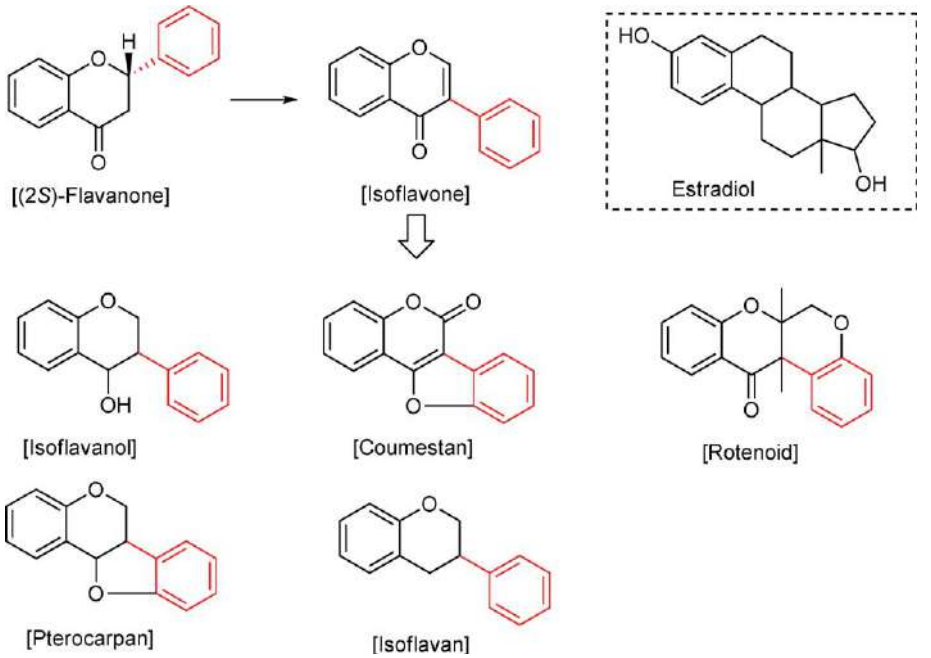
use of transcriptomic and metabolomic data on the *pap1-D* mutant. These approaches were used independently by two research groups.[204,205]

D'Auria *et al.*[204] narrowed down the candidates by phylogenetic analysis of the BAHD family of *A. thaliana* and screening for the AAT-specific motif 2 sequence (YFGNC). Among them, *At3g29590* and *At1g03940* were reported to be upregulated in the constitutive anthocyanin-accumulating *pap1-D* background.[185] The T-DNA insertional knockout mutant of *At3g29590* thus selected was investigated for the anthocyanin composition and was shown to be defective in producing malonylated anthocyanins. Biochemical analyses of the recombinant gene product of *At3g29590* showed it as malonyl-CoA: anthocyanin 5-*O*-glucoside-6″-*O*-malonyltransferase (At5MaT).[204]

On the other hand, Luo *et al.*[205] selected five candidate BAHD genes including *At3g29590* and *At1g03940* whose transcript levels were upregulated in the plants grown on high concentration of sucrose or low concentration of phosphate, which were known to strongly induce anthocyanin production. They functionally characterized the *At3g29590* gene by analyses of the knockout mutants and biochemical characterization of the recombinant enzyme, yielding the results showing that *At3g29590* encodes At5MaT. The functions of *At1g03940* and *At1g03195*, the third candidate, whose nucleotide sequences are 97% identical, were also analyzed using recombinant proteins expressed in *E. coli*. They were active with three hydroxycinnamoyl-CoA donors, *p*-coumaroyl-CoA, feruloyl-CoA, and caffeoyl-CoA. Anthocyanin 3-*O*-glucosides (**69**, **86**, **87**) were more preferable acyl acceptors than anthocyanin 3,5-*O*-diglucosides (**88**, **89**, **90**). No activity was detected when using cyanidin 3-*O*-rutinoside (**113**) as the acyl acceptor. It is thus likely that these enzymes specifically transfer acyl groups to the 6″-position of the 3-*O*-glucose, to which rhamnose is linked in 3-*O*-rutinosides. The enzymes also utilized flavonol glucosides, such as quercetin 3-*O*-glucoside (**93**), kaempferol 3-*O*-glucoside (**100**), and kaempferol 7-*O*-glucoside (**114**), as acyl acceptors, but they are not likely to be the native substrates judging from the kinetic parameters. The single knockout of *At1g03940* or *At1g03195* had little effect on the anthocyanin components, possibly because of functional redundancy of the two genes. Instead, transgenic tobacco plants overexpressing each of the genes were produced, and the production of unusual anthocyanins, which was likely to be cyanidin coumaroylglucoside, was detected.[205]

1.24.3.5 Isoflavonoids

Isoflavonoids constitute a diverse flavonoid subclass based on the 3-arylchromone parent skeleton distinct from most of the other flavonoids based on 2-arylchromone structure (**Scheme 17**). The major distribution of

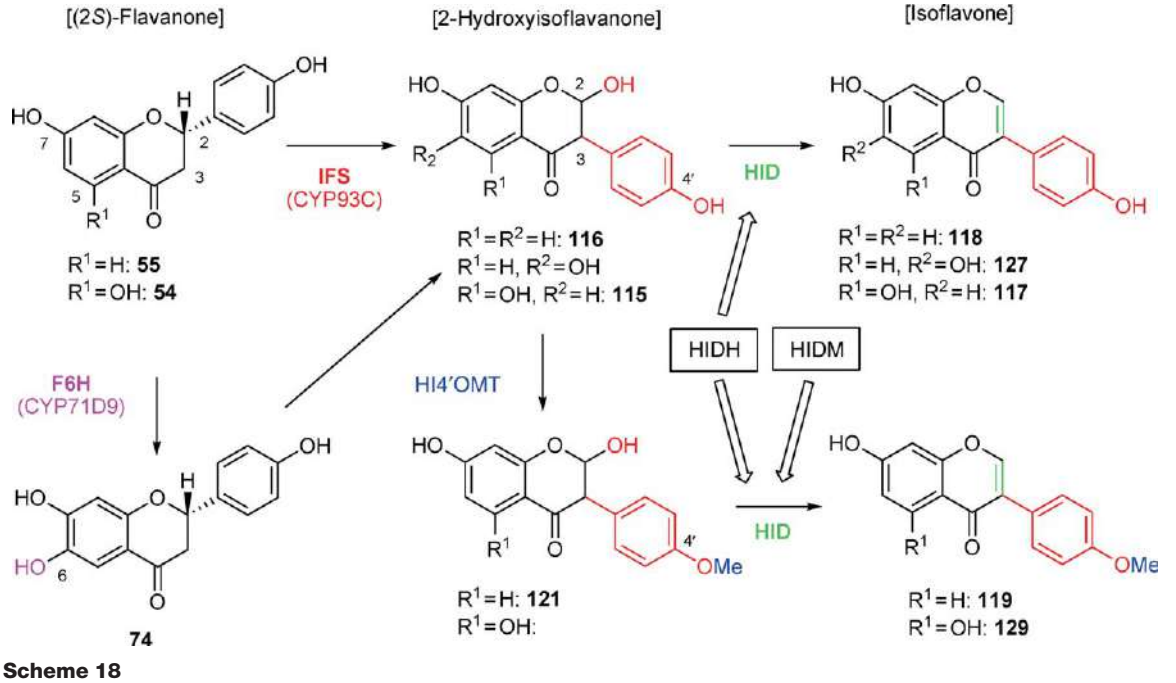


Scheme 17

isoflavonoids is in the Papilionoideae of Fabaceae, and significant ecophysiological functions of isoflavonoids in these plants as defensive agents (as phytoalexin, insecticidal chemicals, and allelochemicals) and symbiotic signals to microorganisms are noted. Estrogenic property of leguminous isoflavonoids can be regarded as another protective function against herbivores, while it can be utilized as health-promoting nutraceuticals in humans. The diverse structures of isoflavonoids and their useful physiological properties have been documented in a review by Dixon,[206] with the detailed description of their biosynthesis, although little molecular biological knowledge for the construction of this unique class of the flavonoids was available at the time. Undoubtedly, the characterization of the cDNA of the P-450 enzyme that constructs the isoflavonoid skeleton in leguminous plants in 1999–2000 was a breakthrough (see below), and, since then, remarkable advances both in the understanding of the biosynthetic mechanism of isoflavonoids and in the metabolic engineering incorporating isoflavonoid biosynthetic enzymes have been witnessed. The major plants for the biochemical and molecular biological studies on isoflavonoid biosynthesis have been agricultural legumes, soybean, pea and alfalfa, and medicinal legumes such as licorice and *Pueraria*. More recently, the established two model legumes, *M. truncatula* (barrel medic) and *L. japonicus*, have been tremendously useful as resources for the biosynthetic studies as well as for other scientific purposes. Review articles concerning isoflavonoid structures, biosynthesis, and biotechnology have recently been published.[207–209] Nonleguminous isoflavonoids are also interesting targets of research from chemotaxonomical and plant physiological viewpoints, and some reviews cover this topic.[210,211]

1.24.3.5.1 Biosynthesis of isoflavones in leguminous plants

Isoflavones are the early products of the isoflavonoid branching pathway and believed to be the common precursor of leguminous phytoalexins with pterocarpan and related isoflavonoid skeletons. The first intermediate possessing the rearranged isoflavonoid structure is produced from a general flavonoid intermediate, (2*S*)-flavanone, through a unique 1,2-aryl migration by a P-450. This P-450 has traditionally been designated isoflavone synthase, but detailed biochemical analysis with soybean[212] and *Pueraria lobata* cells[213,214] and more recent molecular biological studies discussed below have shown that the direct product of the P-450 action is 2-hydroxyisoflavanone, that is, 2,5,7,4′-tetrahydroxyisoflavanone (**115**) or 2,7,4′-trihydroxyisoflavanone (**116**), depending on the substrate, naringenin (**54**) or liquiritigenin (**55**) (**Scheme 18**). We use the abbreviation IFS for 2-hydroxyisoflavanone synthase (or isoflavonoid synthase) in this chapter, but in some papers IFS is still used



Scheme 18

for P-450 isoflavone synthase that actually represents 2-hydroxyisoflavanone synthase. The IFS products are then dehydrated to produce a double bond between C-2 and C-3, yielding genistein (**117**) or daidzein (**118**). Methyl transfer to 4′-hydroxy of **116** before dehydration leads to the formation of a 4′-methoxyisoflavone, formononetin (**119**). Post IFS processes to isoflavones, that is, the reactions catalyzed by 2,7,4′-trihydroxyisoflavanone 4′-*O*-methyltransferase (HI4′OMT) and 2-hydroxyisoflavanone dehydratase (HID), were first characterized with heterologously expressed cDNAs of elicited leguminous cells in *E. coli* through the examination of the catalytic activity in the repeatedly fractionated clones.

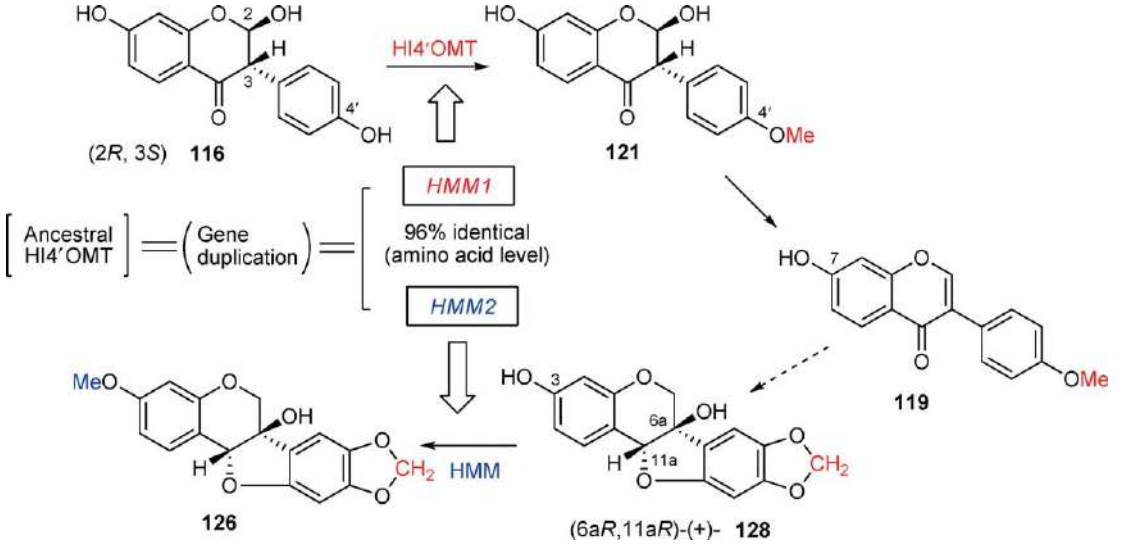
1.24.3.5.1(i) 2-Hydroxyisoflavanone synthase 2-Hydroxyisoflavanone synthase (IFS) cDNAs were initially characterized from licorice (*G. echinata*) and soybean. CYP93B1 of licorice has been shown to be F2H before (see Section 1.24.3.4.1), and another CYP93 family protein (assigned as CYP93C2 later) of licorice induced by elicitation concomitant with the phytoalexin (medicarpin (**120**)) accumulation was a candidate of IFS.[215] The recombinant yeast microsome expressing CYP93C2 converted (2*S*)-liquiritigenin (**55**) into a labile product that is readily transformed to daidzein (**118**) by acid treatment. The direct reaction product was identified as 2,7,4′-trihydroxyisoflavanone (**116**) by NMR measurement,[216] and a by-product, 3,7,4′-trihydroxyflavanone, was also noted conforming with the biochemical study with *P. lobata*.[213] From naringenin (**54**), 2,5,7,4′-tetrahydroxyisoflavanone (**115**) was produced. Independently, from soybean, CYP93C1 has been isolated earlier.[217] A functional genomics approach, that is, search for EST databases of cDNA from pathogen-infected hypocotyls and developing seeds, suggested that this P-450 was soybean IFS, and it was clearly confirmed by biochemical characterization.[218,219] Thereafter, CYP93Cs were cloned from several leguminous plants, *L. japonicus*,[220] *Trifolium pratense*,[221] and *M. truncatula*,[222] and their catalytic function as IFS has been biochemically verified.

CYP93B (FNS II and F2H; see Section 1.24.3.4.1(i)) and CYP93C (IFS) P-450 proteins recognize (2*S*)-flavanone molecules and abstract hydrogen from C-2 (CYP93B) or C-3 (CYP93C). Alignment of amino acid sequences of licorice CYP93B1 (F2H) and CYP93C2 (IFS) and the putative homology modeled active site structures of CYP93C2 resulted in the identification of several candidate residues important for the characteristic 1,2-aryl migration during the IFS catalysis.[10] Single and multiple site-directed mutagenesis of CYP93C2 regarding these residues and *in vitro* assay of the activities showed a dramatically changed reaction specificity accompanied with the alteration of protein stability in the yeast heterologous expression system.[10,223] The following functional residues were then identified: Ser305 at the center of I-helix near the C ring of the substrate that provides a wide space for aryl migration, Leu371 apparently controlling the substrate accommodation in favor of hydrogen abstraction from C-3 and Lys 375 near 7-OH essential for the migration. While the wild-type CYP93C2 yields the major 2-hydroxyisoflavanone and a minor 3-hydroxyflavanone at the ratio of ca. 9: 1, all the mutants showed the reduced ratios or loss of 2-hydroxyisoflavanone among the products, and the triple mutant displayed a complete flavone synthase activity. Although the specific activity of the triple mutant was very low (0.02 mol min^{-1} P-450 mol^{-1} vs. 12 mol min^{-1} P-450 mol^{-1} for the wild type), a much increased expression level (453 nmol P-450 per gram microsome vs. 117 nmol P-450 per gram microsome for the wild type) and the prolonged stability of the triple mutant P-450 protein in the host yeast cells were observed. From these results, it was proposed that the IFS gene has evolved from the common ancestor of the FNS II gene by the substitutions of amino acid residues in the Fabaceae to provide the specific physiologically important products (isoflavonoids), albeit sacrificing the protein stability.

1.24.3.5.1(ii) 2-Hydroxyisoflavanone dehydratase 2-Hydroxyisoflavanones spontaneously lose water to produce isoflavones *in vitro*, especially under acidic conditions. However, dehydration yielding isoflavones by a soluble enzyme *in vivo* has been predicted, and actually 2-hydroxyisoflavanone dehydratase (HID) of *P. lobata*, a leguminous plant yielding C-glycosylisoflavones, has been partially purified (**Scheme 18**).[224] HID activity producing formononetin (**119**) from 2,7-dihydroxy-4′-methoxyisoflavanone (**121**) was detected in the extract of *E. coli* expressing cDNAs of elicitor-treated licorice cells, and repeated fractionation and enzyme assay of the positive *E. coli* population resulted in the identification of the cDNA encoding licorice HID (HIDM).[3] Through searching the homologous sequences, a cDNA of soybean HID (HIDH) was also cloned. Recombinant HIDM protein was specific to the 4′-methoxylated substrate, whereas HIDH was highly active to 2,7,4′-trihydroxyisoflavanone (**116**) (yielding daidzein (**118**)) but possessing additional lower activity to the 4′- methoxylated substrate. The substrate specificity of HIDs agreed with the profiles of isoflavonoid

phytoalexins produced by each plant species – medicarpin (**120**) of licorice and glyceollins (**122–124**) of soybean. The kinetic study with purified HIDH showed that spontaneous (nonenzymatic) dehydration of **116** in neutral solution was much slower (1×10^5-fold) than HIDH-mediated dehydration *in vitro*, supporting the physiological role of HID in plant cells. HID proteins were found to be the members of a large and ubiquitous carboxylesterase family to which many plant proteins involved in defense responses but with unknown catalytic activity belong.[225,226] Site-directed mutagenesis with soybean HID protein suggested that amino acid residues consisting of the oxyanion hole of Gly and the catalytic triad of Thr, His, and Asp are essential for the catalytic activity of HID as well as the residual carboxylesterase activity. Plant carboxylesterases genes constitute a single clade distinct from those of microbe and mammals, and HIDs should be recruited from enzymes of the primary metabolism during evolution. An interesting finding was that a soluble GA receptor protein is also a member of the carboxylesterase family, although His of the catalytic triard is substituted to Val and thus no esterase activity can be assumed.[227]

1.24.3.5.1(iii) 2-Hydroxyisoflavanone 4′-O-methyltransferase Formononetin (**119**) is the precursor of isoflavonoid phytoalexins possessing substituents derived from 4′-OMe, for example, medicarpin (**120**), vestitol (**125**), and pisatin (**126**). **119** could be synthesized from daidzein (**118**) by 4′-*O*-methylation, but this activity has not been demonstrated *in vitro* using protein extracts of plants producing formononetin derivatives, while daidzein 7-*O*-methyltransferase (D7OMT) yielding isoformononetin (4′-hydroxy-7-methoxyisoflavone) has been characterized and its cDNA cloned from alfalfa.[228,229] A possible involvement of D7OMT in **119** biosynthesis has been suggested from indirect evidence: for example, overexpression of D7OMT in alfalfa led to greater accumulation of **119** and medicarpin (**120**).[230] A new scheme of **119** biosynthesis involving 2,7,4′-trihydroxyisoflavanone (**116**) as the methyl acceptor and subsequent dehydration was then proposed by the *in vitro* assay using the protein extracts of elicited licorice cells.[231] While subcellular localization studies on transgenic D7OMT in alfalfa indicated its association with IFS on endoplasmic reticulum and probable change in its reaction specificity was suggested,[68,232] a cDNA encoding 2,7,4′-trihydroxyisoflavanone 4′-*O*-methyltransferase (HI4′OMT) that is distinct from D7OMT was isolated by the screening of functionally expressed licorice cDNAs (**Scheme 18**).[233] Recombinant HI4′OMT did not display D7OMT activity, and licorice D7OMT did not show HI4′OMT activity. HI4′OMT showed the highest amino acid identity (83%) with (+)-6a-hydroxymaackiain 3-*O*-methyltransferase (HMM) of pea, the final enzyme in the biosynthesis of the phytoalexin (+)-pisatin (**126**) (**Scheme 19**).[234] Together with the finding of HID specifically producing 4′-OMe isoflavone from licorice (see above) and characterization of HI4′OMT from other leguminous plants like *M. truncatula* [235,236] and *L. japonicus*,[233] the biosynthetic scheme of **119** through the action of HI4′OMT is now generally accepted.



Scheme 19

1.24.3.5.1(iii)(a) Isoflavone and isoflavanone OMTs of Medicago truncatula *and pea* A comprehensive analysis of isoflavonoid OMTs related to D7OMT and HI4′OMT in a leguminous model plant *M. truncatula* has been performed.[236] Phylogenetic analysis showed that eight OMT cDNAs are grouped into two distinct clades: one clade includes MtIOMTs 1–4 and alfalfa and licorice D7OMTs and another includes MtIOMTs 5–8, HI4′OMTs of licorice and *L. japonicus* and pea HMMs. Purified recombinant MtIOMTs displayed promiscuous activities to several isoflavonoid compounds, but, in general, those in the D7OMT clade prefer isoflavone substrates with a relatively flat conformation, whereas those in the HI4′OMT clade prefer substrates with chiral centers (isoflavanones and isoflavans) with a bent molecular shape. The former enzymes mainly transferred methyl groups to position 7 of isoflavones like glycitein (7,4′-dihydroxy-6-methoxyisoflavone) and daidzein (**118**), although low activity of MtIOMT3 to methylate the 4′-position of genistein (**117**) and 7,3′,4′-trihydroxyisoflavone in addition to 7-position of 6,7,4′-trihydroxyisoflavone (**127**) was observed. Genes of MtIOMTs 1–3 were located within a 20-kb region on a single BAC clone, suggesting that these OMT genes have arisen from ancient gene duplication event. MtIOMT5 was the typical HI4′OMT with a high specific activity, and this *M. truncatula* HI4′OMT also displayed a high specific activity of 3-*O*-methylation of (+)-6a-hydroxymaackiain (**128**) to produce (+)-pisatin (**126**), the pea phytoalexin not produced by *M. truncatula*.[235] Protein X-ray crystal structure of *M. truncatula* HI4′OMT substrate complexes showed that (2*S*, 3*R*)-stereoisomer of 2,7,4′-trihydroxyisoflavanone (in contrast to the predicted (2*R*, 3*S*)-stereoisomer (**502**) of the IFS reaction product[10]) was accommodated in the active site and it revealed identically bound conformation of (6a*R*,11a*R*)-(+)-6a-hydroxymaackiain (**128**) to be methylated to produce (+)-pisatin (**126**), whose overall conformation is enantiomeric to *Medicago* pterocarpans. From these observations, the evolution mechanism of pea HMM as a plant defense machinery taking advantage of the similar conformation of the substrates of the OMT was proposed. Further interesting finding came from the information that pea contains two copies of HMM, HMM1 and HMM2, with 96% amino acid identity (**Scheme 19**).[234] When the reactions with recombinant HMMs and licorice HI4′OMT were compared, HI4′OMT and HMM1 showed high efficiencies for the methylation of 2,7,4′-trihydroxyisoflavanone (**116**), while HMM2 had high efficiency for the methylation of (+)-6a-hydroxymaackiain (**128**).[237] HMM1 was thus concluded to be the pea HI4′OMT and that HMM2 evolved by the duplication of a gene encoding a general enzyme involved in formononetin (**119**) biosynthesis and mutation of a very limited number of amino acid residues (see Section 1.24.3.5.2(iii)).

1.24.3.5.1(iv) Flavanone 6-hydroxylase of soybean involved in 6-methoxyisoflavone biosynthesis Soybean accumulates a 6-methoxyisoflavone, glycitein (7,4′-dihydroxy-6-methoxyisoflavone), in addition to daidzein (**118**) and genistein (**117**), as well as their malonylglucosyl conjugates. A P-450 (CYP71D9) obtained from elicited soybean cells was shown to encode flavonoid 6-hydroxylase (F6H; **Scheme 18**) that introduces the C-6 hydroxyl during glycitein biosynthesis.[161] F6H hydroxylates C-6 of flavanones (liquiritigenin (**55**), eriodictyol (**73**), and naringenin (**54**)), flavones (apigenin (**60**) and luteolin (**72**)), and (dihydro)flavonols, but not isoflavones, and the most preferred substrate was **55** yielding 6,7,4′-trihydroxyflavanone (**74**). Recombinant soybean IFS (CYP93C1v2) efficiently used **74** and produced 6,7,4′-trihydroxyisoflavone (**127**) after acid treatment. Clearly, 6-hydroxylation occurs before B ring migration in glycitein biosynthesis.

1.24.3.5.1(v) Modification of isoflavones

1.24.3.5.1(v)(a) Glycosyltransferase A cDNA encoding UDP-glucose: formononetin 7-*O*-glucosyltransferase (UGT73F1) of licorice cell cultures has been identified using probes from *Scutellaria baicalensis* flavonoid 7-*O*-glucosyltransferase.[238] UDP-glucose: isoflavone 7-*O*-glucosyltransferase (UGT88E3) was then purified from soybean seedlings, and its cDNA was cloned.[239] Another soybean cDNA (UGT73F2) also encoding isoflavone 7-*O*-glucosyltransferase was isolated from seeds.[240] Naturally, all these isoflavonoid-specific glycosyltransferases (GTs) are classified as plant secondary product glycosyltransferases (PSPGs) of group 1 GT family. Phylogenetic analysis showed that UGT73F2 is closely related to UGT73F1, but UGT73F1 and UGT73F2 are distantly related to UGT88E3, which was clustered with snapdragon chalcone 4′-*O*-glucosyltransferase (UGT88D3; see Section 1.24.3.1.2(iii)) and others to form UGT88-related PSGSs. However, substrate preference for UGT88E3 and UGT73F2 was similar; genistein (**117**) > daidzein (**118**) > formononetin (**119**): in contrast, UGT73F1 most preferred **119**. Interestingly, the purified protein of GmIF7GT (UGT88E3) was devoid of N-terminal 49-residue segment and lacked the His15 and Asp125 representing the catalytically important residues identified in other

GTs (see below).[239] The site-directed mutagenesis showed that these residues are not important for this enzyme, indicating a distinct strategy of catalysis to known GTs.

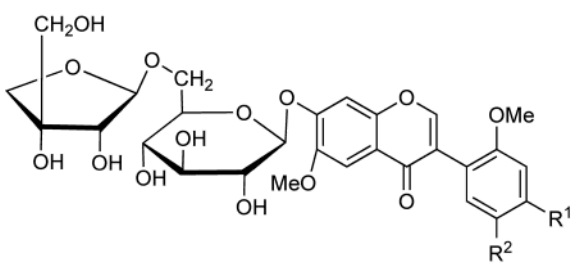
UGT71G1 of *M. truncatula* was identified as a GT that transfers glucose to a triterpene medicagenic acid, but it also transfers glucose to 7-hydroxyl of genistein (**117**) and biochanin A (**129**) and 3′-hydroxyl of a flavonol quercetin (**59**) with higher efficiencies than the triterpene, albeit its primary function in elicited cell cultures being attributed to triterpene biosynthesis from the transcript and metabolite profiling.[241] The crystal structure of UGT71G1 revealed key residues (His22 and Asp121) involved in the recognition of donor substrate, and point mutation confirmed the role of these residues in donor binding.[242] Furthermore, mutation of UGT71G1 revealed the residues that control regioselectivity for flavonoid glycosylation.[11] F148V and Y202A mutants changed regioselectivity of glucosylation of quercetin from 3′-*O* to 3-*O* position. Y202A also acquired the activity of 5-*O*-glycosylation of **117**. More recent EST database analysis of *M. truncatula* resulted in the identification of 150 different UGT genes, and *in vitro* assay revealed that 8 UGTs among 19 recombinant proteins showed glucosyltransferase activity to isoflavones, flavones, flavonols, and anthocyanidins.[243] In particular, three of the enzymes (UGT73C8, UGT88E1, and UGT88E2) showed an overall preference for isoflavonoids including pterocarpan phytoalexins.

1.24.3.5.1(v)(b) Malonyltransferase Two cDNAs showing 99% identity and encoding malonyl-CoA: isoflavone 7-*O*-glucoside-6″-*O*-malonyltransferase (IFMT) of soybean producing isoflavone 7-*O*-(6″-*O*-malonyl)-glucosides have been independently isolated by two groups.[244,240] IFMT belongs to BAHD acyltransferase family (see Section 1.24.3.4.2(v)).[245]

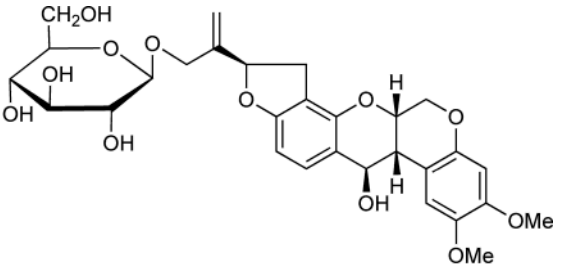
1.24.3.5.1(v)(c) Glycosidase Isoflavone conjugate-hydrolyzing β-glucosidase has been partially purified,[246] and a cDNA has been isolated from soybean.[247] The enzyme acts on isoflavone 7-*O*-glucosides and isoflavone 7-*O*-(6″-*O*-malonyl)-glucosides and release corresponding aglycons (genistein (**117**) or daidzein (**118**)) and glucose or malonylglucose. The enzyme is localized in the cell wall and intercellular space of roots. A comprehensive analysis of methyl jasmonate-treated *M. truncatula* cell cultures identified several β-glucosidases that act on diverse isoflavone glucosides and possibly function in defense response of the cells (see Section 1.24.4.1). [248] Isoflavonoid glycosidases were also purified and their cDNAs cloned from *Dalbergia* sp. (**Scheme 20**). Deduced amino acid sequences of both glycosidases showed 81% identity, but each enzyme showed unique activity toward isoflavonoid glycosides. A glycosidase of *Dalbergia nigrescens* prefers isoflavone diglycosides ((**130, 131**) dalpatein and dalnigrein 7-*O*-β-D-apiofuranosyl-(1 → 6)-β-D-glucoside) as the substrates and releases disaccharide (6-*O*-β-D-apiofuranosyl-D-glucose) and their aglycones.[249,250] Another one obtained from *Dalbergia cochinchinensis* hydrolyzes rotenoid glucoside (dalcochinin 8′-*O*-glucoside (**132**)) to release D-glucose and its aglycone more efficiently than isoflavone diglycosides.[250–252]

1.24.3.5.2 Isoflavonoid phytoalexins

1.24.3.5.2(i) Biosynthesis of pterocarpan and isoflavan skeletons Isoflavonoids possessing pterocarpan and isoflavan skeletons are the most frequently found phytoalexins of leguminous plants.[253] After the formation

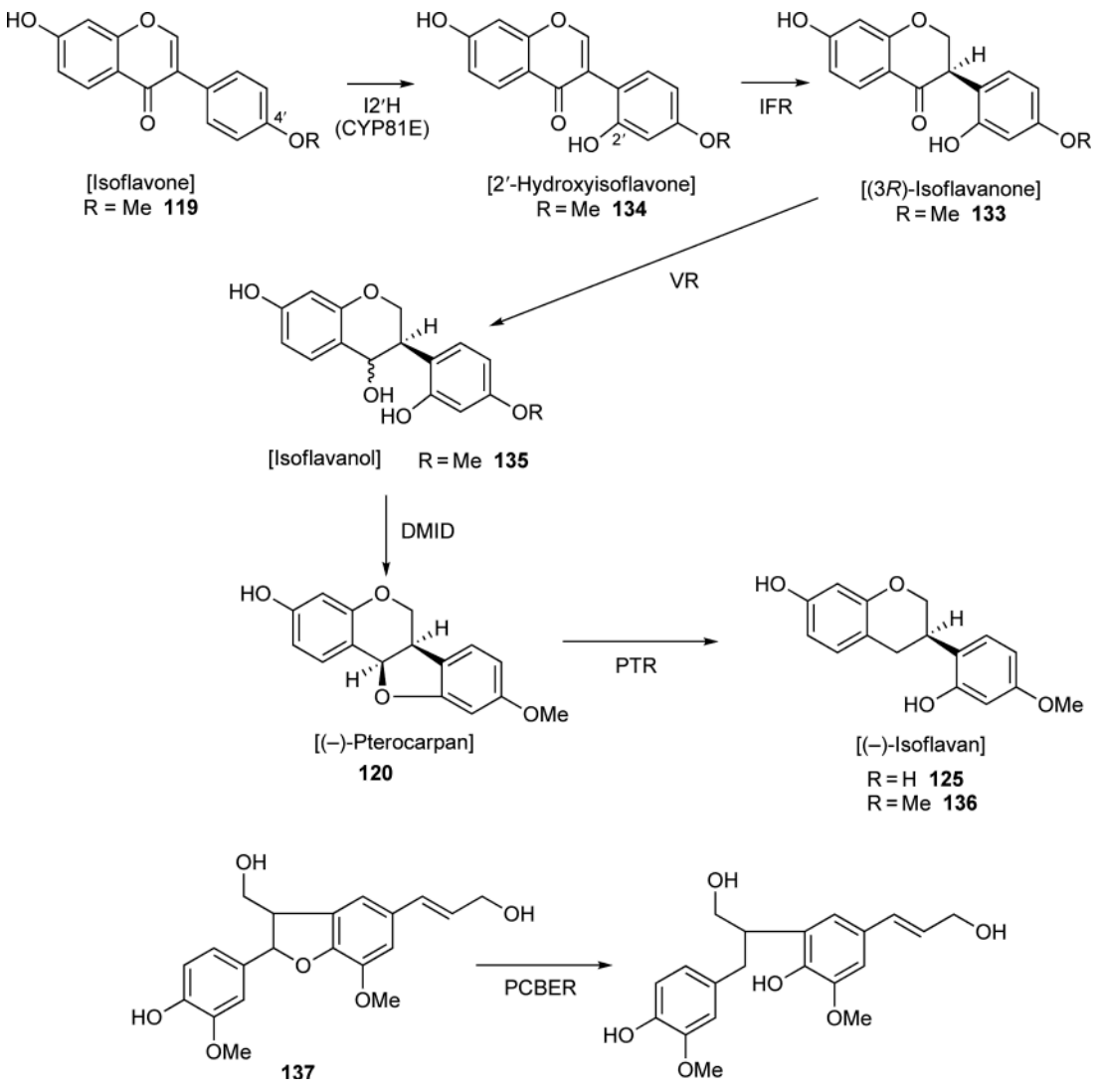


R^1-R^2 = O-CH_2-O: dalpatein glycoside **130**
R^1, R^2 = OMe: dalnigrein glycoside **131**



132

Scheme 20



Scheme 21

of isoflavone skeleton, pterocarpans are constructed by several reactions including oxidation, reduction, and dehydrative cyclization (**Scheme 21**).[254]

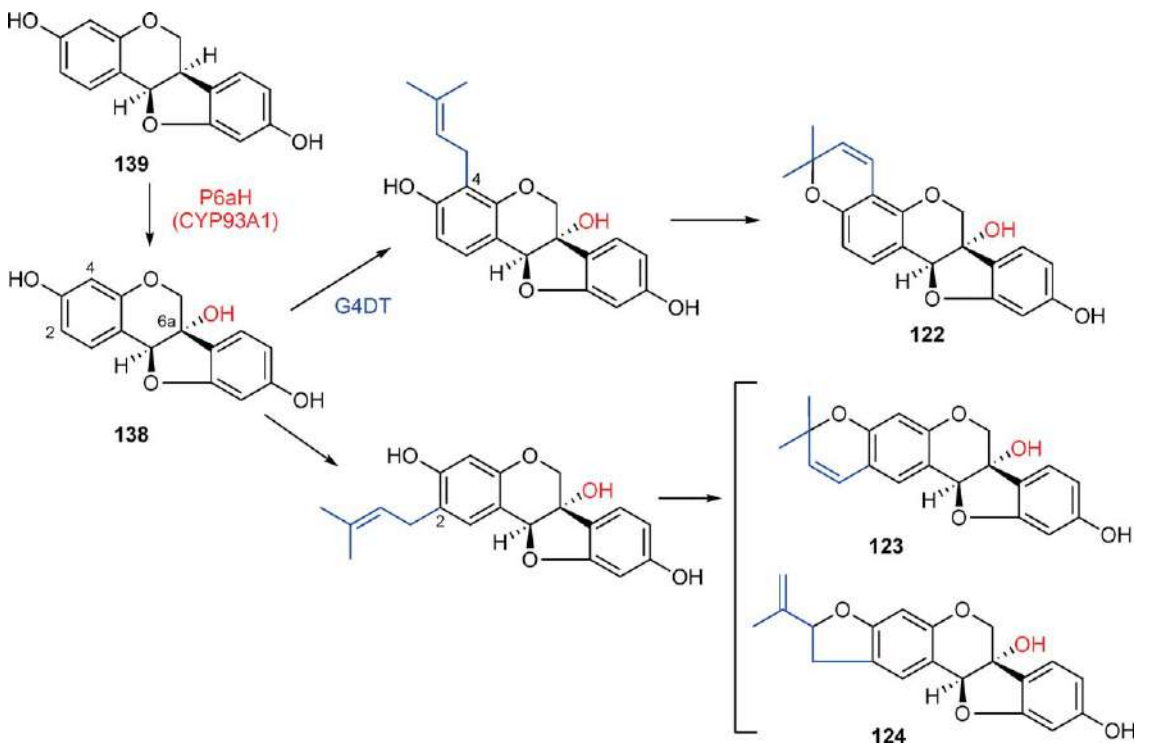
1.24.3.5.2(i)(a) Isoflavone 2′- and 3′-hydroxylases Hydroxylation at C-2′ of isoflavone is essential for the formation of pterocarpan and isoflavan skeletons. Hydroxylation at C-3′ of isoflavone is important for the formation of methylenedioxy functional group at B ring. These reactions have been attributed to P-450s using crude extract of elicited chickpea cells that accumulate maackiain as the phytoalexin.[255,256] One of the licorice P-450 (CYP81E1) induced after elicitor treatment of the cells was identified to be isoflavone 2′-hydroxylase (I2′H) using genistein (**117**), daidzein (**118**), and formononetin (**119**) as the substrates.[257] CYP81E isoforms were also identified in *L. japonicus*, chickpea, and *M. truncatula*.[250,258,259] Among the two CYP81E isoforms of *M. truncatula*, CYP81E7 encodes I2′H, while CYP81E9 encodes isoflavone 3′-hydroxylase (I3′H).[259] I2′H is expressed mainly in the roots and in elicited cultured cells and leaves, and I3′H in the overground part and is inducible by jasmonic acid. Differential tissue-specific and stress-dependent expression patterns suggested different functional roles of 2′- and 3′-hydroxyisoflavonoids in *M. truncatula*, the latter being suggested to be involved in insect resistance.

1.24.3.5.2(i)(b) Isoflavone reductase, isoflavanone 4-reductase, and 2′-hydroxyisoflavanol dehydratase 2′-Hydroxyisoflavones are subjected to a series of NADPH-dependent reductions, first by isoflavone reductase (IFR) to form (3*R*)-isoflavanones, and then by isoflavanone 4-reductase like vestitone reductase (VR) to yield 2′-hydroxyisoflavanol. Dehydrative formation of dihydrofuran ring from 2′-hydroxyisoflavanol by a dehydratase yields a pterocarpan skeleton. In alfalfa, (3*R*)-vestitone (**133**) produced from 2′-hydroxyformononetin (**134**) by IFR is converted to 7,2′-dihydroxy-4′-methoxyisoflavanol (DMI) (**135**) by VR, and DMI dehydratase (DMID) produces medicarpin (**120**). Both IFR and VR cDNAs have been cloned from alfalfa,[260,261] and crystal structures of IFR[262] and VR[263] proteins were revealed. IFR and VR belong to the SDR (or RED) family (see Section 1.24.2.2), and their three-dimensional structures were very similar, in spite of their low (∼10%) sequence identity. A catalytic triad (Ser129–Tyr164–Lys168) is located at the VR active site, and Tyr164 is assumed to be a general base for catalysis, while there is no catalytic triad in IFR, in which a single Lys144 residue was identified as the catalytic residue. For the final enzyme of pterocarpan skeleton synthesis, alfalfa DMID protein has been purified,[264,265] but no information about its gene is available.

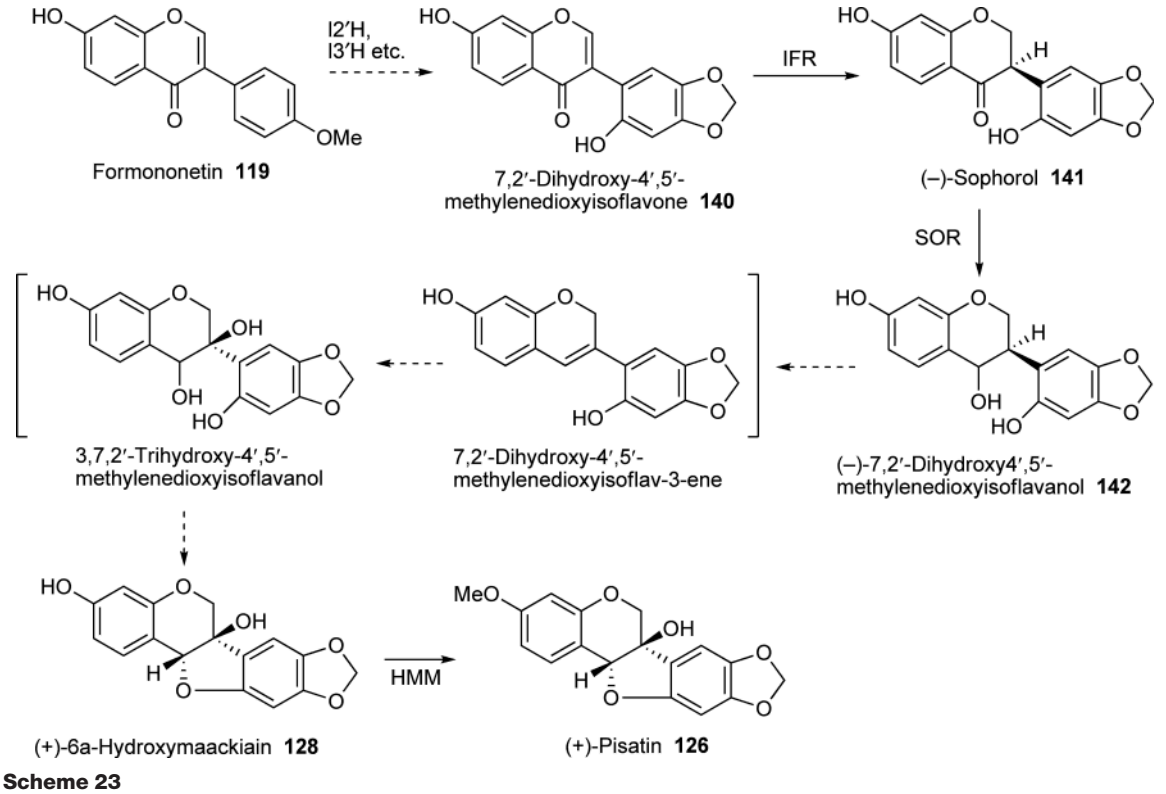
1.24.3.5.2(i)(c) Isoflavan biosynthesis by pterocarpan reductase (−)-Vestitol (**125**) and (−)-sativan (**136**) are typical isoflavan phytoalexins produced by leguminous plants including a model legume *L. japonicus*.[220,266] Earlier studies have suggested a close biosynthetic relationship between isoflavan and pterocarpan: interconversion of the pterocarpan medicarpin (**120**) and the isoflavan **125** has been demonstrated by feeding ^{14}C-labeled precursors to alfalfa seedlings.[267] PCBER, an SDR (or RED) family enzyme of lignan pathway, catalyzes the cleavage of the ether linkage of the dihydrofuran ring of (±)-dehydrodiconiferyl alcohol (**137**),[268] and this reaction mimics the feasible isoflavan skeleton formation from a pterocarpan. Therefore, EST databases of *L. japonicus* were mined for PCBER-like SDR sequences, and four clones (*PTR1–PTR4*) homologous to PCBER were isolated.[269] The recombinant proteins (PTR1–PTR4) catalyzed the formation of vestitol from medicarpin in the presence of NADPH. Among these, PTR1 and PTR2 displayed a high specific activity and mediated the enantiospecific reduction of (−)-medicarpin (**120**), whereas PTR3 and PTR4 showed a low activity without enantiospecificity. It is thus concluded that PTR1 and PTR2 are the enzymes (pterocarpan reductase) producing (−)-vestitol (**125**) in *L. japonicus*. Both PTR3 and PTR4 had a high identity (65–72%) to PCBER, which is known to lack enantiospecificity toward (±)-dehydrodiconiferyl alcohol (**137**), while PTR1 and PTR2 showed lower (about 60%) identity to PCBER. Probably, PTR3 and PTR4 have first been recruited from the ancestral PCBER, and then PTR1 and PTR2 more specialized for phytoalexin synthesis emerged during evolution.

1.24.3.5.2(ii) Glyceollin biosynthesis in soybean (−)-Glyceollins (**122–124**) are the soybean phytoalexins that are pterocarpans with cyclic ether decoration originated from a prenyl substitution (**Scheme 22**). During glyceollin biosynthesis, (−)-3,9,6a-trihydroxypterocarpan ((−)-glycinol (**138**)) is biosynthesized by a P-450-mediated hydroxylation at C-6a of 3,9-dihydroxypterocarpan (**139**), and then a dimethylallyl group is introduced to either C-4 or C-2 of **138**. The final formation of the ether ring from the prenyl side chain has also been attributed to a P-450 action. A P-450 (CYP93A1) isolated from elicited soybean cells was demonstrated to encode 3,9-dihydroxypterocarpan 6a-hydroxylase (P6aH).[270] This was the first identification of the P-450 catalysis involved in isoflavonoid metabolism. Recently, a cDNA selected from soybean EST database homologous to homogentisate phytyltransferase of tocopherol biosynthetic pathway was shown to encode dimethylallyl diphosphate: (−)-glycinol 4-dimethylallyltransferase (G4DT), yielding the direct precursor of (−)-glyceollin I (**122**).[279] The full-length cDNA has a plastid-targeting signal sequence, and plastidal localization of the protein was verified using GFP fusion protein. Furthermore, the prenyl part of **122** was shown to be methylerythritol phosphate pathway origin by a tracer experiment using [1-^{13}C]glucose.

1.24.3.5.2(iii) Pisatin biosynthesis in pea In contrast to major leguminous plants that produce (−)-stereoisomer of pterocarpans such as soybean producing (−)-glyceollins (**122–124**) and chickpea, alfalfa, and licorice producing (−)-medicarpin (**120**), pea produces (+)-pisatin (**126**), a (+)-stereoisomer of pterocarpan, as the phytoalexin. **126** is likely to be biosynthesized from an achiral 2′-hydroxyisoflavone, but the accurate mechanism of the introduction of (+)-chirality is still not known (**Scheme 23**).[272] Two reductases that produce direct intermediates of (−)-pterocarpan synthesis, that is, an IFR catalyzing the conversion of



Scheme 22



Scheme 23

7,2′-dihydroxy-4′,5′-methylenedioxyisoflavone (**140**) to (−)-sophorol (**141**) and sophorol reductase (SOR) mediating the conversion of **141** to (−)-7,2′-dihydroxy-4′,5′-methylenedioxyisoflavonol (**142**), have been characterized in pea and their cDNAs were isolated.[273,274] From RNAi studies, IFR and SOR were shown to be involved in **126** biosynthesis,[272,274] meaning the stereochemical inversion taking place between isoflavanol and (+)-6a-hydroxymaackiain(**128**). Tracer experiments have suggested the involvement of achiral isoflavene as the precursor of **126**.[272] The last step of pisatin (**126**) biosynthesis is the *O*-methylation of (+)-pterocarpan, and the reaction is mediated by HMM, which should have evolved from HI4′OMT, the enzyme of 4′-methoxylated isoflavone biosynthesis (see Section 1.24.3.5.1(iii)(a)).[237,275]

1.24.3.5.3 Transformation

A metabolic engineering using IFS genes aiming at the production of isoflavonoids in nonleguminous plants that are not known to produce isoflavonoids has attracted attention from the viewpoints of exploring plant physiological functions of isoflavonoids as well as of future practical applications.[276,277] It can be expected that the introduction of IFS in nonlegumes may convert them to isoflavonoid producers, because the substrate of IFS, (2*S*)-flavanone, is a common constituent of plants. Transgenic arabidopsis, tobacco, rice, petunia, lettuce, and tomato expressing soybean IFS produced genistein (**117**) and/or its glucoconjugate,[93,272,278–282] although the productivity of isoflavones was much lower than in seeds of leguminous plants. The metabolic flow from flavanone to 3-hydroxyflavanone by FHT (see Section 1.24.3.4.1(ii)(a)) should compete with the isoflavone production in transgenic plants: actually, the isoflavone production in *A. thaliana* F3H mutant transformed with IFS was increased 6- to 31-folds. The suppression of F3H by RNAi and the introduction of soybean IFS in tobacco also resulted in increased accumulation of isoflavones than single IFS transformants.[282] In another trial, overexpression of PAL and introduction of soybean IFS in tobacco and lettuce led to about 1.5-fold enhanced **117** accumulation than single IFS transformant.[282] Daidzein (**118**) was produced in maize cells coexpressing a transcription factor CRC together with PKR.[278] Increase of carbon flow to isoflavone pathway was achieved using artificial bifunctional IFS/CHI in frame fusion protein expressed in tobacco.[283]

Modifications of isoflavonoid contents in legumes have also been accomplished. The overexpression of IFS in soybean elevated isoflavone levels,[284] while suppression of IFS in soybean by RNAi resulted in a high degree of isoflavone decrease and caused the enhancement of susceptibility to the pathogen *Phytophthora sojae*[285] and the severely reduced nodulation.[286,287] Silencing of IFS or CHR in soybean by RNAi led to large decreases of 5-deoxyisoflavone and glyceollin with a breakdown of race-specific resistance to *P. sojae* and the suppression of cell death and reactive oxygen responses associated with race-specific resistance (see also Section 1.24.3.1.1).[61] Isoflavonoids preferably with 5-deoxy-type structure were thus shown to be critical for defense against pathogens and the establishment of nodulation in soybean. In transgenic alfalfa expressing *M. truncatula* IFS, glycosides of **117** and biochanin A (**129**) accumulated, and medicarpin (**120**) contents were elevated ~25-fold compared to control plants after the pathogen infection.[222] Modification of isoflavone contents in transgenic hairy root cultures by the introduction of soybean HID was also accomplished.[288] Soybean HID has broad substrate specificity toward 4′-hydroxylated and 4′-methoxylated 2-hydroxyisoflavanones, and the introduction of soybean HID in *L. japonicus* resulted in the accumulation of **118** and genistein **117**, showing that HID is a critical determinant of isoflavone productivity.

1.24.4 Functional Genomics and Evolution of Phenylpropanoid/Flavonoid Biosynthesis

1.24.4.1 Metabolomics and Transcriptomics

Limited examples of recent functional genomics approaches in relation to the phenylpropanoids/flavonoids area are described here. (More detailed coverage should be found in other volumes of this series (see Chapters 3.24, 9.15, and 9.16.) Transcriptome and metabolome analysis and combination of these -ome methodologies such as coexpression network analysis of the plants under specific conditions and developmental stages, for example, abiotic and biotic stresses, during symbiosis and overexpression of MYB transcription factors, are very useful to reveal gene functions regarding the specialized metabolism.[289] Thus, several flavonoid

glycosyltransferases and acyltransferases of *A. thaliana* have been identified (also see Sections 1.24.3.4.2(iv)(b) and 1.24.3.4.2(v)(e)), and the distinct roles of PAL isozymes have been elucidated.[185,205,290]

To have a comprehensive picture of secondary metabolism reprogramming after exogenous stimuli to plant cells, cultured cells of a model legume, *M. truncatula*, treated with microbial-derived elicitor (yeast elicitor (YE)) or wound signal methyl jasmonate (MJ) were subjected to metabolome and transcriptome analysis.[248,291] DNA microarray analysis displayed rapid and massive induction of many of early (iso)flavonoid biosynthetic genes (PAL, C4H, 4CL, CHS, PKR, and CHI) by YE but not by MJ, while both elicitors induced genes for enzymes that convert isoflavone formononetin (**119**) into pterocarpan medicarpin (**120**) (those annotated as HI4′OMT, HID, I2′H, IFR, VR). YE-treated cells produced known (formononetin (**119**) and biochanin A (**129**)) and novel (afrormosin, alfalone, and irisolidone) isoflavones from *M. truncatula* in addition to **120**, daidzin and genistein diglucoside, and most compounds were excreted to the culture medium. In contrast, MJ-treated cells also accumulated **120**, but no coordinated increase in isoflavonoid precursors and, instead, decline of vacuolar isoflavone glucoside were observed. During these processes, some β-glucosidases and ABC transporters were also induced, and a scheme of hydrolysis of preformed conjugated intermediates in vacuole to be used for phytoalexin by wound signal is postulated. In *L. japonicus*, another model legume, comprehensive analysis of flavonoid glycosides has been performed.[292] Also, expression of CHS7 and CHS8 for isoflavone conjugate accumulation in soybean has been shown to be critical through transcriptome analysis during seed development.[293]

1.24.4.2 Organization and Phylogenetic Relationship of Genes of Phenylpropanoid and (Iso)Flavonoid Pathway

Genomic organization and evolution of the respective enzymes involved in the biosynthesis of (iso)flavonoids are mentioned in the above sections. In recent years, genome-wide analyses of the structural gene families involved in the phenylpropanoid pathway have been carried out. Genomic organization, gene expression pattern, and molecular phylogeny have been discussed in the lignin and flavonoid biosynthesis of *Populus*[294,295] and in the legume-specific 5-deoxyisoflavonoid biosynthesis of *L. japonicus*.[296] Characteristically, many paralogous structural genes encoding biosynthetic enzymes tandemly clustered on separate segments of the *L. japonicus* chromosomes as shown for CHI and DFR genes.

Comparative genomics has revealed that P-450 genes (CYP71, CYP73, CYP75, CYP81, or CYP93 family genes) associated with the phenylpropanoid biosynthesis belong to a single cluster (CYP71 clan) in the phylogenetic tree of the plant P-450 gene family.[297,298] Also, acyltransferases[299] and *O*-methyltransferases[69] involved in the modification of flavonoids are clustered in the respective phylogenetic trees, whereas, the enzymes involved in flavonoid glycosylation are variable and are divided into several groups in the phylogenetic tree.[243,300,301] Development of bioinformatic tools such as the biochemical pathway databases, AraCyc[302] and MedicCyc,[303] must be valuable for the functional and evolutionary genomics.

1.24.4.3 Evolution of Phenylpropanoid/Flavonoid Biosynthesis

Phenylpropanoids have a significant role in the evolution of land plants. Lignin is the rigid polymer functional in mechanical support in air and in UV-B absorption, and flavonoids are functional in UV-B absorption and in interactions with other organisms. Lignin is confined to tracheophytes, although lignans are found in bryophytes. Flavonoids seem to be restricted to land plants (embryophytes), and are not found in charophycean algae, which is considered to be an ancestral group of land plants.[304] Meanwhile, it is argued that the initial function of flavonoids is the internal physiological regulators or chemical messengers, than as a filter against UV irradiation.[305] Elucidation of flavonoid biosynthesis pathway is imperative to understand the evolutionary role of flavonoids.[306] The development of genomic analysis for the moss *Physcomitrella patens* in addition to the model plant *Arabidopsis* and the model legumes has been remarkable, and it is very significant that the gene of naringenin chalcone (**39**) producing CHS was cloned and characterized from the moss.[307]

Gene duplication is considered to be a major mechanism for evolutionary innovations and functional divergence. Positive Darwinian selection after gene duplication is considered to cause the rapid evolution in gene families. To examine whether the genes underwent adaptive evolution, the nonsynonymous/synonymous

rate ratio ($\omega = d_N/d_S$) is used as an indicator of selective pressure. The ω ratio $<1, =1$, and >1 indicates negative (purifying) selection, neutral evolution, and positive (adaptive) selection, respectively.

Yang and coworkers investigated adaptive evolution of CHS gene families of *Dendranthema* (Asteraceae)[308] and *Ipomoea* (Convolvulaceae).[309] The functional CHS genes of *Dendranthema* are composed of three subfamilies, and two of them, SF1 and SF3, code for typical CHS enzymes and the remaining one, SF2, codes for another enzyme that is different from CHS in substrate specificity and reaction. The ω ratio for the lineage ancestral to SF2 was much higher than for other lineages, and some codons in the gene were under positive selection with $\omega > 1$. Thus, positive selection appeared to have driven the divergence of SF2 genes from SF1 and SF3. Like *Dendranthema*, among five functional CHS genes (A–E) of *Ipomoea*, positive selection appeared to have promoted the divergence of CHS A, B, and C genes (CHS A and B produce bisnoryangonin but not naringenin chalcone (**39**)) compared with CHS D and E genes (CHS D and E produce (**39**)). In the case of Asteraceae-specific F3′5′H gene that evolved from F3′H gene within the Asteraceae (see Section 1.24.3.4.2(ii)(a)),[159] the ω ratios for the F3′5′H sequences were elevated around 1.4 times than for the F3′Hs, although the ω ratio was substantially smaller than 1. The requirement for blue or purple flower colors may have exerted a directed positive selection on the evolution of Asteraceae-specific F3′5′H gene, since F3′5′H has provided the basis for purple to blue flower colors by the 3′,4′,5′-hydroxylation of delphinidin derivatives.

On the other hand, escape from adaptive conflict (EAC), in which a single-copy gene is selected to perform a novel function while maintaining its ancestral function, has been proposed as adaptive process involving gene duplication. Adaptive evolutionary change in a duplicated gene of DFR in the genus *Ipomoea* was demonstrated as EAC.[310] In the genus *Ipomoea*, DFR consists of three gene families, DFR A, B, and C, and they form a clade in a phylogenetic tree indicating a lineage-specific duplication in the Convolvulaceae. The base of the clade containing the Convolvulaceae DFR gene family was under purifying selection with low ω ratio like the other DFRs. After the first duplication, the lineage subtending DFR B has a ratio statistically indistinguishable from the base of the Convolvulaceae clade. By contrast, the ratio has significantly increased on the lineage subtending DFR A and C clade with $\omega > 1$, indicating the action of positive selection. Although DFR A and C exhibit essentially no activity on any of the DFR substrates (dihydrokaempferol (**61**), dihydroquercetin (**81**), dihydromyricetin (**82**), naringenin (**54**), and erodyctiol (**73**)), activity for the DFR B on all five substrates is higher than for the DFR belonging to the sister group of the Convolvulaceae. The increased DFR B activities without a signature of positive selection indicate that adaptive change in DFR of the Convolvulaceae is best explained as EAC.

Evolution of enzymes involved in the biosynthesis of flavonoid skeletons and modification of flavonoids, and furthermore, of transcription factors that regulate metabolic pathways[311] must have achieved an extensive diversification of flavonoids, and then diversified flavonoids have come to play a wide range of roles in plant growth and development. Functional roles of flavonoids have been reviewed for plant resistance against biotic and abiotic stress[312–314] and for developmental regulators and chemical messengers.[313,315,316]

References

1. K. Kai; M. Mizutani; N. Kawamura; R. Yamamoto; M. Tamai; H. Yamaguchi; K. Sakata; B. I. Shimizu, *Plant J.* **2008**, *55*, 989–999.
2. T. Nakayama; K. Yonekura-Sakakibara; T. Sato; S. Kikuchi; Y. Fukui; M. Fukuchi-Mizutani; T. Ueda; M. Nakao; Y. Tanaka; T. Kusumi; T. Nishino, *Science* **2000**, *290*, 1163–1166.
3. T. Akashi; T. Aoki; S. Ayabe, *Plant Physiol.* **2005**, *137*, 882–891.
4. H. Ritter; G. E. Schulz, *Plant Cell* **2004**, *16*, 3426–3436.
5. M. J. MacDonald; G. B. D'Cunha, *Biochem. Cell Biol.* **2007**, *85*, 273–282.
6. J. L. Ferrer; M. B. Austin; C. Stewart, Jr.; J. P. Noel, *Plant Physiol. Biochem.* **2008**, *46*, 356–370.
7. S. Rupasinghe; J. Baudry; M. A. Schuler, *Protein Eng.* **2003**, *16*, 721–731.
8. G. A. Schoch; R. Attias; M. Le Ret; D. Werck-Reichhart, *Eur. J. Biochem.* **2003**, *270*, 3684–3695.
9. G. A. Schoch; R. Attias; M. Belghazi; P. M. Dansette; D. Werck-Reichhart, *Plant Physiol.* **2003**, *133*, 1198–1208.
10. Y. Sawada; K. Kinoshita; T. Akashi; T. Aoki; S. Ayabe, *Plant J.* **2002**, *31*, 555–564.
11. X. Z. He; X. Wang; R. A. Dixon, *J. Biol. Chem.* **2006**, *281*, 34441–34447.
12. I. E. Burbulis; B. Winkel-Shirley, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 12929–12934.
13. S. Rasmussen; R. A. Dixon, *Plant Cell* **1999**, *11*, 1537–1551.
14. L. Achnine; E. B. Blancaflor; S. Rasmussen; R. A. Dixon, *Plant Cell* **2004**, *16*, 3098–3109.
15. B. S. Winkel, *Annu. Rev. Plant Biol.* **2004**, *55*, 85–107.

16. K. Jørgensen; A. V. Rasmussen; M. Morant; A. H. Nielsen; N. Bjarnholt; M. Zagrobelny; S. Bak; B. L. Møller, *Curr. Opin. Plant Biol.* **2005**, *8*, 280–291.
17. D. K. Ro; C. J. Douglas, *J. Biol. Chem.* **2004**, *279*, 2600–2607.
18. D. E. Saslowsky; U. Warek; B. S. Winkel, *J. Biol. Chem.* **2005**, *280*, 23735–23740.
19. K. Sasaki; K. Mito; K. Ohara; H. Yamamoto; K. Yazaki, *Plant Physiol.* **2008**, *146*, 1075–1084.
20. E. Ono; M. Hatayama; Y. Isono; T. Sato; R. Watanabe; K. Yonekura-Sakakibara; M. Fukuchi-Mizutani; Y. Tanaka; T. Kusumi; T. Nishino; T. Nakayama, *Plant J.* **2006**, *45*, 133–143.
21. A. M. Boudet, *Phytochemistry* **2007**, *68*, 2722–2735.
22. M. Petersen, *Phytochemistry* **2007**, *68*, 2847–2860.
23. G. Schoch; S. Goepfert; M. Morant; A. Hehn; D. Meyer; P. Ullmann; D. Werck-Reichhart, *J. Biol. Chem* **2001**, *276*, 36566–36574.
24. R. B. Nair; Q. Xia; C. J. Kartha; E. Kurylo; R. N. Hirji; R. Datla; G. Selvaraj, *Plant Physiol.* **2002**, *130*, 210–220.
25. R. Franke; J. M. Humphreys; M. R. Hemm; J. W. Denault; M. O. Ruegger; J. C. Cusumano; C. Chapple, *Plant J.* **2002**, *30*, 33–45.
26. M. Morant; G. A. Schoch; P. Ullmann; T. Ertunc; D. Little; C. E. Olsen; M. Petersen; J. Negrel; D. Werck-Reichhart, *Plant Mol. Biol.* **2007**, *63*, 1–19.
27. M. Matsuno; A. Nagatsu; Y. Ogihara; B. E. Ellis; H. Mizukami, *FEBS Lett.* **2002**, *514*, 219–224.
28. D. R. Gang; T. Beuerle; P. Ullmann; D. Werck-Reichhart; E. Pichersky, *Plant Physiol.* **2002**, *130*, 1536–1544.
29. D. R. Gang, *Annu. Rev. Plant. Biol.* **2005**, *56*, 301–325.
30. E. Pichersky; J. P. Noel; N. Dudareva, *Science* **2006**, *311*, 808–811.
31. T. Koeduka; E. Fridman; D. R. Gang; D. G. Vassao; B. L. Jackson; C. M. Kish; I. Orlova; S. M. Spassova; N. G. Lewis; J. P. Noel; T. J. Baiga; N. Dudareva; E. Pichersky, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10128–10133.
32. T. Koeduka; G. V. Louie; I. Orlova; C. M. Kish; M. Ibdah; C. G. Wilkerson; M. E. Bowman; T. J. Baiga; J. P. Noel; N. Dudareva; E. Pichersky, *Plant J.* **2008**, *54*, 362–374.
33. Y. Kallberg; U. Oppermann; H. Jornvall; B. Persson, *Protein Sci.* **2002**, *11*, 636–641.
34. U. Oppermann; C. Filling; M. Hult; N. Shafqat; X. Wu; M. Lindh; J. Shafqat; E. Nordling; Y. Kallberg; B. Persson; H. Jornvall, *Chem. Biol. Interact.* **2003**, *143–144*, 247–253.
35. G. Labesse; A. Vidal-Cros; J. Chomilier; M. Gaudry; J. P. Mornon, *Biochem. J.* **1994**, *304*, 95–99.
36. D. R. Gang; N. Lavid; C. Zubieta; F. Chen; T. Beuerle; E. Lewinsohn; J. P. Noel; E. Pichersky, *Plant Cell* **2002**, *14*, 505–519.
37. T. Koeduka; T. J. Baiga; J. P. Noel; E. Pichersky, *Plant Physiol.* **2009**, *149*, 384–394.
38. C. Ramirez-Ahumada Mdel; B. N. Timmermann; D. R. Gang, *Phytochemistry* **2006**, *67*, 2017–2029.
39. S. Brand; D. Holscher; A. Schierhorn; A. Svatos; J. Schroder; B. Schneider, *Planta* **2006**, *224*, 413–428.
40. Y. Katsuyama; M. Matsuzawa; N. Funa; S. Horinouchi, *J. Biol. Chem.* **2007**, *282*, 37702–37709.
41. Y. Katsuyama; T. Kita; N. Funa; S. Horinouchi, *J. Biol. Chem.* **2009**, *284*, 11160–11170.
42. U. Matern; P. Lüer; D. Kreusch, Biosynthesis of Coumarins. In *Comprehensive Natural Products Chemistry, Vol. 1: Polyketides and Other Secondary Metabolites Including Fatty Acids and Their Derivatives*; U. Sankawa, Ed.; Elsevier: Oxford, 1999pp 623–638.
43. K. Kai; B. Shimizu; M. Mizutani; K. Watanabe; K. Sakata, *Phytochemistry* **2006**, *67*, 379–386.
44. R. Larbat; S. Kellner; S. Specker; A. Hehn; E. Gontier; J. Hans; F. Bourgaud; U. Matern, *J. Biol. Chem.* **2007**, *282*, 542–554.
45. C. NDong; D. Anzellotti; R. K. Ibrahim; N. P. Huner; F. Sarhan, *J. Biol. Chem.* **2003**, *278*, 6854–6861.
46. M. Hehmann; R. Lukacin; H. Ekiert; U. Matern, *Eur. J. Biochem.* **2004**, *271*, 932–940.
47. J. Chong; R. Baltz; C. Schmitt; R. Beffa; B. Fritig; P. Saindrenan, *Plant Cell* **2002**, *14*, 1093–1107.
48. C. Gachon; R. Baltz; P. Saindrenan, *Plant Mol. Biol.* **2004**, *54*, 137–146.
49. A. Matros; H. P. Mock, *Plant Cell Physiol.* **2004**, *45*, 1185–1193.
50. G. Taguchi; T. Yazawa; N. Hayashida; M. Okazaki, *Eur. J. Biochem.* **2001**, *268*, 4086–4094.
51. G. Taguchi; T. Ubukata; N. Hayashida; H. Yamamoto; M. Okazaki, *Arch. Biochem. Biophys.* **2003**, *420*, 95–102.
52. G. Forkmann; W. Heller, Biosynthesis of Flavonoids. In *Comprehensive Natural Products Chemistry, Vol. 1: Polyketides and Other Secondary Metabolites Including Fatty Acids and Their Derivatives*; U. Sankawa, Ed.; Elsevier: Oxford, 1999pp 713–748.
53. M. B. Austin; A. J. P. Noel, *Nat. Prod. Rep.* **2003**, *20*, 79–110.
54. T. Aoki; T. Akashi; S. Ayabe, *J. Plant Res.* **2000**, *113*, 475–488.
55. S. Oguro; T. Akashi; S. Ayabe; H. Noguchi; I. Abe, *Biochem. Biophys. Res. Commun.* **2004**, *325*, 561–567.
56. E. K. Bomati; M. B. Austin; M. E. Bowman; R. A. Dixon; J. P. Noel, *J. Biol. Chem.* **2005**, *280*, 30496–30503.
57. N. Shimada; S. Sato; T. Akashi; Y. Nakamura; S. Tabata; S. Ayabe; T. Aoki, *DNA Res.* **2007**, *14*, 25–36.
58. K. M. Davies; S. J. Bloor; G. B. Spiller; S. C. Deroles, *Plant J.* **1998**, *13*, 259–266.
59. J. Y. Joung; G. M. Kasthuri; J. Y. Park; W. J. Kang; H. S. Kim; B. S. Yoon; H. Joung; J. H. Jeon, *Biochem. Biophys. Res. Commun.* **2003**, *303*, 326–331.
60. N. Shimada; T. Nakatsuka; M. Nishihara; S. Yamamura; S. Ayabe; T. Aoki, *Plant Biotechnol.* **2006**, *23*, 509–513.
61. T. L. Graham; M. Y. Graham; S. Subramanian; O. Yu, *Plant Physiol.* **2007**, *144*, 728–740.
62. M. A. Ramadan; M. S. Kamel; K. Ohtani; R. Kasai; K. Yamasaki, *Phytochemistry* **2000**, *54*, 891–896.
63. A. M. do Nascimento; D. C. R. de Oliveira, *Biochem. Syst. Ecol.* **2004**, *32*, 1079–1081.
64. J. R. Dimmock; D. W. Elias; M. A. Beazely; N. M. Kandepu, *Curr. Med. Chem.* **1999**, *6*, 1125–1149.
65. M. L. Go; X. Wu; X. L. Liu, *Curr. Med. Chem.* **2005**, *12*, 483–499.
66. N. C. Veitch; R. J. Grayer, Chalcones, Dihydrochalcones, and Aurones. In *Flavonoids: Chemistry, Biochemistry and Applications*; Ø.M. Andersen, K. R. Markham, Eds.; CRC Press: Boca Raton, FL, 2006pp 1003–1100.
67. C. A. Maxwell; M. J. Harrison; R. A. Dixon, *Plant J.* **1993**, *4*, 971–981.
68. C. Zubieta; X. Z. He; R. A. Dixon; J. P. Noel, *Nat. Struct. Biol.* **2001**, *8*, 271–279.
69. K. C. Lam; R. K. Ibrahim; B. Behdad; S. Dayanandan, *Genome* **2007**, *50*, 1001–1013.
70. M. Haga; T. Akashi; T. Aoki; S. Ayabe, *Plant Physiol.* **1997**, *113*, 663.
71. G. Wimmer; H. Halbwirth; F. Wurst; G. Forkmann; K. Stich, *Phytochemistry* **1998**, *47*, 1013–1016.
72. K. Schlangen; S. Miosic; F. Topuz; G. Muster; T. Marosits; C. Seitz; H. Halbwirth, *Plant Sci.* **2009**, *117*, 97–102.

73. H. Halbwirth; G. Wimmer; F. Wurst; G. Forkmann; K. Stich, *Plant Sci.* **1997**, *122*, 125–131.
74. E. Ono; M. Fukuchi-Mizutani; N. Nakamura; Y. Fukui; K. Yonekura-Sakakibara; M. Yamaguchi; T. Nakayama; T. Tanaka; T. Kusumi; Y. Tanaka, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11075–11080.
75. J. Ogata; Y. Yoshio Itoh; M. Ishida; H. Yoshida; Y. Ozeki, *Plant Biotechnol.* **2004**, *21*, 367–375.
76. Y. Itoh; D. Higeta; A. Suzuki; H. Yoshida; Y. Ozeki, *Plant Cell Physiol.* **2002**, *43*, 578–585.
77. H. Yoshida; Y. Itoh; Y. Ozeki; T. Iwashina; M. Yamaguchi, *Sci. Hortic.* **2004**, *99*, 175–186.
78. K. Marinova; K. Kleinschmidt; G. Weissenbock; M. Klein, *Plant Physiol.* **2007**, *144*, 432–444.
79. S. Reuber; B. Jende-Strid; V. Wray; G. Weissenbock, *Physiol. Plantarum* **1997**, *101*, 827–832.
80. L. Li; L. V. Modolo; L. L. Escamilla-Trevino; L. Achnine; R. A. Dixon; X. Wang, *J. Mol. Biol.* **2007**, *370*, 951–963.
81. A. Vitali; B. Giardina; G. Delle Monache; F. Rocca; A. Silvestrini; A. Tafi; B. Botta, *FEBS Lett.* **2004**, *557*, 33–38.
82. J. F. Stevens; J. E. Page, *Phytochemistry* **2004**, *65*, 1317–1330.
83. J. Nagel; L. K. Culley; Y. P. Lu; E. W. Liu; P. D. Matthews; J. F. Stevens; J. E. Page, *Plant Cell* **2008**, *20*, 186–200.
84. X. C. Li; H. N. ElSohly; A. C. Nimrod; A. M. Clark, *J. Nat. Prod.* **1999**, *62*, 767–769.
85. J. B. Harborne; H. Baxter, *The Handbook of Natural Flavonoids*; John Wiley & Sons: Chichester, 1999.
86. Atta-Ur-Rahman; M. I. Choudhary; S. Hayat; A. M. Khan; A. Ahmed, *Chem. Pharm. Bull.* **2001**, *49*, 105–107.
87. T. Nakayama; T. Sato; Y. Fukui; K. Yonekura-Sakakibara; H. Hayashi; Y. Tanaka; T. Kusumi; T. Nishino, *FEBS Lett.* **2001**, *499*, 107–111.
88. Y. Tanaka; N. Sasaki; A. Ohmiya, *Plant J.* **2008**, *54*, 733–749.
89. M. A. Farag; B. E. Deavours; A. de Fatima; M. Naoumkina; R. A. Dixon; L. W. Sumner, *Plant Physiol.* **2009**, in press.
90. Y. Kimura; T. Aoki; S. Ayabe, *Plant Cell Physiol.* **2001**, *42*, 1169–1173.
91. N. Shimada; T. Aoki; S. Sato; Y. Nakamura; S. Tabata; S. Ayabe, *Plant Physiol.* **2003**, *131*, 941–951.
92. D. Ober, *Trends Plant Sci.* **2005**, *10*, 444–449.
93. L. Ralston; S. Subramanian; M. Matsuno; O. Yu, *Plant Physiol.* **2005**, *137*, 1375–1388.
94. H. B. Kim; J. H. Bae; J. D. Lim; C. Y. Yu; C. S. An, *Mol. Cells* **2007**, *23*, 405–409.
95. Y. H. Xiao; Z. S. Zhang; M. H. Yin; M. Luo; X. B. Li; L. Hou; Y. Pei, *Biochem. Biophys. Res. Commun.* **2007**, *358*, 73–78.
96. M. Yamazaki; M. Shibata; Y. Nishiyama; K. Springob; M. Kitayama; N. Shimada; T. Aoki; S. I. Ayabe; K. Saito, *FEBS J.* **2008**, *275*, 3494–3502.
97. R. Stracke; H. Ishihara; G. H. A. Barsch; F. Mehrtens; K. Niehaus; B. Weisshaar, *Plant J.* **2007**, *50*, 660–677.
98. Y. Morita; M. Saitoh; A. Hoshino; E. Nitasaka; S. Iida, *Plant Cell Physiol.* **2006**, *47*, 457–470.
99. J. Boddu; C. H. Jiang; V. Sangar; T. Olson; T. Peterson; S. Chopra, *Plant Mol. Biol.* **2006**, *60*, 185–199.
100. S. R. Muir; G. J. Collins; S. Robinson; S. Hughes; A. Bovy; C. H. R. De Vos; A. J. van Tunen; M. E. Verhoeyen, *Nat. Biotechnol.* **2001**, *19*, 470–474.
101. F. Li; Z. Jin; W. Qu; D. Zhao; F. Ma, *Plant Physiol. Biochem.* **2006**, *44*, 455–461.
102. F. X. Li; Z. P. Jin; D. X. Zhao; L. Q. Cheng; C. X. Fu; F. S. Ma, *Phytochemistry* **2006**, *67*, 553–560.
103. M. Nishihara; T. Nakatsuka; S. Yamamura, *FEBS Lett.* **2005**, *579*, 6074–6078.
104. J. M. Jez; M. E. Bowman; R. A. Dixon; J. P. Noel, *Nat. Struct. Biol.* **2000**, *7*, 786–791.
105. J. M. Jez; M. E. Bowman; J. P. Noel, *Biochemistry* **2002**, *41*, 5168–5176.
106. J. M. Jez; J. P. Noel, *Journal of Biol. Chem.* **2002**, *277*, 1361–1369.
107. S. Hur; Z. E. R. Newby; T. C. Bruice, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2730–2735.
108. J. J. Ruiz-Pernia; E. Silla; I. Tunon, *J. Am. Chem. Soc.* **2007**, *129*, 9117–9124.
109. C. Herles; A. Braune; M. Blaut, *Arch. Microbiol.* **2004**, *181*, 428–434.
110. M. Gensheimer; A. Mushegian, *Protein Sci.* **2004**, *13*, 540–544.
111. H. Mathews; S. K. Clendennen; C. G. Caldwell; X. L. Liu; K. Connors; N. Matheis; D. K. Schuster; D. J. Menasco; W. Wagoner; J. Lightner; D. R. Wagner, *Plant Cell* **2003**, *15*, 1689–1703.
112. S. Martens; A. Mithöfer, *Phytochemistry* **2005**, *66*, 2399–2407.
113. R. Lukacin; U. Matern; K. T. Junghanns; M. L. Heskamp; L. Britsch; G. Forkmann; S. Martens, *Arch. Biochem. Biophys.* **2001**, *393*, 177–183.
114. S. Martens; G. Forkmann; U. Matern; R. Lukacin, *Phytochemistry* **2001**, *58*, 43–46.
115. T. Akashi; T. Aoki; S. Ayabe, *FEBS Lett.* **1998**, *431*, 287–290.
116. S. Martens; G. Forkmann, *Plant J.* **1999**, *20*, 611–618.
117. T. Akashi; T. Aoki; Si. Ayabe, *Plant Physiol.* **1999**, *121*, 821–828.
118. C. Kitada; Z. Gong; Y. Tanaka; M. Yamazaki; K. Saito, *Plant Cell Physiol.* **2001**, *42*, 1338–1344.
119. T. Nakatsuka; M. Nishihara; K. Mishiba; S. Yamamura, *Plant Sci.* **2005**, *168*, 1309–1318.
120. J. Zhang; S. Subramanian; Y. Zhang; O. Yu, *Plant Physiol.* **2007**, *144*, 741–751.
121. L. Lepiniec; I. Debeaujon; J.-M. Routaboul; A. Baudry; L. Pourcel; N. Nesi; M. Caboche, *Annu. Rev. Plant Biol.* **2006**, *57*, 405–430.
122. E. Wisman; U. Hartmann; M. Sagasser; E. Baumann; K. Palme; K. Hahlbrock; H. Saedler; B. Weisshaar, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12432–12437.
123. R. Lukacin; C. Urbanke; I. Gröning; U. Matern, *FEBS Lett.* **2000**, *467*, 353–358.
124. F. Wellmann; R. Lukacin; T. Moriguchi; L. Britsch; E. Schiltz; U. Matern, *Eur. J. Biochem.* **2002**, *269*, 4134–4142.
125. A. G. Prescott; N. P. J. Stamford; G. Wheeler; J. L. Firmin, *Phytochemistry* **2002**, *60*, 589–593.
126. R. Lukacin; F. Wellmann; L. Britsch; S. Martens; U. Matern, *Phytochemistry* **2003**, *62*, 287–292.
127. R. Takahashi; S. M. Githiri; K. Hatayama; E. G. Dubouzet; N. Shimada; T. Aoki; S. I. Ayabe; T. Iwashina; K. Toda; H. Matsumura, *Plant Mol. Biol.* **2006**, *63*, 125–135.
128. D. K. Owens; A. B. Alerding; K. C. Crosby; A. B. Bandara; J. H. Westwood; B. S. Winkel, *Plant Physiol.* **2008**, *147*, 1046–1061.
129. K. Saito; M. Kobayashi; Z. Gong; Y. Tanaka; M. Yamazaki, *Plant J.* **1999**, *17*, 181–189.
130. J. Nakajima; Y. Tanaka; M. Yamazaki; K. Saito, *J. Biol. Chem.* **2001**, *276*, 25797–25803.
131. J. J. Turnbull; W. J. Sobey; R. T. Aplin; A. Hassan; C. J. Schofield; J. L. Firmin; A. G. Prescott, *Chem. Commun.* **2000**, 2473–2474.

132. R. W. D. Welford; J. J. Turnbull; T. D. W. Claridge; C. J. Schofield; A. G. Prescott, *Chem. Commun.* **2001**, *2001*, 1828–1829.
133. R. W. D. Welford; I. J. Clifton; J. J. Turnbull; S. C. Wilson; C. J. Schofield, *Org. Biomol. Chem.* **2005**, *3*, 3117–3126.
134. J. -i Nakajima; Y. Sato; T. Hoshino; M. Yamazaki; K. Saito, *J. Biol. Chem.* **2006**, *281*, 21387–21398.
135. F. Wellmann; M. Griesser; W. Schwab; S. Martens; W. Eisenreich; U. Matern; R. Lukacin, *FEBS Lett.* **2006**, *580*, 1642–1648.
136. N. Shikazono; Y. Yokota; S. Kitamura; C. Suzuki; H. Watanabe; S. Tano; A. Tanaka, *Genetics* **2003**, *163*, 1449–1455.
137. S. Abrahams; E. Lee; A. R. Walker; G. J. Tanner; P. J. Larkin; A. R. Ashton, *Plant J.* **2003**, *35*, 624–636.
138. S. Shimada; Y. T. Inoue; M. Sakuta, *Plant J.* **2005**, *44*, 950–959.
139. S. Martens; G. Forkmann; L. Britsch; F. Wellmann; U. Matern; R. Lukacin, *FEBS Lett.* **2003**, *544*, 93–98.
140. J. J. Turnbull; J. -I. Nakajima; R. W. D. Welford; M. Yamazaki; K. Saito; C. J. Schofield, *J. Biol. Chem.* **2004**, *279*, 1206–1216.
141. Y. H. Gebhardt; S. Witte; H. Steuber; U. Matern; S. Martens, *Plant Physiol.* **2007**, *144*, 1442–1454.
142. J. Bernhardt; K. Stich; Z. Schwarz-Sommer; H. Saedler; U. Wienand, *Plant J.* **1998**, *14*, 483–488.
143. M. Chen; P. SanMiguel; J. L. Bennetzen, *Genetics* **1998**, *148*, 435–444.
144. Y. Inagaki; Y. Johzuka-Hisatomi; T. Mori; S. Takahashi; Y. Hayakawa; S. Peyachoknagul; Y. Ozeki; S. Iida, *Gene* **1999**, *226*, 181–188.
145. E. Himi; K. Noda, *J. Exp. Bot.* **2004**, *55*, 365–375.
146. N. Shimada; R. Sasaki; S. Sato; T. Kaneko; S. Tabata; T. Aoki; S. -i Ayabe, *J. Exp. Bot.* **2005**, *56*, 2573–2585.
147. R. A. Dixon; D. -Y. Xie; S. B. Sharma, *New Phytol.* **2005**, *165*, 9–28.
148. S. Albert; M. Delseny; M. Devic, *Plant J.* **1997**, *11*, 289–299.
149. D. -Y. Xie; S. B. Sharma; N. L. Paiva; D. Ferreira; R. A. Dixon, *Science* **2003**, *299*, 396–399.
150. D. Y. Xie; S. B. Sharma; R. A. Dixon, *Arch. Biochem. Biophys.* **2004**, *422*, 91–102.
151. G. J. Tanner; K. T. Francki; S. Abrahams; J. M. Watson; P. J. Larkin; A. R. Ashton, *J. Biol. Chem.* **2003**, *278*, 31647–31656.
152. Y. Pang; G. J. Peel; E. Wright; Z. Wang; R. A. Dixon, *Plant Physiol.* **2007**, *145*, 601–615.
153. F. Brugliera; G. Barri-Rewell; T. A. Holton; J. G. Mason, *Plant J.* **1999**, *19*, 441–451.
154. C. Schoenbohm; S. Martens; C. Eder; G. Forkmann; B. Weisshaar, *Biol. Chem.* **2000**, *381*, 749–753.
155. M. Yamazaki; Z. Gong; M. Fukuchi-Mizutani; Y. Fukui; Y. Tanaka; T. Kusumi; K. Saito, *J. Biol. Chem.* **1999**, *274*, 7405–7411.
156. M. Yamazaki; M. Shibata; Y. Nishiyama; K. Springob; M. Kitayama; N. Shimada; T. Aoki; S. Ayabe; K. Saito, *FEBS J.* **2008**, *275*, 3494–3502.
157. K. Toda; D. Yang; N. Yamanaka; S. Watanabe; K. Harada; R. Takahashi, *Plant Mol. Biol.* **2002**, *50*, 187–196.
158. G. Zabala; L. Vodkin, *Genetics* **2003**, *163*, 295–309.
159. C. Seitz; C. Eder; B. Deiml; S. Kellner; S. Martens; G. Forkmann, *Plant Mol. Biol.* **2006**, *61*, 365–381.
160. C. Seitz; S. Ameres; G. Forkmann, *FEBS Lett.* **2007**, *581*, 3429–3434.
161. A. O. Latunde-Dada; F. Cabello-Hurtado; N. Czittrich; L. Didierjean; C. Schopfer; N. Hertkorn; D. Werck-Reichhart; J. Ebel, *J. Biol. Chem.* **2001**, *276*, 1688–1695.
162. H. Yamamoto; A. Yatou; K. Inoue, *Phytochemistry* **2001**, *58*, 671–676.
163. H. Halbwirth; K. Stich, *Phytochemistry* **2006**, *67*, 1080–1087.
164. A. B. Christensen; P. L. Gregersen; C. E. Olsen; D. B. Collinge, *Plant Mol. Biol.* **1998**, *36*, 219–227.
165. T. S. Huang; D. Anzellotti; F. Dedaldechamp; R. K. Ibrahim, *Plant Physiol.* **2004**, *134*, 1366–1376.
166. A. Gauthier; P. J. Gulick; R. K. Ibrahim, *Arch. Biochem. Biophys.* **1998**, *351*, 243–249.
167. I. Muzac; J. Wang; D. Anzellotti; H. Zhang; R. K. Ibrahim, *Arch. Biochem. Biophys.* **2000**, *375*, 385–388.
168. B. -G. Kim; Y. Lee; H. -G. Hur; Y. Lim; J. -H. Ahn, *Phytochemistry* **2006**, *67*, 387–394.
169. S. Cacace; G. Schröder; E. Wehinger; D. Strack; J. Schmidt; J. Schröder, *Phytochemistry* **2003**, *62*, 127–137.
170. G. Schröder; E. Wehinger; R. Lukacin; F. Wellmann; W. Seefelder; W. Schwab; J. Schröder, *Phytochemistry* **2004**, *65*, 1085–1094.
171. J. M. Zhou; N. D. Gold; V. J. Martin; E. Wollenweber; R. K. Ibrahim, *Biochim. Biophys. Acta* **2006**, *1760*, 1115–1124.
172. S. Paquette; B. L. Møller; S. Bak, *Phytochemistry* **2003**, *62*, 399–413.
173. A. Kubo; Y. Arai; S. Nagashima; T. Yoshikawa, *Arch. Biochem. Biophys.* **2004**, *429*, 198–203.
174. M. Brazier-Hicks; K. M. Evans; M. C. Gershater; H. Puschmann; P. G. Steel; R. Edwards, *J. Biol. Chem.* **2009**, *284*, 17926–17934.
175. C. M. Ford; P. K. Boss; P. B. Hoj, *J. Biol. Chem.* **1998**, *273*, 9224–9233.
176. M. Yamazaki; E. Yamagishi; Z. Gong; M. Fukuchi-Mizutani; Y. Fukui; Y. Tanaka; T. Kusumi; M. Yamaguchi; K. Saito, *Plant Mol. Biol.* **2002**, *48*, 401–411.
177. M. Fukuchi-Mizutani; H. Okuhara; Y. Fukui; M. Nakao; Y. Katsumoto; K. Yonekura-Sakakibara; T. Kusumi; T. Hase; Y. Tanaka, *Plant Physiol.* **2003**, *132*, 1652–1663.
178. J. Ogata; Y. Kanno; Y. Itoh; H. Tsugawa; M. Suzuki, *Nature* **2005**, *435*, 757–758.
179. C. M. Kramer; R. T. N. Prata; M. G. Willits; V. De Luca; J. C. Steffens; G. Graser, *Phytochemistry* **2003**, *64*, 1069–1076.
180. J. Isayenkova; V. Wray; M. Nimtz; D. Strack; T. Vogt, *Phytochemistry* **2006**, *67*, 1598–1612.
181. K. Kogawa; N. Kato; K. Kazuma; N. Noda; M. Suzuki, *Planta* **2007**, *226*, 1501–1509.
182. Y. Pang; G. J. Peel; S. B. Sharma; Y. Tang; R. A. Dixon, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 14210–14215.
183. P. Jones; B. Messner; J. Nakajima; A. R. Schäffner; K. Saito, *J. Biol. Chem.* **2003**, *278*, 43910–43918.
184. J. O. Borevitz; Y. Xia; J. Blount; R. A. Dixon; C. Lamb, *Plant Cell* **2000**, *12*, 2383–2394.
185. T. Tohge; Y. Nishiyama; M. Y. Hirai; M. Yano; J. Nakajima; M. Awazuhara; E. Inoue; H. Takahashi; D. B. Goodenowe; M. Kitayama; M. Noji; M. Yamazaki; K. Saito, *Plant J.* **2005**, *42*, 218–235.
186. K. Yonekura-Sakakibara; T. Tohge; R. Niida; K. Saito, *J. Biol. Chem.* **2007**, *282*, 14932–14941.
187. K. Yonekura-Sakakibara; T. Tohge; F. Matsuda; R. Nakabayashi; H. Takayama; R. Niida; A. Watanabe-Takahashi; E. Inoue; S. Saito, *Plant Cell* **2008**, *20*, 2160–2176.
188. M. Mato; Y. Ozeki; Y. Itoh; D. Higeta; K. Yoshitama; S. Teramoto; R. Aida; N. Ishikura; M. Shibata, *Plant Cell Physiol.* **1998**, *39*, 1145–1155.
189. K. D. Miller; V. Guyon; J. N. Evans; W. A. Shuttleworth; L. P. Taylor, *J. Biol. Chem.* **1999**, *274*, 34011–34019.
190. K. D. Miller; J. Strommer; L. P. Taylor, *Plant Mol. Biol.* **2002**, *48*, 233–242.

191. S. Masada; K. Terasaka; Y. Oguchi; S. Okazaki; T. Mizushima; H. Mizukami, *Plant Cell Physiol.* **2009**, *50*, 1401–1415.
192. A. Frydman; O. Weisshaus; M. Bar-Peled; D. V. Huhman; L. W. Sumner; F. R. Marin; E. Lewinsohn; R. Fluhr; J. Gressel; Y. Eyal, *Plant J.* **2004**, *40*, 88–100.
193. Y. Morita; A. Hoshino; Y. Kikuchi; H. Okuhara; E. Ono; Y. Tanaka; Y. Fukui; N. Saito; E. Nitasaka; H. Noguchi; S. Iida, *Plant J.* **2005**, *42*, 353–363.
194. S. Sawada; H. Suzuki; F. Ichimaida; M. Yamaguchi; T. Iwashita; Y. Fukui; H. Hemmi; T. Nishino; T. Nakayama, *J. Biol. Chem.* **2005**, *280*, 899–906.
195. T. Nakayama; H. Suzuki; T. Nishino, *J. Mol. Catal. B Enzym.* **2003**, *23*, 117–132.
196. Y. Abe; M. Tera; N. Sasaki; M. Okamura; N. Umemoto; M. Momose; N. Kawahara; H. Kamakura; Y. Goda; K. Nagasawa; Y. Ozeki, *Biochem. Biophys. Res. Commun.* **2008**, *373*, 473–477.
197. H. Fujiwara; Y. Tanaka; K. Yonekura-Sakakibara; M. Fukuchi-Mizutani; M. Nakao; Y. Fukui; M. Yamaguchi; T. Ashikari; T. Kusumi, *Plant J.* **1998**, *16*, 421–431.
198. K. Yonekura-Sakakibara; Y. Tanaka; M. Fukuchi-Mizutani; H. Fujiwara; Y. Fukui; T. Ashikari; Y. Murakami; M. Yamaguchi; T. Kusumi, *Plant Cell Physiol.* **2000**, *41*, 495–502.
199. H. Suzuki; T. Nakayama; K. Yonekura-Sakakibara; Y. Fukui; N. Nakamura; M. Nakao; Y. Tanaka; M. A. Yamaguchi; T. Kusumi; T. Nishino, *J. Biol. Chem.* **2001**, *276*, 49013–49019.
200. H. Suzuki; T. Nakayama; T. Nishino, *Biochemistry* **2003**, *42*, 1764–1771.
201. H. Suzuki; S. Sawada; K. Watanabe; S. Nagae; M. A. Yamaguchi; T. Nakayama; T. Nishino, *Plant J.* **2004**, *38*, 994–1003.
202. H. Suzuki; T. Nakayama; K. Yonekura-Sakakibara; Y. Fukui; N. Nakamura; M. -A. Yamaguchi; Y. Tanaka; T. Kusumi; T. Nishino, *Plant Physiol.* **2002**, *130*, 2142–2151.
203. H. Suzuki; T. Nakayama; S. Nagae; M. -A. Yamaguchi; T. Iwashita; Y. Fukui; T. Nishino, *J. Mol. Catal. B Enzym.* **2004**, *28*, 87–93.
204. J. C. D'Auria; M. Reichelt; K. Luck; A. Svatos; J. Gershenzon, *FEBS Lett.* **2007**, *581*, 872–878.
205. J. Luo; Y. Nishiyama; C. Fuell; G. Taguchi; K. Elliott; L. Hill; Y. Tanaka; M. Kitayama; M. Yamazaki; P. Bailey; A. Parr; A. J. Michael; K. Saito; C. Martin, *Plant J.* **2007**, *50*, 678–695.
206. R. A. Dixon, Isoflavonoids: Biochemistry, Molecular Biology, and Biological Functions. In *Comprehensive Natural Products Chemistry, Vol. 1: Polyketides and Other Secondary Metabolites Including Fatty Acids and Their Derivatives*; U. Sankawa, Ed.; Elsevier: Oxford, 1999pp 773–823.
207. O. Yu; B. McGonigle, *Adv. Agron.* **2005**, *86*, 147–190.
208. N. C. Veitch, *Nat. Prod. Rep.* **2007**, *24*, 417–464.
209. L. Tian; Y. Pang; R. A. Dixon, *Phytochem. Rev.* **2008**, *7*, 445–465.
210. J. Reynaud; D. Guilet; R. Terreux; M. Lussignol; N. Walchshofer, *Nat. Prod. Rep.* **2005**, *22*, 504–515.
211. O. Lapcik, *Phytochemistry* **2007**, *68*, 2909–2916.
212. G. Kochs; H. Grisebach, *Eur. J. Biochem.* **1986**, *155*, 311–318.
213. M. F. Hashim; T. Hakamatsuka; Y. Ebizuka; U. Sankawa, *FEBS Lett.* **1990**, *271*, 219–222.
214. T. Hakamatsuka; M. F. Hashim; Y. Ebizuka; U. Sankawa, *Tetrahedron* **1991**, *47*, 5969–5978.
215. T. Akashi; T. Aoki; S. Ayabe, *Plant Physiol.* **1999**, *121*, 821–828.
216. S. Ayabe; T. Akashi; T. Aoki, *Methods Enzymol.* **2002**, *357*, 360–369.
217. B. Siminszky; F. T. Corbin; E. R. Ward; T. J. Fleischmann; R. E. Dewey, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1750–1755.
218. C. L. Steele; M. Gijzen; D. Qutob; R. A. Dixon, *Arch. Biochem. Biophys.* **1999**, *367*, 146–150.
219. W. Jung; O. Yu; S. M. Lau; D. P. O'Keefe; J. Odell; G. Fader; B. McGonigle, *Nat. Biotechnol.* **2000**, *18*, 208–212.
220. N. Shimada; T. Akashi; T. Aoki; S. Ayabe, *Plant Sci.* **2000**, *160*, 37–47.
221. B. G. Kim; S. Y. Kim; H. S. Song; C. Lee; H. G. Hur; S. I. Kim; J. H. Ahn, *Mol. Cells* **2003**, *15*, 301–306.
222. B. Deavours; R. Dixon, *Plant Physiol.* **2005**, *138*, 2245–2259.
223. Y. Sawada; S. Ayabe, *Biochem. Biophys. Res. Commun.* **2005**, *330*, 907–913.
224. T. Hakamatsuka; K. Mori; S. Ishida; Y. Ebizuka; U. Sankawa, *Phytochemistry* **1998**, *49*, 497–505.
225. S. D. Marshall; J. J. Putterill; K. M. Plummer; R. D. Newcomb, *J. Mol. Evol.* **2003**, *57*, 487–500.
226. M. C. Gershater; I. Cummins; R. Edwards, *J. Biol. Chem.* **2007**, *282*, 21460–21466.
227. M. Ueguchi-Tanaka; M. Ashikari; M. Nakajima; H. Itoh; E. Katoh; M. Kobayashi; T. Y. Chow; Y. I. Hsing; H. Kitano; I. Yamaguchi; M. Matsuoka, *Nature* **2005**, *437*, 693–698.
228. X. Z. He; R. A. Dixon, *Arch. Biochem. Biophys.* **1996**, *336*, 121–129.
229. X. Z. He; J. T. Reddy; R. A. Dixon, *Plant Mol. Biol.* **1998**, *36*, 43–54.
230. X. Z. He; R. A. Dixon, *Plant Cell* **2000**, *12*, 1689–1702.
231. T. Akashi; Y. Sawada; T. Aoki; S. Ayabe, *Biosci. Biotechnol. Biochem.* **2000**, *64*, 2276–2279.
232. C. J. Liu; R. A. Dixon, *Plant Cell* **2001**, *13*, 2643–2658.
233. T. Akashi; Y. Sawada; N. Shimada; N. Sakurai; T. Aoki; S. Ayabe, *Plant Cell Physiol.* **2003**, *44*, 103–112.
234. Q. Wu; C. L. Preisig; H. D. VanEtten, *Plant Mol. Biol.* **1997**, *35*, 551–560.
235. C. J. Liu; B. E. Deavours; S. B. Richard; J. L. Ferrer; J. W. Blount; D. Huhman; R. A. Dixon; J. P. Noel, *Plant Cell* **2006**, *18*, 3656–3669.
236. B. E. Deavours; C. J. Liu; M. A. Naoumkina; Y. Tang; M. A. Farag; L. W. Sumner; J. P. Noel; R. A. Dixon, *Plant Mol. Biol.* **2006**, *62*, 715–733.
237. T. Akashi; H. D. VanEtten; Y. Sawada; C. C. Wasmann; H. Uchiyama; S. Ayabe, *Phytochemistry* **2006**, *67*, 2525–2530.
238. S. Nagashima; R. Inagaki; A. Kubo; M. Hirotani; T. Yoshikawa, *Planta* **2004**, *218*, 456–459.
239. A. Noguchi; A. Saito; Y. Homma; M. Nakao; N. Sasaki; T. Nishino; S. Takahashi; T. Nakayama, *J. Biol. Chem.* **2007**, *282*, 23581–23590.
240. S. Dhaubhadel; M. Farhangkhoee; R. Chapman, *J. Exp. Bot.* **2008**, *59*, 981–994.
241. L. Achnine; D. V. Huhman; M. A. Farag; L. W. Sumner; J. W. Blount; R. A. Dixon, *Plant J.* **2005**, *41*, 875–887.
242. H. Shao; X. He; L. Achnine; J. W. Blount; R. A. Dixon; X. Wang, *Plant Cell* **2005**, *17*, 3141–3154.
243. L. V. Modolo; J. W. Blount; L. Achnine; M. A. Naoumkina; X. Wang; R. A. Dixon, *Plant Mol. Biol.* **2007**, *64*, 499–518.
244. H. Suzuki; T. Nishino; T. Nakayama, *Phytochemistry* **2007**, *68*, 2035–2042.

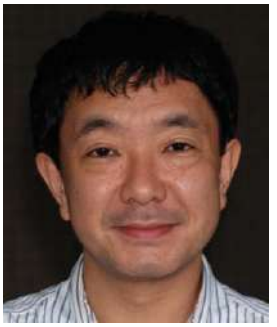
245. J. C. D'Auria, *Curr. Opin. Plant Biol.* **2006**, *9*, 331–340.
246. M. -C. Hsieh; T. L. Graham, *Phytochemistry* **2001**, *58*, 995–1005.
247. H. Suzuki; S. Takahashi; R. Watanabe; Y. Fukushima; N. Fujita; A. Noguchi; R. Yokoyama; K. Nishitani; T. Nishino; T. Nakayama, *J. Biol. Chem.* **2006**, *281*, 30251–30259.
248. M. Naoumkina; M. A. Farag; L. W. Sumner; Y. Tang; C. J. Liu; R. A. Dixon, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 17909–17915.
249. P. Chuankhayan; Y. Hua; J. Svasti; S. Sakdarat; P. A. Sullivan; J. R. Ketudat Cairns, *Phytochemistry* **2005**, *66*, 1880–1889.
250. P. Chuankhayan; T. Rimlumduan; W. Tantanuch; N. Mothong; P. T. Kongsaeree; P. Metheenukul; J. Svasti; O. N. Jensen; J. R. Cairns, *Arch Biochem Biophys* **2007**, *468*, 205–216.
251. C. Srisomsap; J. Svasti; R. Surarit; V. Champattanachai; P. Sawangareetrakul; K. Boonpuan; P. Subhasitanont; D. Chokchaichamnankit, *J. Biochem.* **1996**, *119*, 585–590.
252. J. R. Ketudat Cairns; V. Champattanachai; C. Srisomsap; B. Wittman-Liebold; B. Thiede; J. Svasti, *J. Biochem.* **2000**, *128*, 999–1008.
253. P. M. Dewick, Isoflavonoids. In *The Flavonoids. Advances in Research Since 1986*; J. B. Harborne, Ed.; Chapman and Hall: London, 1994pp 117–238.
254. W. Heller; G. Forkmann, Biosynthesis of Flavonoids. In *The Flavonoids. Advances in Research Since 1986*; J. B. Harborne, Ed.; Chapman and Hall: London, 1994pp 499–535.
255. S. Clemens; W. Hinderer; U. Wittkampf; W. Barz, *Phytochemistry* **1993**, *32*, 653–657.
256. W. Hinderer; U. Flentje; W. Barz, *FEBS Lett.* **1987**, *214*, 101–106.
257. T. Akashi; T. Aoki; S. Ayabe, *Biochem. Biophys. Res. Commun.* **1998**, *251*, 67–70.
258. S. Overkamp; F. Hein; W. Barz, *Plant Sci.* **2000**, *155*, 101–108.
259. C. -J. Liu; D. Huhman; L. W. Sumner; R. A. Dixon, *Plant J.* **2003**, *36*, 471–484.
260. N. L. Paiva; R. Edwards; Y. J. Sun; G. Hrazdina; R. A. Dixon, *Plant Mol. Biol.* **1991**, *17*, 653–667.
261. L. Guo; N. L. Paiva, *Arch. Biochem. Biophys.* **1995**, *320*, 353–360.
262. X. Wang; X. He; J. Lin; H. Shao; Z. Chang; R. A. Dixon, *J. Mol. Biol.* **2006**, *358*, 1341–1352.
263. H. Shao; R. Dixon; X. Wang, *J. Mol. Biol.* **2007**, *369*, 265–276.
264. L. Guo; R. A. Dixon; N. L. Paiva, *J. Biol. Chem.* **1994**, *269*, 22372–22378.
265. L. Guo; R. A. Dixon; N. L. Paiva, *FEBS Lett.* **1994**, *356*, 221–225.
266. M. R. Bonde; R. L. Millar; J. L. Ingham, *Phytochemistry* **1973**, *12*, 2957–2959.
267. P. Dewick; M. Martin, *Phytochemistry* **1979**, *18*, 591–596.
268. D. R. Gang; H. Kasahara; Z. Q. Xia; K. Vander Mijnsbrugge; G. Bauw; W. Boerjan; M. Van Montagu; L. B. Davin; N. G. Lewis, *J. Biol. Chem.* **1999**, *274*, 7516–7527.
269. T. Akashi; S. Koshimizu; T. Aoki; S. Ayabe, *FEBS Lett.* **2006**, *580*, 5666–5670.
270. C. R. Schopfer; G. Kochs; F. Lottspeich; J. Ebel, *FEBS Lett.* **1998**, *432*, 182–186.
271. T. Akashi; K. Sasaki; T. Aoki; S. Ayabe; K. Yazaki, *Plant Physiol.* **2009**, *149*, 683–693.
272. E. Kaimoyo; H. VanEtten, *Phytochemistry* **2008**, *69*, 76–87.
273. N. L. Paiva; Y. Sun; R. A. Dixon; H. D. VanEtten; G. Hrazdina, *Arch. Biochem. Biophys.* **1994**, *312*, 501–510.
274. G. DiCenzo; H. VanEtten, *Phytochemistry* **2006**, *67*, 675–683.
275. Q. Wu; H. D. VanEtten, *Mol. Plant Microbe Interact.* **2004**, *17*, 798–804.
276. R. A. Dixon; C. L. Steele, *Trends Plant Sci.* **1999**, *4*, 394–400.
277. J. M. Humphreys; C. Chapple, *Trends Plant Sci.* **2000**, *5*, 271–272.
278. O. Yu; W. Jung; J. Shi; R. A. Croes; G. M. Fader; B. McGonigle; J. T. Odell, *Plant Physiol.* **2000**, *124*, 781–793.
279. C. J. Liu; J. W. Blount; C. L. Steele; R. A. Dixon, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14578–14583.
280. V. S. Sreevidya; C. Srinivasa Rao; S. B. Sullia; J. K. Ladha; P. M. Reddy, *J. Exp. Bot.* **2006**, *57*, 1957–1969.
281. C. H. Shih; Y. Chen; M. Wang; I. K. Chu; C. Lo, *J. Agric. Food Chem.* **2008**, *56*, 5655–5661.
282. R. Liu; Y. Hu; J. Li; Z. Lin, *Metab. Eng.* **2007**, *9*, 1–7.
283. L. Tian; R. Dixon, *Planta* **2006**, *224*, 496–507.
284. W. Jung; I. -M. Chung; H. -Y. Heo, *J. Plant Biotechnol.* **2003**, *5*, 149–155.
285. S. Subramanian; M. Y. Graham; O. Yu; T. L. Graham, *Plant Physiol.* **2005**, *137*, 1345–1353.
286. S. Subramanian; G. Stacey; O. Yu, *Plant J.* **2006**, *48*, 261–273.
287. S. Subramanian; G. Stacey; O. Yu, *Trends Plant Sci.* **2007**, *12*, 282–285.
288. M. Shimamura; T. Akashi; N. Sakurai; H. Suzuki; K. Saito; D. Shibata; S. Ayabe; T. Aoki, *Plant Cell Physiol.* **2007**, *48*, 1652–1657.
289. K. Saito; M. Y. Hirai; K. Yonekura-Sakakibara, *Trends Plant Sci.* **2008**, *13*, 36–43.
290. A. Rohde; K. Morreel; J. Ralph; G. Goeminne; V. Hostyn; R. De Rycke; S. Kushnir; J. Van Doorsselaere; J. P. Joseleau; M. Vuylsteke; G. Van Driessche; J. Van Beeumen; E. Messens; W. Boerjan, *Plant Cell* **2004**, *16*, 2749–2771.
291. M. A. Farag; D. V. Huhman; R. A. Dixon; L. W. Sumner, *Plant Physiol.* **2008**, *146*, 387–402.
292. H. Suzuki; R. Sasaki; Y. Ogata; Y. Nakamura; N. Sakurai; M. Kitajima; H. Takayama; S. Kanaya; K. Aoki; D. Shibata; K. Saito, *Phytochemistry* **2008**, *69*, 99–111.
293. S. Dhaubhadel; M. Gijzen; P. Moy; M. Farhangkhoee, *Plant Physiol.* **2007**, *143*, 326–338.
294. C. J. Tsai; S. A. Harding; T. J. Tschaplinski; R. L. Lindroth; Y. Yuan, *New Phytol.* **2006**, *172*, 47–62.
295. B. Hamberger; M. Ellis; M. Friedmann; C. D. A. Souza; B. Barbazuk; C. J. Douglas, *Can. J. Bot.* **2007**, *85*, 1182–1201.
296. N. Shimada; S. Sato; T. Akashi; Y. Nakamura; S. Tabata; S. I. Ayabe; T. Aoki, *DNA Res.* **2007**, *14*, 25–36.
297. D. R. Nelson; M. A. Schuler; S. M. Paquette; D. Werck-Reichhart; S. Bak, *Plant Physiol.* **2004**, *135*, 756–772.
298. L. Y. Li; H. Cheng; J. Y. Gai; D. Y. Yu, *Planta* **2007**, *226*, 109–123.
299. J. C. D'Auria, *Curr. Opin. Plant Biol.* **2006**, *9*, 331–340.
300. J. Ross; Y. Li; E. Lim; D. J. Bowles, *Genome Biol.* **2001**, *2*, REVIEWS3004.
301. C. M. M. Gachon; M. Langlois-Meurinne; P. Saindrenan, *Trends Plant Sci.* **2005**, *10*, 542–549.
302. L. A. Mueller; P. F. Zhang; S. Y. Rhee, *Plant Physiol.* **2003**, *132*, 453–460.
303. E. Urbanczyk-Wochniak; L. W. Sumner, *Bioinformatics* **2007**, *23*, 1418–1423.

304. J. A. Raven, *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* **2000**, *355*, 833–846.
305. H. A. Stafford, *Plant Physiol.* **1991**, *96*, 680–685.
306. B. Winkel-Shirley, *Plant Physiol.* **2001**, *126*, 485–493.
307. C. G. Jiang; C. K. Schommer; S. Y. Kim; D. Y. Suh, *Phytochemistry* **2006**, *67*, 2531–2540.
308. J. Yang; J. Huang; H. Gu; Y. Zhong; Z. Yang, *Mol. Biol. Evol.* **2002**, *19*, 1752–1759.
309. J. Yang; H. Gu; Z. Yang, *J. Mol. Evol.* **2004**, *58*, 54–63.
310. D. L. Des Marais; M. D. Rausher, *Nature* **2008**, *454*, 762–765.
311. E. Grotewold, *Trends Plant Sci.* **2005**, *10*, 57–62.
312. D. Treutter, *Environ. Chem. Lett.* **2006**, *4*, 147–157.
313. K. S. Gould; C. Lister, Flavonoids Functions in Plants. In *Flavonoids: Chemistry, Biochemistry and Applications*; Ø.M. Andersen, K. R. Markham, Ed.; CRC Press: Boca Raton FL, 2006pp 397–469.
314. M. S. Simmonds, *Phytochemistry* **2003**, *64*, 21–30.
315. L. P. Taylor; E. Grotewold, *Curr. Opin. Plant Biol.* **2005**, *8*, 317–323.
316. W. A. Peer; A. S. Murphy, *Trends Plant Sci.* **2007**, *12*, 556–563.

Biographical Sketches

Shin-ichi Ayabe, Ph.D. graduated from the Department of Chemistry, University of Tokyo, and after taking the masters degree of the graduate school there, he started his study of the constituents, mainly flavonoids of medicinal legume licorice, and their biosynthesis in cultured plant cells at the School of Pharmaceutical Sciences, Kitasato University (Tokyo) in 1973. He then moved to the Department of Applied Biological Sciences, Nihon University, at Fujisawa, Kanagawa, in 1989 where he is a professor of plant biochemistry since 1998. He was a postdoctoral research associate at St. Louis University, USA, from 1982 to 1983, and worked on morphinan alkaloid biosynthesis. His research has expanded from the investigation of the biosynthesis of unusual retrochalcone in the 1980s to biochemical and molecular biological approaches to (iso)flavonoid metabolism and led to the discovery and mechanistic studies of flavonoid-specific cytochrome P-450s in the 1990s–2000s. He has also been interested in the elicitation mechanism of phytoalexin production in plant cells, and more recently, in the functions of plant secondary metabolites in plant–microbe symbiosis.

Hiroshi Uchiyama is an associate professor of plant cell science at the Department of Applied Biological Sciences, College of Bioresource Sciences, Nihon University. He graduated from the Botanical Institute, Faculty of Science, Hiroshima University in 1989 with a Ph.D. His interest lies in the evolutionary biology of some plant groups, such as aquatic plants and carnivorous plants, through molecular cytogenetic and molecular phylogenetic approaches.



Toshio Aoki is an associate professor of plant biology in the Department of Applied Biological Sciences at Nihon University. He received his Ph.D. in 1991 from the University of Tokyo. He completed a his postdoctoral research fellowship funded by Japan Health Sciences Foundation from 1991 to 1993 at Tsukuba Medicinal Plant Research Station, National Institute of Health Sciences. From 1993 to 1996, he served as a research fellow at the Iwate Biotechnology Research Center. He has been an editorial board member of the *Journal of Plant Research* and an editor of *Plant Biotechnology*. His research interests include functional and structural analyses of genes involved in the biosynthesis of natural products, their biological significance in plant–microbe interactions and stress responses, and the identification of stress-tolerant genes.

Tomoyoshi Akashi graduated from the Department of Applied Biological Sciences, Nihon University in 1995, and obtained his Ph.D. degree from the same university in 1998. He worked as a JSPS Research Fellow at the same university (1997–1999). He was a postdoctoral research associate at the Department of Chemistry, Tokyo Institute of Technology (1999–2000) and worked on the biosynthesis of macrolide antibiotics in actinomycetes. He then moved back to the Department of Applied Biological Sciences, Nihon University in 2000 as an assistant professor. He is working on flavonoid biosynthesis, and his research has led to the discovery of P-450s in (iso)flavonoid metabolism and the complete characterization of the pathway and genes of isoflavone biosynthesis. His research interests include biosynthesis, molecular evolution, and metabolic engineering of plant natural products.

1.25 Alkaloids

Sarah E. O'Connor, Massachusetts Institute of Technology, Cambridge, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

1.25.1	What Is an Alkaloid?	977
1.25.2	Classes of Alkaloids	977
1.25.3	Function and Diversity of Alkaloids	978
1.25.4	Strategies for Elucidating Alkaloid Biosynthesis	982
1.25.5	Benzylisoquinoline Alkaloid Biosynthesis	987
1.25.6	Monoterpene Indole Alkaloid Biosynthesis	992
1.25.7	Tropane Alkaloid Biosynthesis	999
1.25.8	Purine Alkaloid Biosynthesis	1001
1.25.9	Conclusions and Outlook	1001
References		1003

1.25.1 What Is an Alkaloid?

Alkaloids encompass an enormous class of approximately 12 000 natural products.[1] The principal requirement for classification as an alkaloid is the presence of a basic nitrogen atom at any position in the molecule, which does not include nitrogen in an amide or peptide bond. As implied by this exceptionally broad definition, the alkaloids form a group of structurally diverse and biogenically unrelated molecules. Most classes of natural products are composed of similar chemical structures, in which the same starting materials are assembled in related biosynthetic pathways. For example, all polyketides derive from acetate and propionate building blocks that undergo a series of Claisen condensation reactions.[2] However, no biochemical paradigm is centrally applied throughout alkaloid biosynthesis. Instead, the biosynthetic pathways of alkaloids are as diverse as the chemical structures found within this class of natural products.

Historical reasons account for this broad chemical definition of an alkaloid.[3] Prior to the nineteenth century, all compounds purified from plants – such as tartaric acid, oxalic acid, and tannins – exhibited acidic properties. However, alkaline material called potash extracted from burnt wood was later found to contain basic compounds of pharmacological interest. Meissner suggested in 1819 that these compounds be referred to as alkaloids, meaning a plant-derived substance that displays alkaline properties.[3]

In 1806, Friedrich Wilhelm Serturner, for the first time, isolated a pure compound that exhibited the same pharmacological sleep-inducing properties of the crude opium extract from which this compound was isolated. The discovery of this alkaloid compound, subsequently named morphine, played a key role in the development of what was to become the modern pharmaceutical industry, and the purification of many other pharmacologically important alkaloids such as strychnine, quinine, and caffeine rapidly followed suit (**Figure 1**).[3] Notably, accurate structural elucidation of the alkaloids was much more difficult. Many alkaloid structures remained unknown until well into the twentieth century when X-ray spectroscopy became widely available, and after organic chemistry had advanced to the point where these molecules could be produced synthetically.

Although plants were the first known source of alkaloid compounds, it is now known that fungi, bacteria, insects, and animals also produce a wide array of alkaloids. Therefore, the definition of alkaloids has been expanded to state "alkaloids are nitrogen-containing organic substances of natural origin with a greater or lesser degree of basic character".[3]

1.25.2 Classes of Alkaloids

Alkaloids are most commonly constructed from amino acid starting materials, although some purine-derived alkaloids are also known. The structural class of the alkaloid is typically defined by the substrate starting material. For example, tyrosine is used for the production of tetrahydroisoquinoline alkaloids (**Figure 2(a)**).

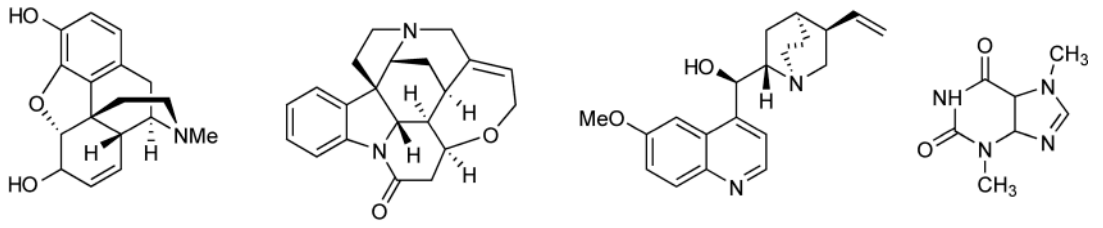


Figure 1 Pharmacologically important alkaloids discovered early in the history of alkaloids. Morphine, strychnine, quinine, and caffeine are shown.

Tryptophan is the starting material for indole-containing alkaloids, such as alkaloids containing a β-carboline moiety (**Figure 2(b)**). Additionally, the indole of tryptophan can be modified to form the quinoline alkaloids (**Figure 2(b)**). Ornithine, a nonproteogenic amino acid derived from glutamate or arginine, is used to produce the pyrrolizidine- and tropane-type alkaloids (**Figure 2(c)**). Nicotinic acid can be used in combination with other amino acids to yield the nicotinic alkaloids (**Figure 2(d)**). Lysine, which contains one extra methylene group compared to ornithine, produces the structurally analogous piperidine, quinolizidine, and indolizidine alkaloids (**Figure 2(e)**). Known alkaloids derived from other amino acids are more rare. In addition to amino acid building blocks, a number of other nitrogen-containing starting materials can serve as alkaloid precursors. Anthranilic acid, a precursor to tryptophan, is used to produce quinazoline-, quinoline-, and acridine-type alkaloids (**Figure 2(f)**). A number of purine-derived alkaloids have also been isolated (**Figure 2(g)**). Finally, some alkaloids acquire the requisite basic nitrogen through transamination of an existing polyacetate or terpenoid framework. These compounds are referred to as pseudoalkaloids.

1.25.3 Function and Diversity of Alkaloids

The function of alkaloids is still not entirely clear. Although the pharmacological uses of many alkaloids are well defined, the specific roles that these compounds play in the producing organism are not well elucidated in most cases. At one point it was speculated that alkaloids were simply waste products derived from the degradation of primary metabolites. However, although the exact roles of many alkaloids remain poorly understood, these compounds are now believed to play an important ecological role, enabling the producing organism to defend itself and interact with its environment. In fact, although natural products are often termed secondary metabolites, the vital role that many natural products play in signaling and development renders the term 'secondary' a misnomer.

Most obviously, many alkaloids are toxic, so biosynthesis of these compounds provides a general defensive mechanism for the producing organism. For example, caffeine has been shown to act as an insecticide. Caffeine may act as a natural insecticide in plants. When the three *N*-methyltransferase genes involved in caffeine biosynthesis were overexpressed in tobacco, the resulting increase in caffeine production improved the tolerance of the plants to certain pests.[4] Occasionally, alkaloids that are present in animals are acquired by predation on a plant alkaloid producer. In one example, a moth species *Tyria jacobaeae* feeds on the plant *Senecio jacobaea*, which produces a number of pyrrolizidine alkaloids. The moth detoxifies the molecule by oxidizing the nitrogen to an N-oxide using a flavin-dependent monooxygenase (**Figure 3**). If the moth is ingested by a predator, the N-oxide is reduced in the gut of the predator where it is degraded to highly toxic pyrrole moieties. The regulation of natural products also suggests a defensive function for these compounds. Many natural products, including alkaloids, are upregulated under stressful conditions. For example, sterilized preparations of fungal cell wall extracts, called fungal elicitors, can be added to plant cell culture to increase the levels of alkaloid production in many plant species.

Nevertheless, it remains unclear why such a diversity of vastly different defensive, toxic molecules is required. This diversity is particularly apparent in the tetrahydroisoquinoline and monoterpene indole classes of alkaloids, where thousands of alkaloid products are generated from a single, central biosynthetic

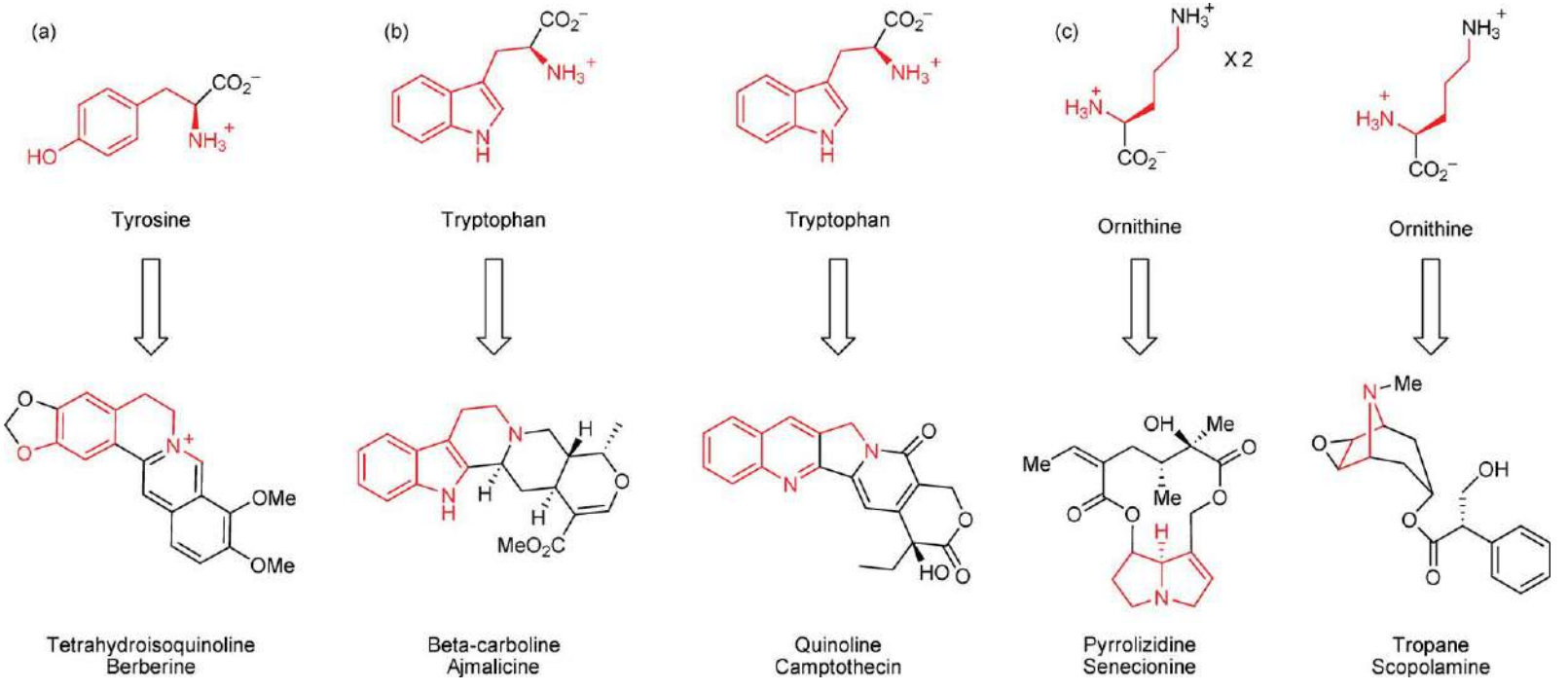


Figure 2 (Continued)

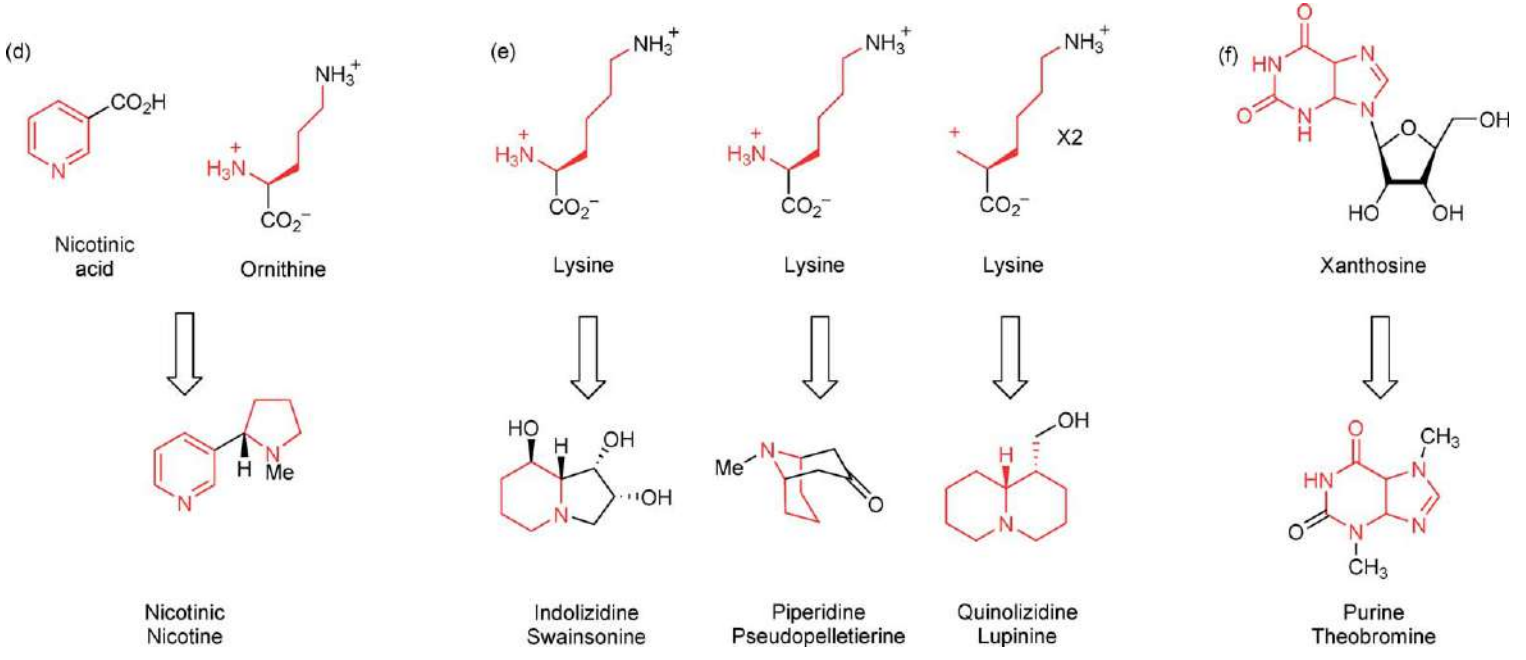


Figure 2 (Continued)

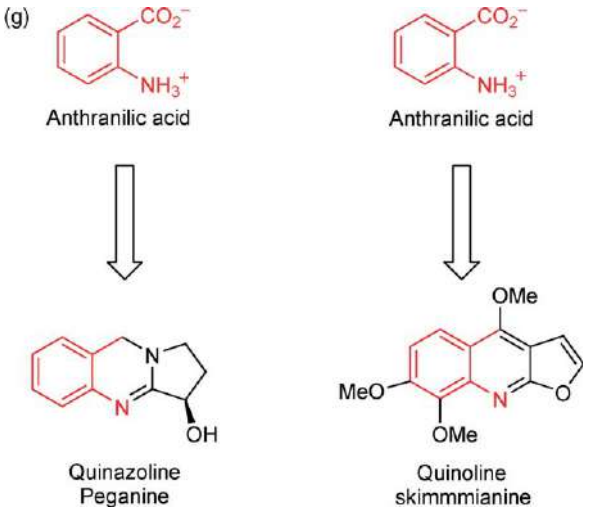
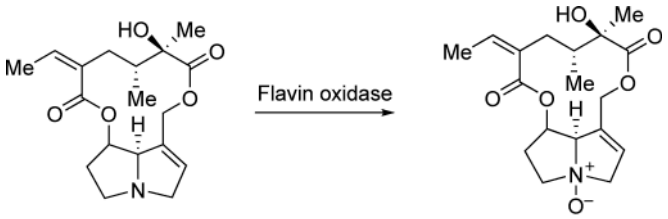


Figure 2 Representative members of the major structural classes of alkaloids. (a) Tyrosine-derived tetrahydroisoquinoline alkaloids. (b) Tryptophan-derived monoterpene indole alkaloids containing a β-carboline or quinoline moiety. (c) Ornithine-derived pyrrolizidine- and tropane-type alkaloids. (d) Nicotinic acid-derived alkaloid. (e) Lysine-derived piperdine, quinolizidine, and indolizidine alkaloids. (f) Anthranilic acid-derived quinazoline-, quinoline-, and acridine-type alkaloids. (g) Purine-derived alkaloid.



Sennecione Sennecione oxide

Figure 3 Oxidation of pyrrolizidine alkaloid for detoxification.

intermediate. A single medicinal plant known to produce these alkaloids typically generate 100 or more members of the alkaloid family. A recent commentary highlights that 'primary metabolic pathways are target-oriented', in other words, designed to produce a single, highly optimized molecule.[5] Many secondary or natural product pathways are, in contrast, 'diversity oriented'. Firn and Jones[6] hypothesized that since potent biological activity is a rare property for any molecule to have, "an organism needs the ability to make multiple molecules in order to hit upon the rare potent ones". In many species, natural products serve as an immune system of sorts. If the producing organism continuously evolves its capacity to generate new structures, the organism may be advantageously positioned for survival in its environment.

1.25.4 Strategies for Elucidating Alkaloid Biosynthesis

First and foremost, a mechanistic elucidation begins with the accurate structural characterization of the alkaloid product. Natural product structure elucidation posed a formidable challenge in the early days of alkaloid isolation. A variety of strategies have made this process much easier, namely, X-ray analysis and high-field 2D nuclear magnetic resonance (NMR) techniques.[7] Additionally, total synthesis of a reported complex alkaloid structure is a critical strategy for confirming the structural features of the natural product. A mechanistic elucidation of a natural product biosynthetic pathway typically begins by establishing which precursors are involved. Isotopically labeled (potential) precursors are fed to the producing organism, and the desired alkaloid product is then isolated. If the alkaloid contains the isotopic label, then it is clear that the precursor was used in the biosynthesis. Older experiments relied on radiolabeled precursors, since detection of radioactivity in the final product is sensitive and straightforward. However, the routine use of powerful, high-resolution mass spectrometry techniques has made the detection of safe, inexpensive stable isotopes such as hydrogen-2 (deuterium), carbon-13, and nitrogen-15 much more practical. After establishing the identity of the correct starting materials, strategic positioning of isotopic labels can be used to probe the mechanism of the transformations and structural rearrangements that occur along the biosynthetic route. These experiments, when placed in the context of previously known biochemical transformations, enable a series of logical enzymatic reactions to be proposed.

Identification of the enzymes, and the corresponding genes that encode them, constitutes the next level of pathway elucidation. How these genes are identified, or cloned, is dependent on the identity of the producing organism. Historically, alkaloid research has focused on plant-derived compounds. Plants are immobile and interact with their environment largely via the release of complex small molecules, so it is not surprising that this evolutionary pressure has resulted in the production of an extremely diverse array of natural products by plants. However, elucidating the genes of a plant pathway poses significant challenges. In contrast to microbes, the genes of plant pathways – with a few exceptions[8,9] – are not clustered on the genome, so each gene of a plant pathway must be discovered individually. Additionally, the genome sizes of medicinal plants are much larger (>1000 Mbp) than the typical natural product-producing bacteria (~8 Mbp), which makes finding and screening putative biosynthetic genes difficult. Although spectacular successes have been achieved in elucidating plant pathways, the challenges of plant biology have hindered the study of plant secondary metabolism and most plant-derived pathways remain incompletely elucidated at the genetic level. However, complex, higher plants frequently produce compounds that are not found in any known bacteria or fungi.

The majority of plant biosynthetic enzymes have been identified using a classical approach in which the enzyme of interest is purified from a crude plant lysate by standard protein chromatography.[10] Once a homogenous preparation of enzyme is prepared, partial protein sequence is obtained from the purified protein, which is then used to identify the gene encoding the desired enzyme. In another approach, if a guess can be made as to the type of enzyme involved in the desired transformation, a homology-based cloning strategy can be used.[11] Enzymes within a given class contain highly conserved regions in the protein sequence. Oligonucleotide primers complementary to these consensus sequences can be used to amplify gene candidates that can be screened for function. In a third approach called subtractive hybridization, cDNA from two types of tissue can be 'subtracted' from one another.[12] The genes in tissue that produces natural products at high levels can be compared with gene expression levels in tissue that produce low levels of alkaloid in question. The genes unique to each tissue type are readily obtained, and presumably, at least some of the genes that are unique to the alkaloid-producing tissue are involved in alkaloid biosynthesis. Finally, suppression (by RNAi)[13] or activation (by T-DNA tagging)[14] of large numbers of genes, followed by screening for changes in the levels of alkaloid production, can also be used to identify biosynthetic enzymes, provided that a fast screen or selection is available to interpret the phenotype of the transformed lines. Many plant alkaloid enzymes have been elucidated, but the plant alkaloids that have been studied in the most detail at the genetic level are the isoquinoline alkaloids,[1,15] the terpenoid indole alkaloids,[16] the tropane alkaloids,[17,18] and the purine alkaloids[19] (**Figure 4**).

Bacteria, whether terrestial or marine in habitat, typically export an arsenal of natural products involved in defense and signaling. Bacteria have much smaller genomes than plants, and the genes of bacterial natural product pathways tend to be organized in clusters, which makes identification of an entire metabolic pathway much more straightforward than in plants. Historically, bacteria have not been a rich source of alkaloids, although several complex alkaloid pathways have been recently discovered.[20,21] Undoubtedly, many more alkaloids will be discovered as the technical and financial barriers to sequence whole bacterial genomes continue to decrease, and more bacterial species are readily available for genetic analysis.

The study of microbe-derived natural products underwent a revolution in the 1980s as improved sequencing technologies allowed the rapid discovery of the genes that encode natural product biosynthetic pathways.[22] Since the genes of bacterial-derived pathways are relatively easy to identify, elucidation of these pathways is less dependent on isotopic labeling studies. Bioinformatic analysis of the genes in the cluster can provide clues as to the types of enzymatic transformations that take place. Although traditionally, the desired gene cluster is targeted and isolated in a cosmid vector for sequence analysis, it can now be cost-effective to simply sequence the entire genome of the producing strain and use bioinformatic analysis to search for candidate gene clusters in the genome sequence. After identification of the cluster, heterologous expression of candidate enzymes followed by *in vitro* biochemical assay of various combinations of these enzymes and substrates provide important information. Finally, if the producing organism can be genetically manipulated, targeted gene deletions can be made to definitively validate that the cluster in question is responsible for the production of the compound. Most notably, the prodiginines[23] and the indolocarbazoles[24–26] such as rebeccamycin, staurosporin, and violecin are prokaryotic alkaloids that have been the subject of several recent investigations. Benzodiazapines[27] and saframycins[28–30] have also been the subject of recent study (**Figure 5**).

Fungi, like bacteria, also produce a wealth of natural products. Eukaryotic fungal organisms generally have larger genomes and more complex life cycles, so, while fungi are still much simpler than higher plants they are considerably more complex than bacteria. It appears that the genes of many metabolic pathways are also clustered in fungal genomes, thereby greatly simplifying the study of fungal biosynthetic pathways. Nevertheless, fungal clusters are typically larger than prokaryotic clusters, and the genes often contain short introns. Several fungal alkaloids have also been partially elucidated at the genetic level. The ergot alkaloids[31] and the indole diterpenes[32] are two major classes of fungal alkaloids that have been studied extensively (**Figure 6**).

In addition to plants, bacteria, and fungi, many other organisms produce extraordinary alkaloid structures. Sponges as well as other multicelled marine organisms, insects, amphibians, and even mammals all produce complex alkaloid natural products.[33,34] However, in general, the biosynthetic pathways from these complex organisms are much less well characterized than the pathways from microbial and plant kingdoms. In many of these cases, the producing organisms are not viable in a laboratory environment, which complicates the elucidation of the biosynthetic mechanism. Additionally, genetic manipulation of many of these organisms is not yet developed.

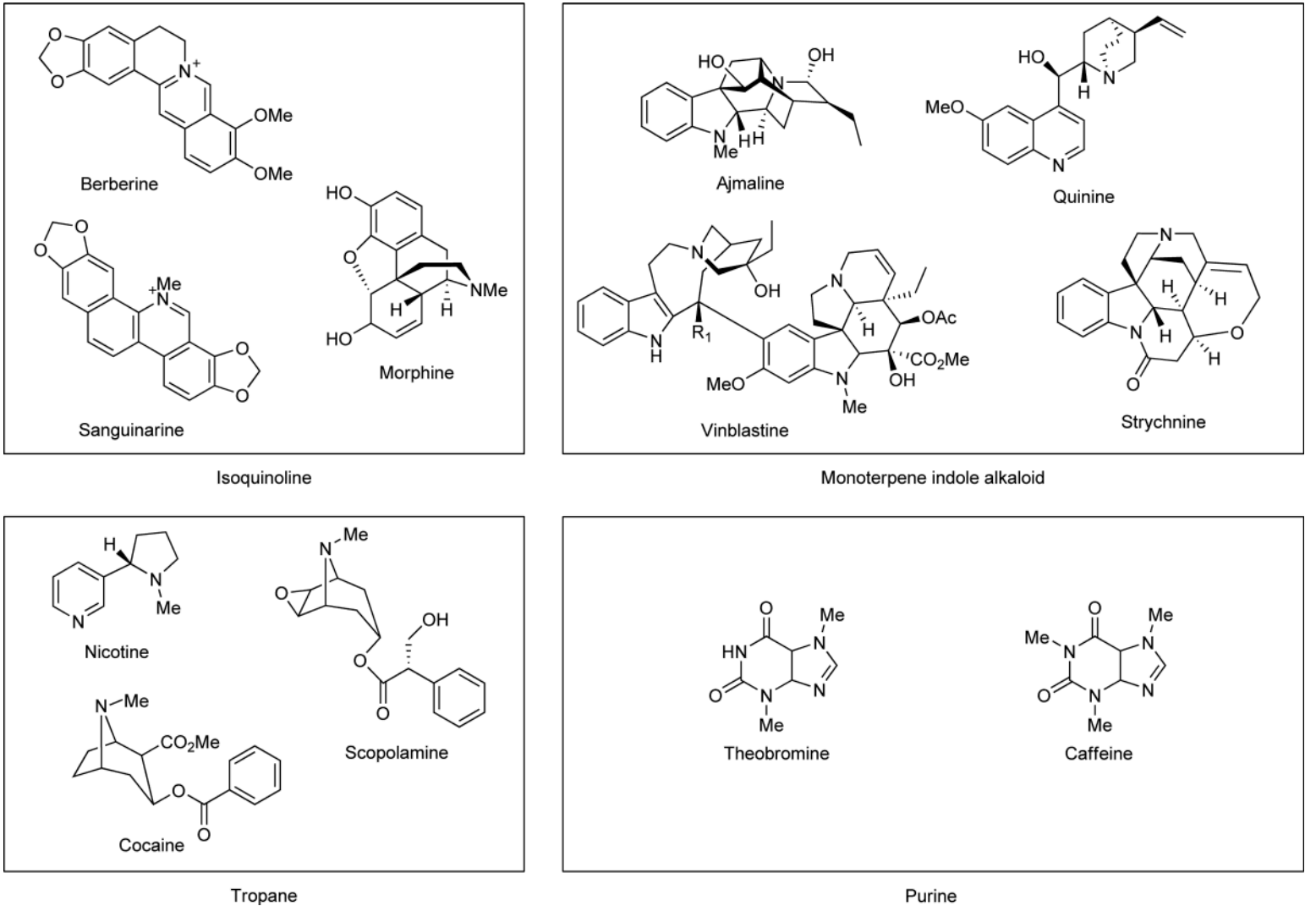


Figure 4 Representative plant-derived alkaloids from the tetrahydroisoquinoline, monoterpene indole, tropane, and purine classes.

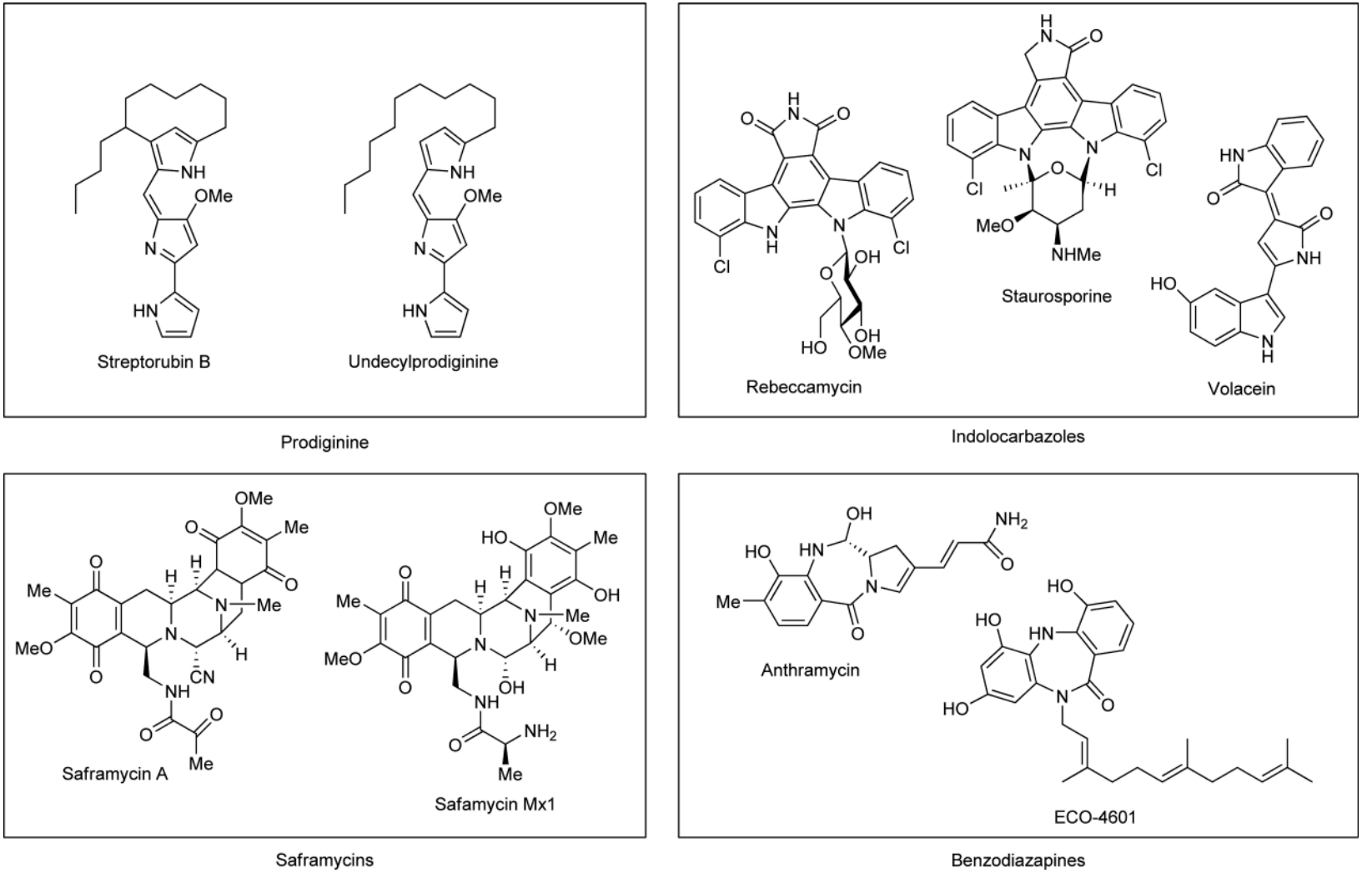


Figure 5 Representative prokaryote-derived alkaloids from the prodiginine, indolocarbazole, saframicin, and benzodiazapine classes.

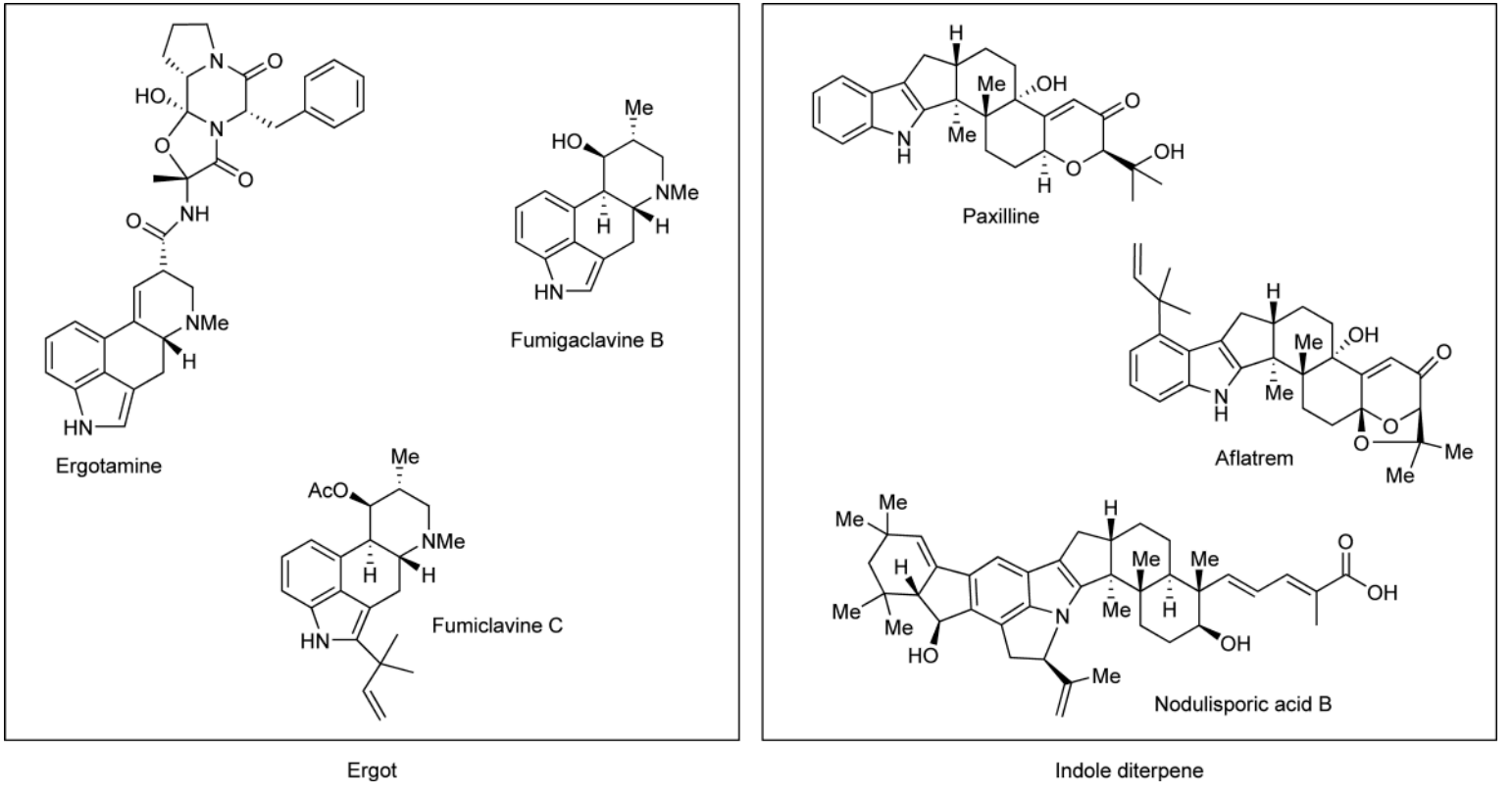


Figure 6 Representative fungal-derived alkaloids of the ergot and indole diterpene types.

A comprehensive discussion of all known alkaloid biosynthetic mechanisms extends well beyond the scope of this chapter. Here, we focus on the small number of plant-derived alkaloids for which genetic information regarding the biosynthetic pathway has been elucidated. Even within this limited subset of alkaloid structures, the structural diversity that is observed among the alkaloids becomes apparent, and the chemistry involved in these relatively few pathways is wide ranging.

1.25.5 Benzylisoquinoline Alkaloid Biosynthesis

The isoquinoline alkaloids include, most famously, the opiates morphine and codeine as well as the antibiotic berberine (**Figure 7**). Morphine and codeine are two of the most important analgesics used in medicine, and plants remain the main commercial source of the alkaloids.[35] Notably, development of plant cell cultures of *Eschscholzia californica*, *Papaver somniferum*, and *Coptis japonica* has aided in the isolation and cloning of many enzymes involved in the biosynthesis of isoquinoline alkaloids.[15]

Isoquinoline biosynthesis begins with the substrates dopamine and *p*-hydroxyphenylacetaldehyde to yield the central intermediate of this biosynthetic pathway, (*S*)-reticuline (**Figure 8**). Tyrosine is hydroxylated and decarboxylated to yield dopamine. Enzymes that catalyze the hydroxylation and decarboxylation steps in either order exist in the plant, and the predominant pathway for the formation of dopamine from tyrosine is not clear. The second substrate, *p*-hydroxyphenylacetaldehyde, is synthesized by transamination and decarboxylation of tyrosine.[36] Dopamine and *p*-hydroxyphenylacetaldehyde are coupled by the enzyme norcoclaurine synthase to form (*S*)-norcoclaurine. Two norcoclaurine synthases with completely unrelated sequences have been cloned (*Thalictrum flavum* and *C. japonica*) and heterologously expressed in *Escherichia coli*.[37,38] One shows homology to iron-dependent dioxygenases, whereas the other is homologous to a pathogenesis-related protein. Recent structural analysis of one of these enzymes has shed light onto the mechanism of this enzymatic transformation. Undoubtedly, future experiments will explain how two such widely divergent sequences can catalyze the same reaction.

One of the hydroxyl groups of (*S*)-norcoclaurine is methylated by an *S*-adenosyl methionine (SAM)-dependent *O*-methyltransferase to yield (*S*)-coclaurine. This enzyme has been cloned, and the heterologously expressed enzyme exhibited the expected activity.[39,40] The next biosynthetic intermediate is N-methylated to yield *N*-methylcoclaurine, an enzyme that has also been cloned.[41,42] *N*-Methylcoclaurine is then hydroxylated by a P-450-dependent enzyme (CYP80B), *N*-methylcoclaurine 3′-hydroxylase, that has been cloned.[43] One of the hydroxyl groups is methylated by the enzyme 3′-hydroxy-*N*-methylcoclaurine 4′-*O*-methyltransferase (4′-OMT) to yield (*S*)-reticuline, the common biosynthetic intermediate for this pathway (**Figure 8**).[44–46] The biosynthetic pathway then diverges to yield the different structural classes of isoquinoline alkaloids.

In one major pathway, (*S*)-reticuline is converted to (*S*)-scoulerine by the action of a well-characterized flavin-dependent enzyme, berberine bridge enzyme (**Figure 9**). This enzyme has been cloned from several plant species,[47–49] and the mechanism of this enzyme has been studied extensively.[50,51] Notably, a structural analysis of this enzyme has been recently reported.[52] (*S*)-Scolerine is then methylated by scoulerine 9-*O*-methyltransferase to yield (*S*)-tetrahydrocolumbamine. Heterologous expression yielded an enzyme that had the expected substrate specificity.[53] The substrate-specific cytochrome P-450 oxidase canadine synthase that generates the methylene dioxy bridge of (*S*)-canadine has been cloned.[54] The final step of berberine

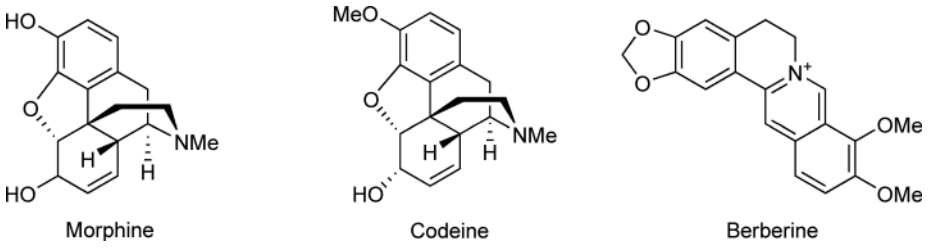


Figure 7 Representative tetrahydroisoquinoline alkaloids morphine, codeine, and berberine.

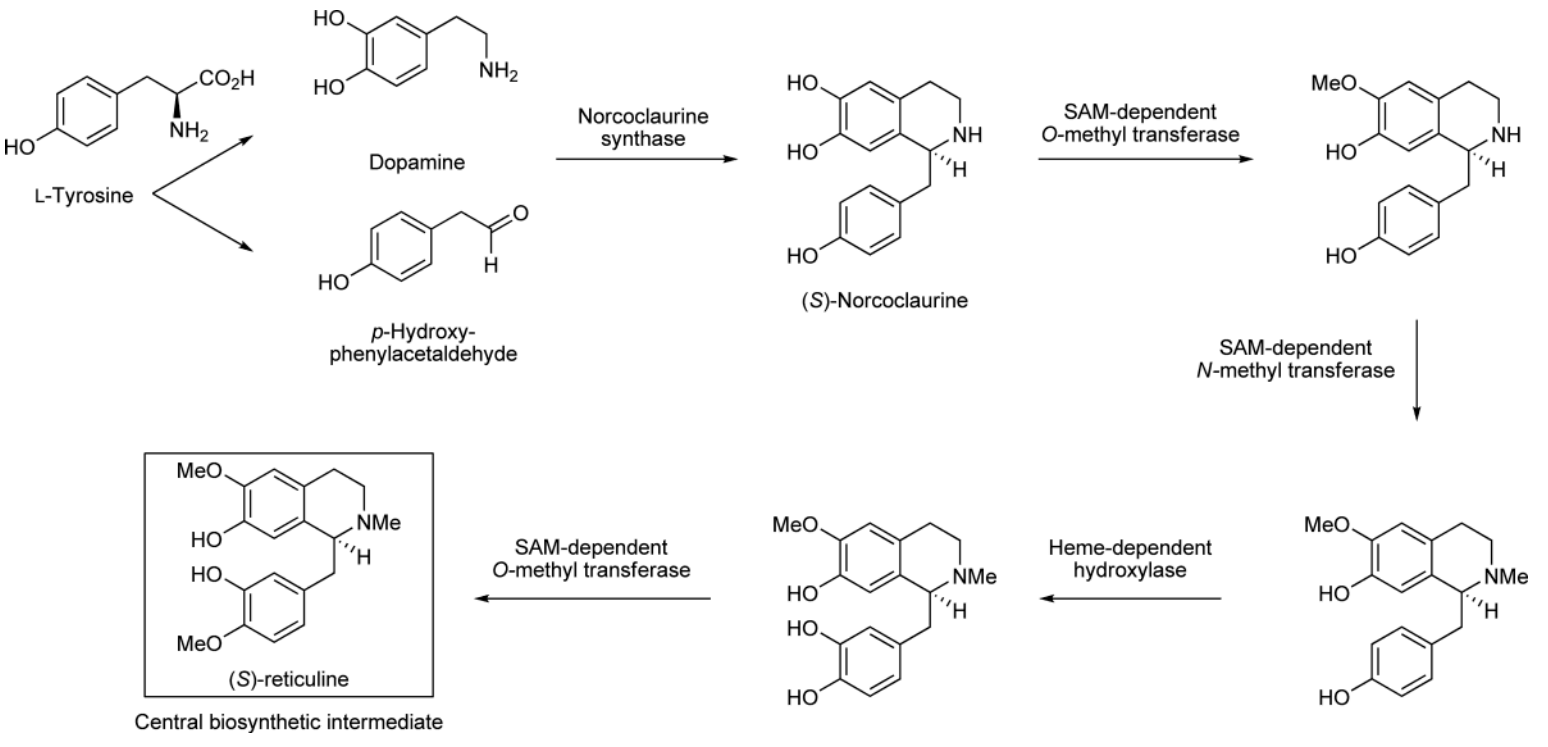


Figure 8 Biosynthesis of (*S*)-reticuline, the central intermediate of tetrahydroisoquinoline alkaloid biosynthesis.

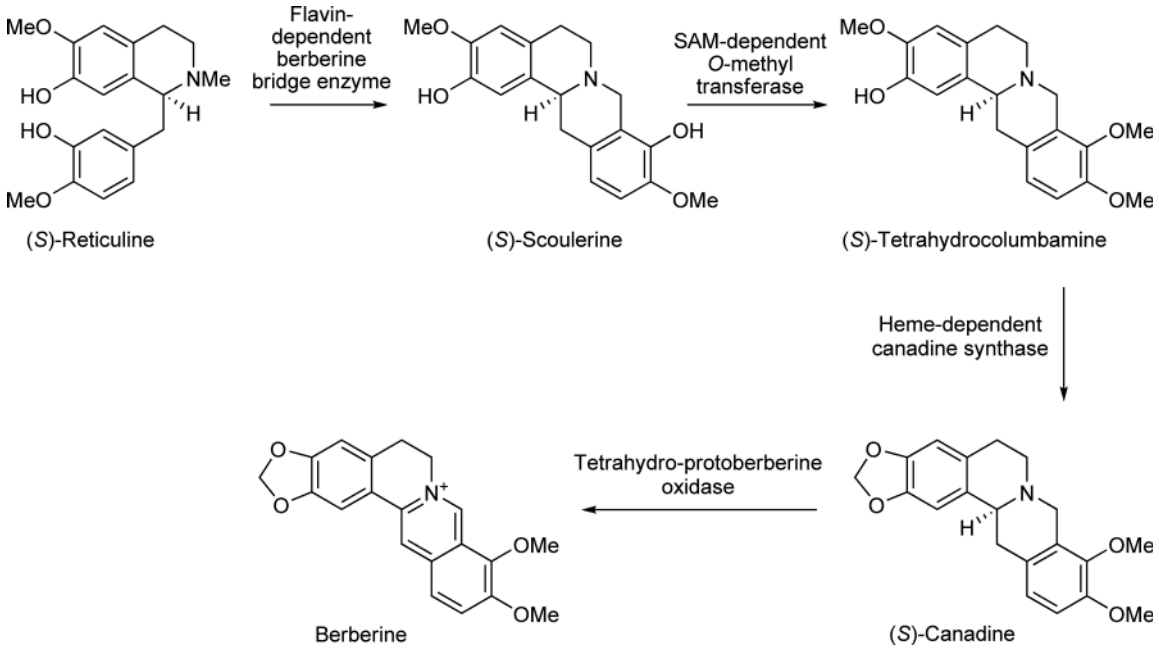


Figure 9 Biosynthesis of berberine.

biosynthesis is catalyzed by a substrate-specific oxidase, tetrahydroprotoberberine oxidase, the sequence of which has not been identified yet.[55] Berberine can be overproduced in *C. japonica* cell suspension cultures with reported productivity of berberine reaching 7 g l^{-1}.[56,57] This overproduction is one of the first demonstrations of production of a benzylisoquinoline alkaloid in cell culture at levels necessary for economic production. Additionally, this cell line has enabled the identification of many of the biosynthetic enzymes.

A second major pathway branch is the biosynthesis of the highly oxidized benzo(*c*)phenanthidine alkaloid sanguinarine, which is produced in a variety of plants and competes with morphine production in opium poppy. The pathway to sanguinarine has been elucidated at the enzymatic level (**Figure 10**).[58] Sanguinarine biosynthesis starts from (*S*)-scoulerine, as in berberine biosynthesis. Methylenedioxy bridge formation is then catalyzed by the P-450 cheilanthifoline synthase to yield cheilanthifoline.[59] A second P-450 enzyme, stylopine synthase, catalyzes the formation of the second methyenedioxy bridge of stylopine. Stylopine synthase from *E. californica* has been cloned recently.[60] Stylopine then is N-methylated by (*S*)-tetrahydroprotoberberine *cis-N*-methyltransferase to yield (*S*)-*cis-N*-methylstylopine, an enzyme that has been cloned recently from opium poppy.[61] A third P-450 enzyme, (*S*)-*cis-N*-methylstylopine hydroxylase, then forms protopine. Protopine is hydroxylated by a fourth P-450 enzyme, protopine 6-hydroxylase, to yield an intermediate that rearranges to dihydrosanguinarine.[62] The copper-dependent oxidase dihydrobenzophenanthridine oxidase, which has been purified,[63,64] then catalyzes the formation of sanguinarine from dihydrosanguinarine.

A third major branch leading to morphine biosynthesis has been investigated in *P. somniferum* cells and tissue. Notably, in morphine biosynthesis, (*S*)-reticuline is converted to (*R*)-reticuline, epimerizing the stereocenter generated by norcoclaurine synthase at the start of the pathway (**Figure 11**). This clearly exemplifies that the shortest enzymatic route to the final product is not always utilized by nature in metabolic pathways. Instead, nature will often use an existing biosynthetic intermediate for the biosynthesis. (*S*)-Reticuline is converted to (*R*)-reticuline through a 1,2-dehydroreticuline intermediate. Dehydroreticuline synthase catalyzes the oxidation of (*S*)-reticuline to 1,2-dehydroreticulinium ion.[65] This enzyme has not been cloned but has been purified partially and shown to be membrane associated. This intermediate then is reduced by dehydroreticuline reductase, an NADPH-dependent enzyme that stereoselectively transfers a hydride to dehydroreticulinium ion to yield (*R*)-reticuline. This enzyme has not been cloned yet but has been purified to homogeneity.[66] The key carbon–carbon bond of the morphinan alkaloids is formed by the cytochrome P-450 enzyme salutaridine

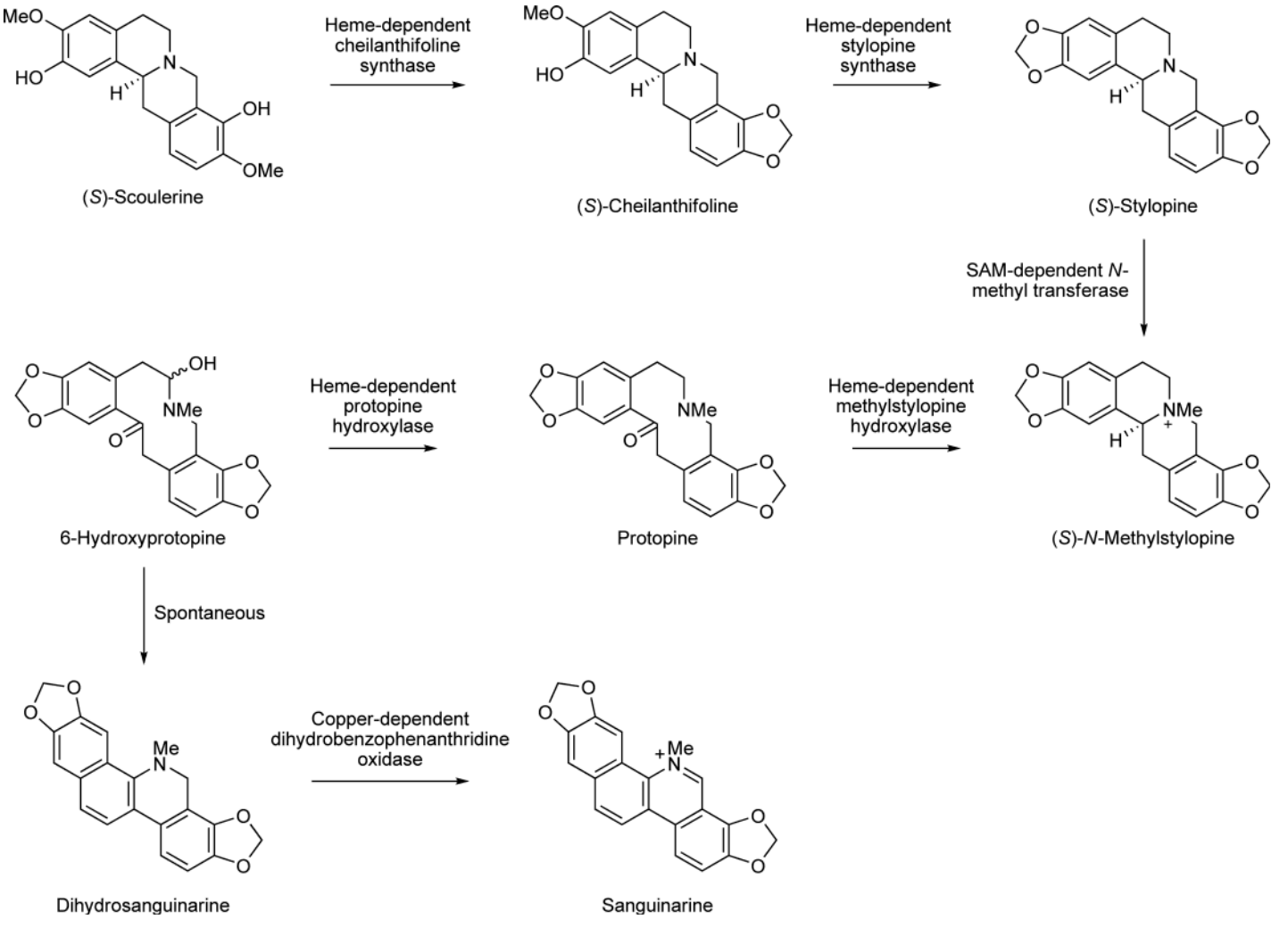


Figure 10 Biosynthesis of sanguinarine.

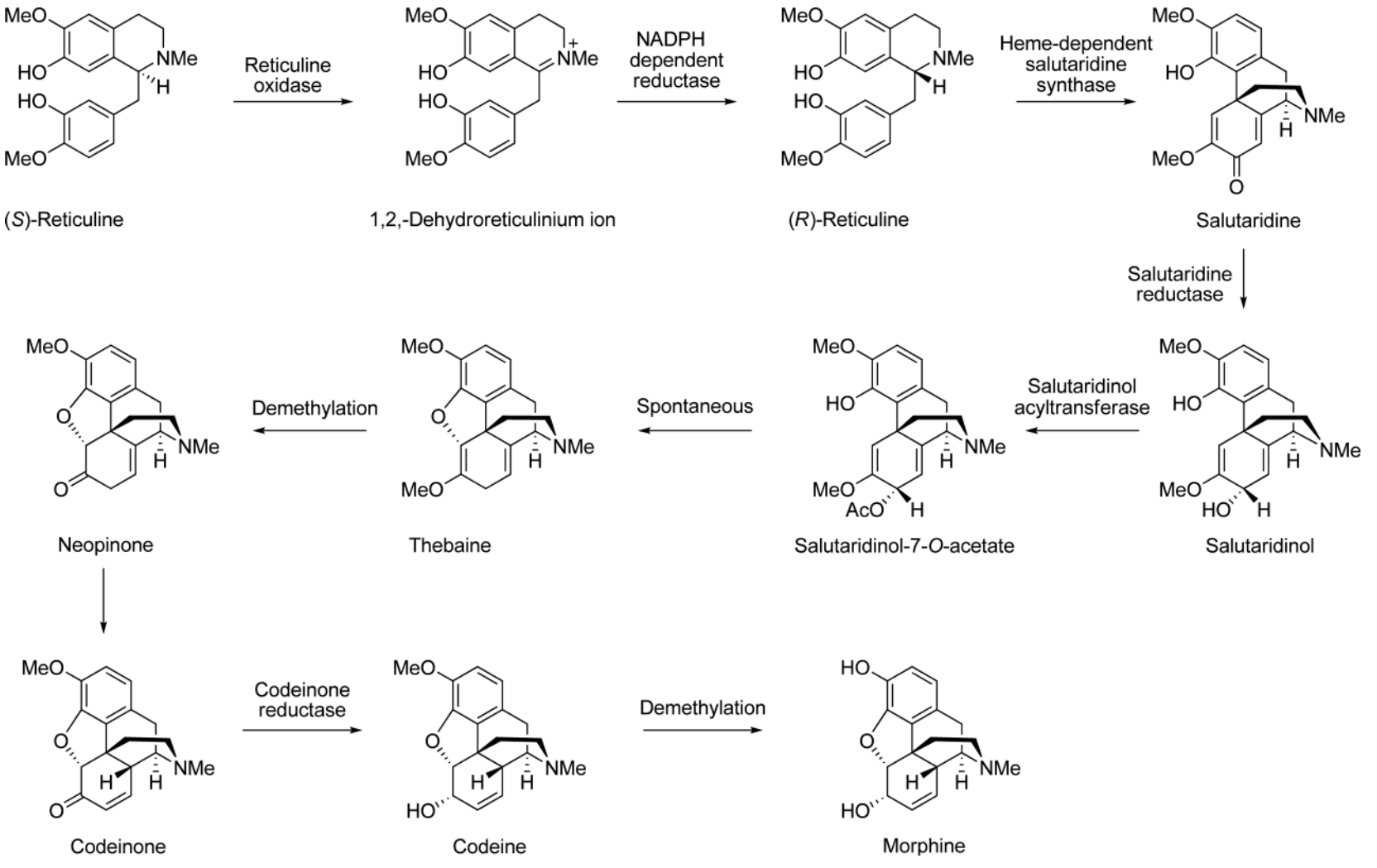


Figure 11 Biosynthesis of morphine and codeine.

synthase. Activity for this enzyme has been detected in microsomal preparations, and the sequence has been recently deposited.[67] The keto moiety of the resulting product, salutaridine, then is stereoselectively reduced by the NADPH-dependent salutaridine reductase to form salutardinol. Recent transcript analysis profile of *P. sominiferum* has resulted in the identification of this gene.[68] Salutaridinol acetyltransferase, also cloned, transfers an acyl group from acetyl-CoA to the newly formed hydroxyl group, which leads to the formation of salutaridinol-7-*O*-acetate.[69] The molecule can then undergo a spontaneous reaction in which the acetate acts as a leaving group. The resulting product, thebaine, then is demethylated by an as yet uncharacterized enzyme to yield neopinione, which exists in equilibrium with its tautomer codeinone. The NADPH-dependent codeinone reductase catalyzes the reduction of codeinone to codeine and has been cloned.[70,71] Finally, codeine is demethylated by an uncharacterized enzyme to yield morphine.

A series of elegant labeling experiments with human cell culture has indicated that mammalian tissue is capable of synthesizing morphine from dopamine precursors.[72] Presumably, morphine may play a role as an endogenous pain reliever in humans and other mammals. The mammalian enzymes involved in morphine biosynthesis have not been extensively investigated.

The localization of morphine biosynthesis has been investigated at the cellular level in intact poppy plants by using *in situ* RNA hybridization and immunofluorescence microscopy. The localization of 4′-*O*-methyltransferase (reticuline biosynthesis), berberine bridge enzyme (saguinarine biosynthesis), salutaridinol acetyltransferase (morphine biosynthesis), and codeinone reductase (morphine biosynthesis) has been probed. 4′-*O*-Methyltransferase and salutaridinol acetyltransferase are localized to parenchyma cells, whereas codeinone reductase is localized to laticifer cells in sections of capsule (fruit) and stem from poppy plants. Berberine bridge enzyme is found in parenchyma cells in roots. Therefore, this study suggests that two cell types are involved in isoquinoline biosynthesis in poppy and that intercellular transport is required for isoquinoline alkaloid biosynthesis.[73] Another study, however, implicates a single cell type (sieve elements and their companion cells) in isoquinoline alkaloid biosynthesis.[74,75] Therefore, it is not clear whether transport of pathway intermediates is required for alkaloid biosynthesis or whether the entire pathway can be performed in one cell type. Undoubtedly, future studies will provide more insight into the trafficking involved in plant secondary metabolism.

The extensive knowledge of the genes of isoquinoline biosynthesis has enabled a variety of metabolic engineering work to be done. In attempts to accumulate thebaine and decrease production of morphine (a precursor to heroin), codeinone reductase in opium poppy plant was downregulated by using RNAi.[35,76] Silencing of codeinone reductase results in the accumulation of (*S*)-reticuline but not the substrate codeinone or other compounds on the pathway from (*S*)-reticuline to codeine (**Figure 11**). The cytochrome P-450 responsible for the oxidation of (*S*)-*N*-methylcoclaurine to (*S*)-3′-hydroxy-*N*-methylcocluarine has been overexpressed in opium poppy plants, and morphinan alkaloid production in the latex is increased subsequently to 4.5 times the level in wild-type plants.[77] Additionally, suppression of this enzyme resulted in a decrease in morphinan alkaloids to 16% of the wild-type level. Notably, analysis of a variety of biosynthetic gene transcript levels in these experiments supports the hypothesis that this P-450 enzyme plays a regulatory role in the biosynthesis of benzylisoquinoline alkaloids. Collectively, these studies highlight that the complex metabolic networks found in plants are not redirected easily or predictably in all cases. Notably, portions of this pathway have been reconstituted in *Saccharomyces cerevisiae*, which is an organism that is much more amenable to rational metabolic engineering efforts.[78,79]

1.25.6 Monoterpene Indole Alkaloid Biosynthesis

The terpene indole alkaloids are a diverse class of natural products, comprising over 2000 members. These complex natural products possess a range of chemical structures and a wealth of biological activities (**Figure 12**).[80,81] The biosynthetic pathways of some classes of terpene indole alkaloids are well understood, and in some branches, many of the enzymes that are responsible for biosynthesis have been actually cloned and mechanistically studied *in vitro*. In other cases, the biosynthetic pathway is only proposed based on the results of feeding studies with isotopically labeled substrates and from the structures of isolated biosynthetic intermediates. Although many biosynthetic genes from this pathway remain unidentified, recent studies have correlated

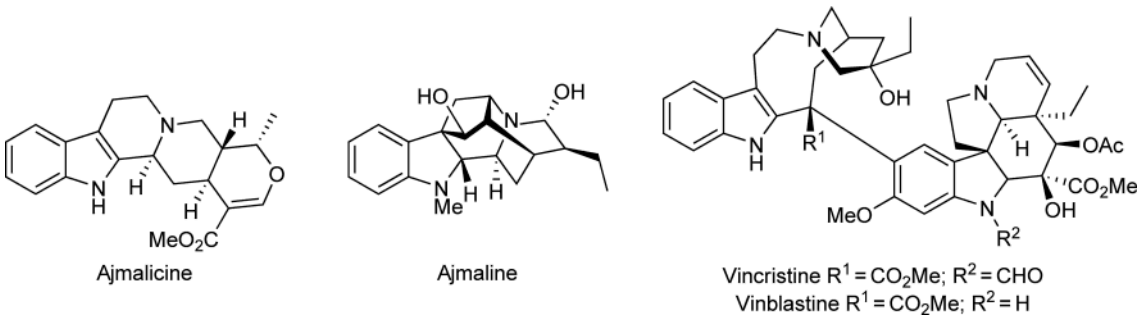


Figure 12 Representative monoterpene indole alkaloids ajmalicine, ajmaline, vinblastine, and vincristine.

terpenoid indole alkaloid production with the transcript profiles of *Catharanthus roseus* cell cultures.[82] Although the genome sequences of none of these alkaloid-producing plants is available, a number of expressed sequence tag (EST) libraries for *C. roseus* have been reported.[83,84]

All terpenoid indole alkaloids are derived from tryptophan and the iridoid terpene secologanin (**Figure 13**). The involvement of an iridoid monoterpene in these indole alkaloid pathways was first proposed after the structures of several iridoid terpenes were elucidated, and secologanin was identified as the specific iridoid precursor.[85–87] Secologanin is itself a natural product, and the biosynthetic pathway for this molecule has not been fully elucidated, although feeding studies with *C. roseus* cell suspension cultures and ^{13}C-glucose strongly suggest that secologanin is ultimately derived from the triose phosphate/pyruvate or 'nonmevalonate'

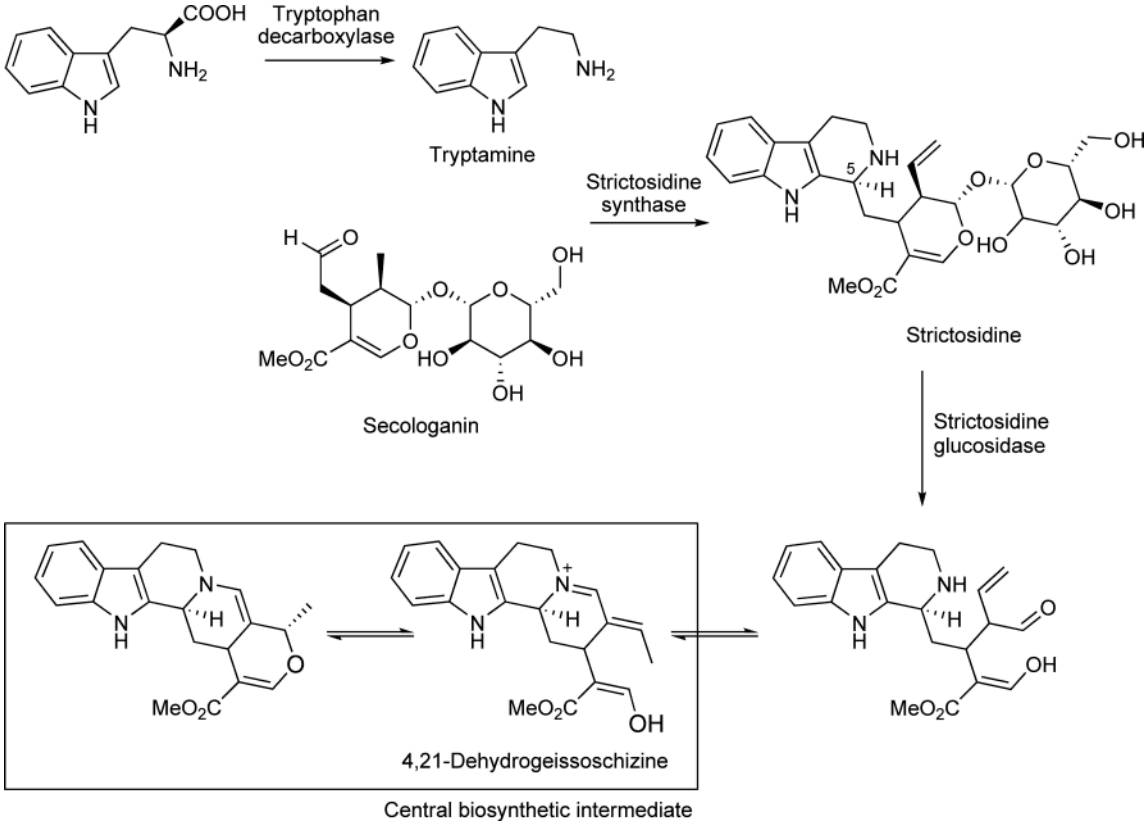


Figure 13 Biosynthesis of deglycosylated strictosidine, the central biosynthetic intermediate in monoterpene indole alkaloid biosynthesis.

pathway.[88] Several enzymes involved in the biosynthesis of Isoprenylpyrophosphate-1-deoxy-D-xylulose 5-phosphate (IPP)–2-Methyl-D-erythritol-4-phosphate (DXP) synthase, DXP reductoisomerase, and MEP synthase – have been cloned from *C. roseus*.[89,90] Several other genes involved in the later steps of secologanin biosynthesis, namely geraniol-10-hydroxylase,[91,92] and secologanin synthase[93,94] have been identified. Tryptophan decarboxylase, a pyridoxal-dependent enzyme, converts tryptophan to tryptamine to yield the amine-containing starting material for this family of alkaloids.[95,96] The enzyme strictosidine synthase catalyzes a stereoselective Pictet–Spengler condensation between tryptamine and secologanin to yield strictosidine. Strictosidine synthase[97] has been cloned from the plants *C. roseus*,[98] *Rauwolfia serpentina*,[99] and *Ophiorrhiza pumila*.[100] A crystal structure of strictosidine synthase from *R. serpentina* has been reported,[101] and a number of reports have indicated that the substrate specificity of the enzyme can be modulated.[102–104]

In most monoterpene indole alkaloids, strictosidine is deglycosylated by a dedicated β-glucosidase, which converts the substrate to a reactive hemiacetal intermediate.[105–107] This hemiacetal opens to form a dialdehyde intermediate, which then forms dehydrogeissoschizine. The enol form of dehydrogeissoschizine can undergo 1,4 conjugate addition to produce the heteroyohimbine cathenamine, as well as a variety of other isomers.[107–111]

The biosynthetic pathway for ajmaline in *R. serpentina* is one of the best-characterized branches of the terpenoid indole alkaloid pathways (**Figure 14**). Much of this progress has been detailed in an extensive review.[112] Like all other terpenoid indole alkaloids, ajmaline, an antiarrhythmic drug with potent sodium channel-blocking properties,[113] is derived from deglycosylated strictosidine. At least eight enzymes are predicted to catalyze the subsequent steps of ajmaline biosynthesis that occur after strictosidine deglycosylation. The sarpagan alkaloid, polyneuridine aldehyde, is a known early intermediate of the ajmaline pathway. A mechanism in which the sarpagan bridge enzyme transforms an isomer of deglycosylated strictosidine to polyneuridine aldehyde has been proposed.[114] A membrane-protein fraction of an *R. serpentina* extract transformed labeled strictosidine into sarpagan-type alkaloids. The enzyme activity was shown to be dependent on NADPH and molecular oxygen, suggesting that sarpagan bridge enzyme may be a cytochrome P-450

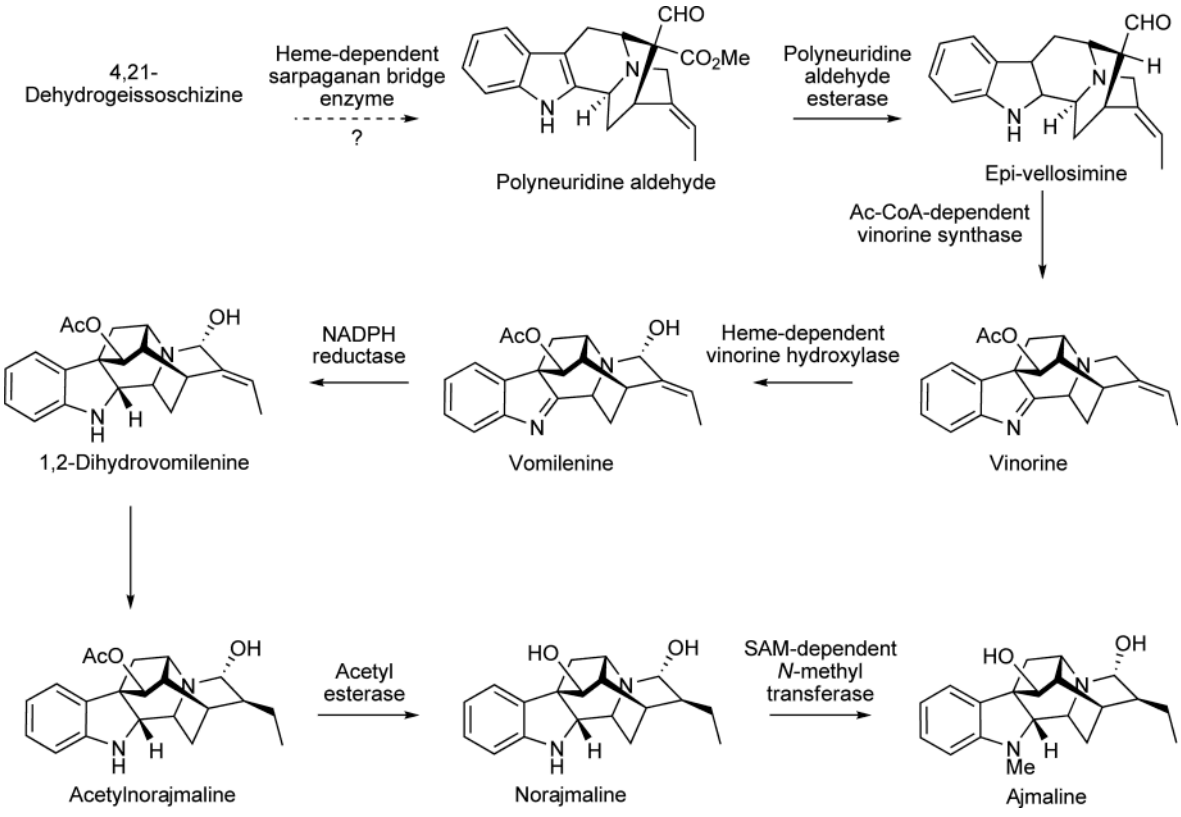


Figure 14 Biosynthesis of ajmaline.

enzyme.[115,116] Isolation of this enzyme will yield further insight into this key step that commits the deglycosylated strictosidine intermediate to the sarpagan- and ajmalan-type alkaloid pathways.

Polyneuridine aldehyde esterase then hydrolyzes the polyneuridine aldehyde methyl ester, which generates an acid that spontaneously decarboxylates to yield epi-vellosamine. This enzyme has been cloned from a *Rauwolfia* cDNA library, heterologously expressed in *E. coli,* and subjected to detailed mechanistic studies.[117–119] Polyneuridine aldehyde esterase appears to be a member of the α/β hydrolase super family and contains a Ser, His, Asp catalytic triad.[117–119] Site-directed mutagenesis indicates that each residue of the catalytic triad is required for activity. Vinorine synthase transforms the sarpagan alkaloid epi-vellosamine to the ajmalan alkaloid vinorine.[120] Vinorine synthase also has been purified from *Rauwolfia* cell culture, subjected to protein sequencing, and cloned from a cDNA library.[121,122] The enzyme, which seems to be an acetyltransferase homolog, has been crystallized and subjected to site-directed mutagenesis studies of this protein, leading to a proposed mechanism.[123] Vinorine hydroxylase hydroxylates vinorine to form vomilene.[124] Vinorine hydroxylase seems to be a P-450 enzyme, and has not been cloned yet. The indolenine bond of vomilene is reduced by an NADPH-dependent reductase to yield 1,2-dihydrovomilenene. A second enzyme, 1,2-dihydrovomilenene reductase, then reduces this product to acetylnorajmaline. Partial protein sequences have been obtained for both of the purified reductases. Although several putative clones that encode these proteins have been isolated, the activity of these clones has not been verified yet.[125,126] An acetylesterase then hydrolyzes the acetyl link of acetylnorajmaline to yield norajmaline. This esterase has been purified from *R. serpentina* cell suspension cultures, and a full-length clone has been isolated from a cDNA library. Expression of the gene in tobacco leaves successfully yielded protein with the expected enzymatic activity.[127] In the final step of ajmaline biosynthesis, an *N*-methyltransferase introduces a methyl group at the indole nitrogen of norajmaline. Although this enzymatic activity has been detected in crude cell extracts, the enzyme has not been characterized additionally.[128] In summary, the enzymatic activities for all steps of ajmaline biosynthesis have been detected. Strictosidine synthase, strictosidine glucosidases, polyneuridine aldehyde esterase, vinorine synthase, and 17-*O*-acetyl-ajmalanesterase have been cloned. Putative clones for vinorine hydroxylase, vomilenine reductase, and 1,2-dihydrovomilenen reductase have been isolated. *N*-Methyltransferase activity and sarpagan bridge enzyme activities have only been detected in crude cell extracts.

Ajmalicine (raubasine) affects smooth muscle function and is used to help prevent strokes,[129] and tetrahydroalstonine exhibits antipsychotic properties (**Figure 15**).[130] These compounds are found in a variety of plants, including *C. roseus* and *R. serpentina.* A partially purified NADPH-dependent reductase isolated from a tetrahydroalstonine that produces a *C. roseus* cell line was shown to catalyze the conversion of cathenamine, a spontaneous reaction product that results after strictosidine deglycosylation, to tetrahydroalstonine *in vitro* (**Figure 15**).[131] A second *C. roseus* cell line contains an additional reductase that produces ajmalicine. Labeling studies performed with crude *C. roseus* cell extracts in the presence of D_2O or deuterated, reduced form of Nicotinamide adenine dinucleotide phosphate (NADPD) support a mechanism in which the reductase acts on the iminium form of cathenamine.[132]

Vindoline, an aspidosperma-type alkaloid produced by *C. roseus*, is a key precursor for vinblastine, an anticancer drug that is the most important pharmaceutical product of *C. roseus* (**Figure 16**). Vindoline, like ajmalicine and ajmaline, is produced from deglycosylated strictosidine. Deglycosylated strictosidine is converted to tabersonine, an aspidosperma-type alkaloid, through a series of biochemical steps for which no enzymatic information exists. Studies by numerous groups in the 1960s and 1970s enabled detailed hypothetical proposals of the biosynthesis of aspidosperma-type alkaloids in *C. roseus.*[133–143] These proposed pathways are based on feeding studies of isotopically labeled substrates to seedlings or shoots, isolation of discrete intermediates from plant materials, and from biomimetic model reactions. More details are known about the elaboration of tabersonine to vindoline.[144] Tabersonine-16-hydroxylase, a cytochrome P-450, hydroxylates tabersonine to 16-hydroxy-tabsersonine.[145,146] This hydroxyl group is then methylated by a SAM-dependent *O*-methyltransferase to yield 16-methoxy-tabersonine; this enzyme (16-hydroxytabersonine-16-*O*-methyltransferase) has recently been cloned.[147] In the next step, hydration of a double bond by an uncharacterized enzyme produces 16-methoxy-2,3-dihydro-3-hydroxytabersonine. Transfer of a methyl group to the indole nitrogen by an *N*-methyltransferase yields desacetoxyvindoline. This methyltransferase activity has been detected only in differentiated plants, not in plant cell cultures.[148] The resulting intermediate, deacetylvindoline, is produced by the oxoglutarate-dependent dioxygenase enzyme desacetylvindoline 4-hydroxylase. This enzyme has been cloned and is also absent from plant cell cultures.[149] In the last step,

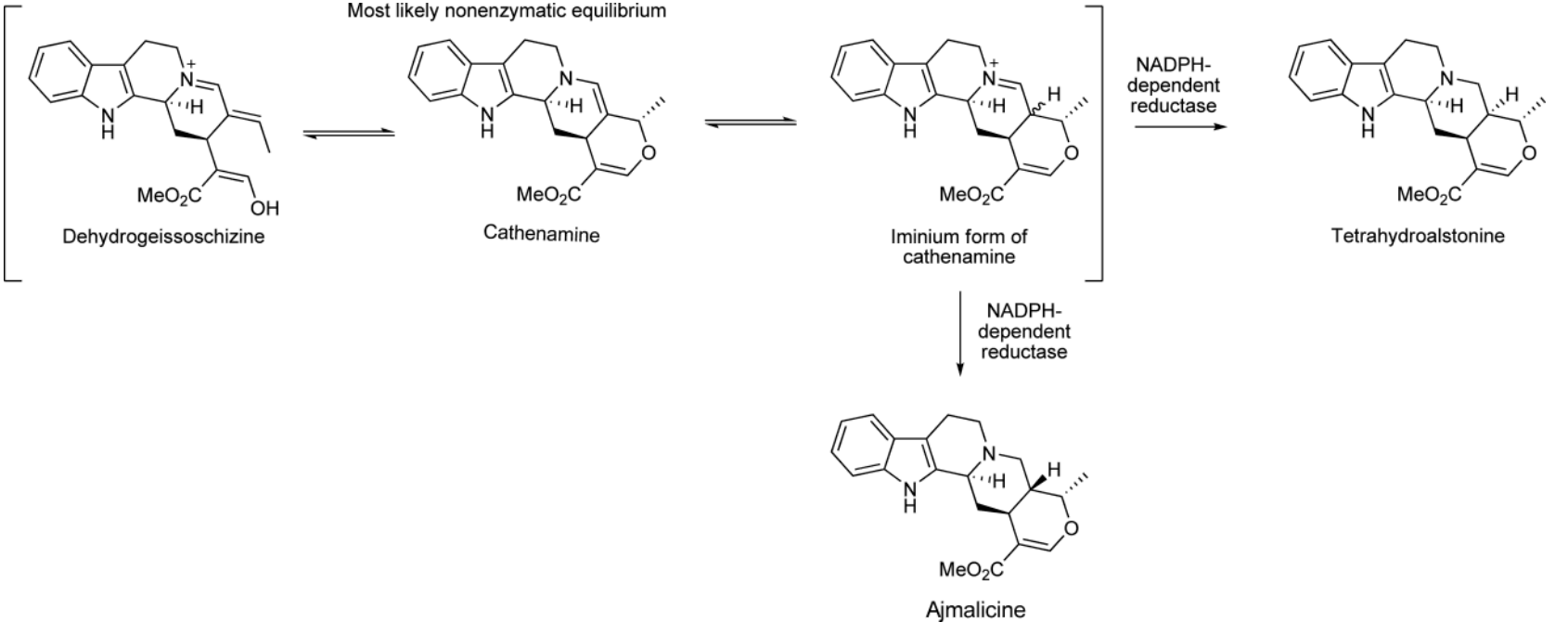


Figure 15 Biosynthesis of ajmalicine and tetrahydroalstonine.

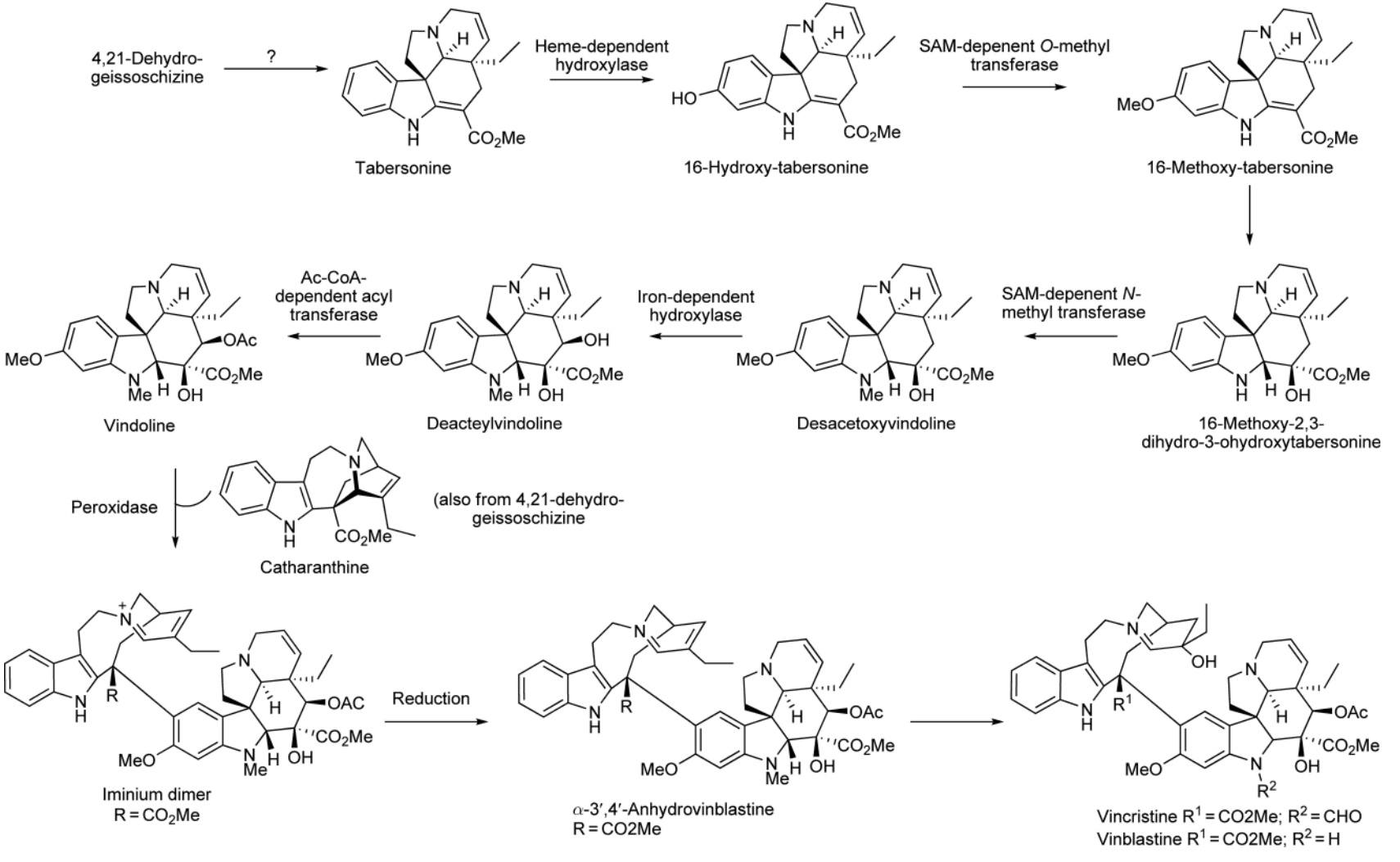


Figure 16 Biosynthesis of vindoline and vinblastine.

desacetylvindoline is acetylated by desacetylvindoline *O*-acetyltransferase. This enzyme, also absent from nondifferentiated plant material, has been cloned successfully.[150]

Vinblastine and the structurally related vincristine are highly effective anticancer agents currently used clinically against leukemia and other cancers.[151,152] Inspection of these bisindole alkaloids indicates that they are derived from coupling of vindoline and catharanthine, which is believed to proceed via the formation of an iminum intermediate with catharanthine (**Figure 16**). This iminium intermediate is reduced to form anhydrovinblastine, a naturally occurring compound in *C. roseus* plants.[153]

Peroxidase-containing fractions of plant extracts were found to catalyze the formation of the bisindole dehydrovinblastine from catharanthine and vindoline.[154,155] The peroxidase CRPRX1 (α-3′,4′-anhydrovinblastine synthase), purified and cloned from *C. roseus* leaves, has been demonstrated to convert vindoline and catharanthine to anhydrovinblastine.[156,157] Catharanthine is most likely oxidized to an iminium ion, which then reacts with the relatively nucleophilic vindoline.[158] Although this peroxidase is not highly substrate specific for catharanthine and vindoline, localization studies strongly suggest that CRPRX1 is the dedicated peroxidase for bisindole formation.[157,158] Finally, after formation of dehydrovinblastine, hydroxylation of the double bond yields vinblastine, and oxidation of the *N*-methyl group yields vincristine.

As in morphine biosynthesis, the knowledge of the enzyme sequences allows a more detailed understanding of the localization of the enzymes.[159] Strictosidine synthase seems to be localized to the vacuole,[98] and strictosidine glucosidase is believed to be associated with the cytosol and the membrane of the endoplasmic reticulum.[160] Tabersonine-16-hydroxylase is associated with the endoplasmic reticulum membrane;[145] *N*-Methyltransferase activity is believed to be associated with the thylakoid, a structure located within the chloroplast;[148,161] and vindoline-4-hydroxylase and desacetylvindoline *O*-acetyltransferase are believed to be localized to the cytosol.[161,162] In addition to subcellular compartmentalization, specific cell types are required for the biosynthesis of some terpenoid alkaloids, as is the case in morphine biosynthesis described above. Several enzymes involved in the early stages of secologanin biosynthesis seem to be localized to the phloem parenchyma, as evidenced by immunocytochemistry and *in situ* RNA hybridization studies.[163] However, additional studies have suggested that these genes also are observed in the epidermis and laticifers.[164] Studies of the localization of vindoline biosynthetic enzymes by using immunocytochemistry and *in situ* RNA hybridization strongly suggest that the midpart of the vindoline pathway (tryptophan decarboxylase, strictosidine synthase, and tabersonine-16-hydroxylase) takes place in epidermal cells of leaves and stems. However, the later steps catalyzed by desacetylvindoline 4-hydroxylase and desacetylvindoline *O*-acetyltransferase take place in specialized cells, the laticifers, and idioblasts.[165] As with isoquinoline alkaloid biosynthesis, deconvolution of the enzyme localization patterns remains a challenging endeavor.

Again, the partially elucidated pathways of monoterpene indole alkaloid biosynthesis have allowed metabolic engineering efforts. These studies have primarily taken place in *C. roseus*. Strictosidine synthase and tryptophan decarboxylase have been overexpressed in *C. roseus* cell cultures.[166,167] Generally, overexpression of tryptophan decarboxylase does not seem to have a significant impact on alkaloid production, although overexpression of strictosidine synthase does seem to improve alkaloid yields. Overexpression of tryptophan and secologanin biosynthetic enzymes in *C. roseus* hairy root cultures resulted in modest increases in terpenoid indole alkaloid production.[168,169] Secologanin biosynthesis seems to be the rate-limiting factor in alkaloid production.[170] Using a combination of unnatural tryptamine analogs and a reengineered strictosidine synthase enzyme, the biosynthetic pathway can be used to produce alkaloid derivatives.[171,172] Although strictosidine synthase and strictosidine glucosidase enzymes have been expressed heterologously in yeast,[173] efforts to express heterologously terpenoid indole alkaloids currently are limited because the majority of the biosynthetic genes remain uncloned.

Transcription factors that upregulate strictosidine synthase,[174] as well as a transcription factor that coordinately upregulates expression of several terpenoid indole alkaloid biosynthetic genes, have been found.[14] Several zinc finger proteins that act as transcriptional repressors to tryptophan decarboxylase and strictosidine synthase also have been identified.[175] Manipulation of these transcription factors may allow tight control of the regulation of terpenoid indole alkaloid production. Interestingly, expression of a transcription factor from *Arabidopsis thaliana* in *C. roseus* cell cultures results in an increase in alkaloid production.[176] The dramatic increases in alkaloid production that have been noted in morphine metabolic engineering efforts have for the most part not been observed in the monoterpene indole alkaloids. Notably, more enzymes in the isoquinoline

biosynthetic pathway are known. Presumably, as more monoterpene indole alkaloid biosynthetic enzymes are identified, one may be identified that has a significant impact on the alkaloid expression levels.

1.25.7 Tropane Alkaloid Biosynthesis

The tropane alkaloids hyoscyamine and scopolamine (**Figure 17**) function as acetylcholine receptor antagonists and are used clinically as parasympatholytics. The illegal drug cocaine is also a tropane alkaloid. The tropane alkaloids are biosynthesized primarily in plants of the family Solonaceae, which includes *Hyoscyamus*, *Duboisia*, *Atropa*, and *Scopolia*.[17,18] Nicotine, although perhaps not apparent immediately from its structure, is related biosynthetically to the tropane alkaloids.

Tropane alkaloid biosynthesis has been studied at the biochemical level, and several enzymes from the biosynthetic pathway have been isolated and cloned, although the pathway has not been elucidated completely at the genetic level (**Figure 18**).[177] In plants, L-arginine is converted to the nonproteogenic amino acid L-ornithine by the urease enzyme arginase. Ornithine decarboxylase then decarboxylates ornithine to yield the diamine putrescine. In *Hyoscyamus*, *Duboisia*, and *Atropa*, putrescine (so-named because of its odor) serves as the common precursor for the tropane alkaloids. Putrescine is N-methylated by a SAM-dependent methyltransferase that has been cloned to yield *N*-methylputrescine.[178,179] Putrescine *N*-methyltransferase now has been cloned from a variety of plant species,[180–182] and site-directed mutagenesis and homology models have led to insights into the structure–function relationships of this enzyme.[182] *N*-Methylputrescine is then oxidized by a diamine oxidase to form 4-methylaminobutanal, which then cyclizes, most likely nonenzymatically, to form the *N*-methyl-D-pyrrolinium ion.[183–185] This enzyme, which recently has been cloned, seems to be a copper-dependent amine oxidase.[186] Immunoprecipitation experiments suggest that this enzyme associates with the enzyme *S*-adenosylhomocysteine hydrolase.[187] The pyrrolinium ion is then converted to the tropanone skeleton by as yet uncharacterized enzymes. Although no enzymatic information is available, chemical labeling studies have indicated that an acetate-derived moiety condenses with the pyrrolinium ion.[17] Tropanone is then reduced via an NADPH-dependent reductase to tropine that has been cloned from *Hyoscyamus niger*.[188,189] All tropane-producing plants seem to contain two tropinone reductases, which create a branch point in the pathway. Tropinone reductase I yields the tropane skeleton, whereas tropinone reductase II yields the opposite stereocenter, pseudotropine.[190] Tropane is converted to scopolamine or hyoscyamine, while the tropinone reductase II product pseudotropine leads to calystegines.[191] These two tropinone reductases have been crystallized, and site-directed mutagenesis studies indicate that the stereoselectivity of the enzymes can be switched by rational protein engineering.[192,193]

The biosynthesis of scopolamine is the best characterized of the tropane alkaloids. After action by tropinone reductase I, tropine is condensed with phenylacetate through the action of a P-450 enzyme to form littorine.[13] The phenyllactate moiety is believed to derive from an intermediate involved in phenylalanine metabolism.[17] Littorine then undergoes rearrangement to form hyoscyamine. The enzyme that catalyzes this rearrangement was originally believed to proceed via a radical mechanism using SAM as the source of an adenosyl radical.[194–197] However, a large-scale RNAi study performed in *H. niger* suggests that a P-450 enzyme, followed by action of a reductase enzyme, is responsible for the rearrangement. Hyoscyamine 6β-hydroxylase

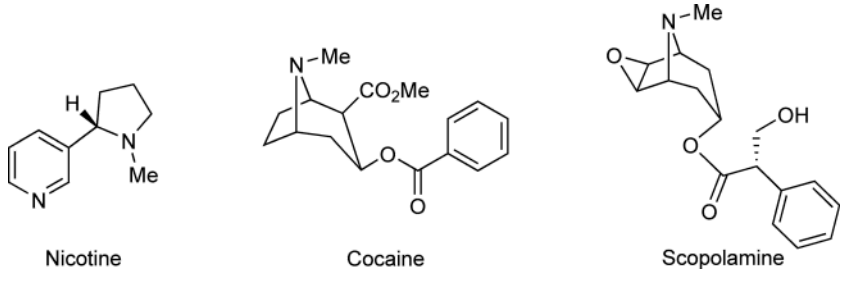


Figure 17 Representative type tropane and nicotinic alkaloids nicotine, cocaine, and scopolamine.

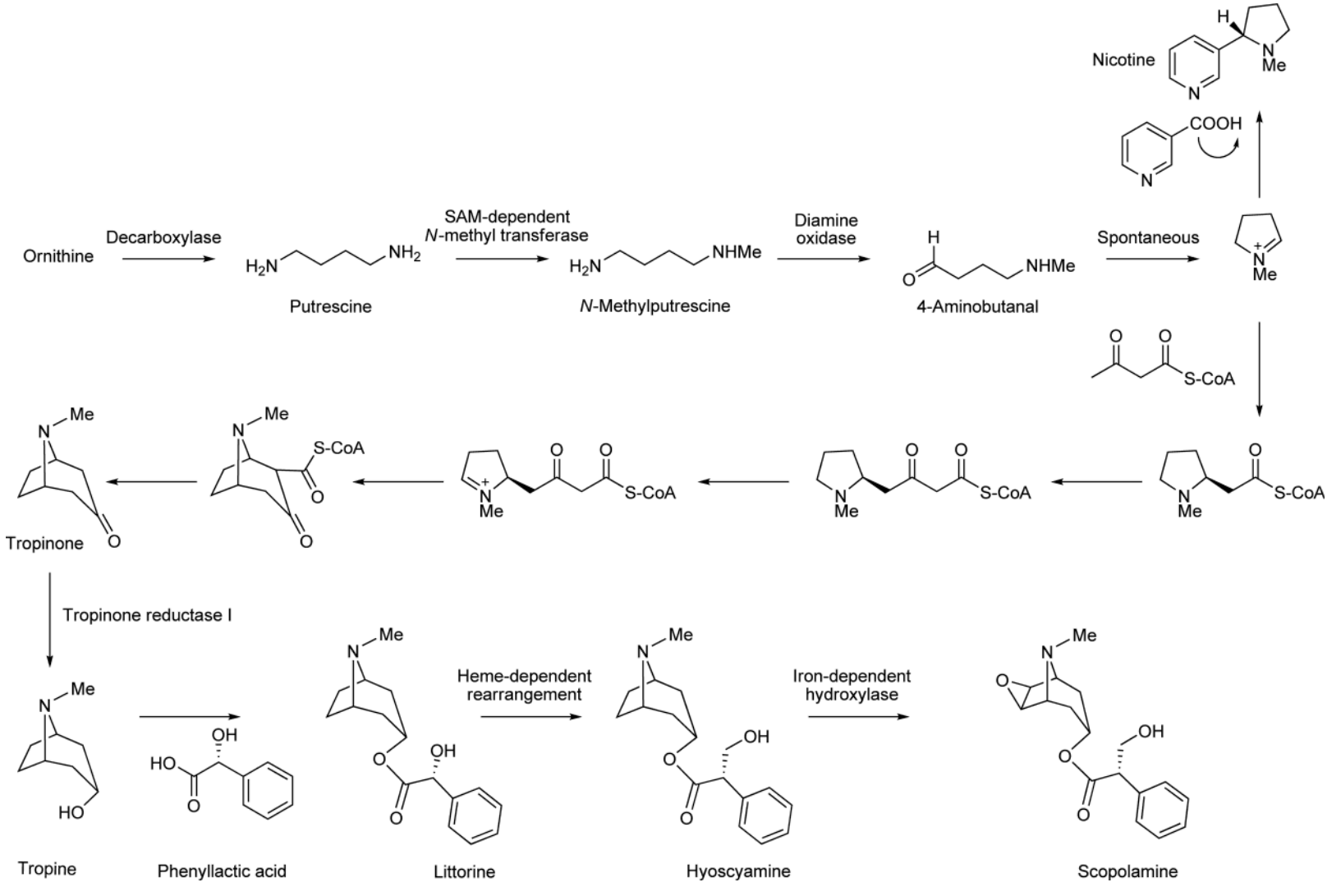


Figure 18 Biosynthesis of scopolamine.

catalyzes the hydroxylation of hyoscyamine to 6β-hydroxyhyoscyamine as well as the epoxidation to scopolamine.[198,199] Hyoscyamine 6β-hydroxylase, which has been cloned and expressed heterologously,[200] is a nonheme, iron-dependent, oxoglutarate-dependent protein. The epoxidation reaction appears to occur more slowly than the hydroxylation reaction. The tropane alkaloids seem to be formed in the roots and then transported to the aerial parts of the plant.[201]

Atropa belladonna plants have been transformed with hyoscyamine 6β-hydroxylase from *H. niger*. *A. belladonna* normally produces high levels of hyoscyamine, the precursor for the more pharmaceutically valuable alkaloid scopolamine. However, in one of the earliest demonstrations of plant natural product metabolic engineering, after transformation with hyoscyamine 6β-hydroxylase, transgenic *A. belladonna* plants were shown to accumulate scopolamine almost exclusively.[202] Additionally, the levels of tropane alkaloid production in a variety of hairy root cultures were altered by overexpression of methyltransferase putrescine-*N*-methyltransferase and hyoscyamine 6β-hydroxylase. Overexpression of both of these enzymes in a hairy root cell culture resulted in significant increases in scopolamine production.[202,203] Fluorinated phenyllactic acid substrates could be incorporated into the pathway,[204] and several substrates derived from putrescine analogs were turned over by the enzymes of several Solonaceae species.[205]

1.25.8 Purine Alkaloid Biosynthesis

Caffeine, a purine alkaloid, is one of the most widely ingested of all natural products. Caffeine is a natural component of coffee, tea, and cocoa, and the impact of caffeine on human health has been studied extensively. The biosynthetic pathway of caffeine has been elucidated recently on the genetic level (**Figure 19**), and most work has focused on the plant species *Coffea* (coffee) and *Camellia* (tea).[19,206] Xanthosine, which is derived from purine metabolites, is the first committed intermediate in caffeine biosynthesis. Xanthosine can be formed from *de novo* purine biosynthesis, SAM cofactor, the adenylate pool, and the guanylate pool.[19] *De novo* purine biosynthesis and the adenosine from SAM are believed to be the most important sources of xanthosine.[19,207] Xanthosine is methylated to yield *N*-methylxanthosine by the enzyme xanthosine *N*-methyltransferase (XMT) (also called 7-methylxanthosine synthase).[208–210] *N*-Methylxanthosine is converted to *N*-methylxanthine by methylxanthine nucleosidase, an enzyme that has not been cloned, but is present in many noncaffeine-producing organisms.[211] *N*-Methylxanthine is converted to theobromine by 7-methylxanthine-*N*-methyltransferase (MXMT) (also called theobromine synthase), a second *N*-methyltransferase.[209,212] Theobromine is converted to caffeine by a final *N*-methyltransferase, dimethylxanthine-*N*-methyltransferase (DXMT) (also called caffeine synthase).[209]

Coffee and tea plants seem to contain a variety of *N*-methyltransferase enzymes that have varying substrate specificity.[19,206] For example, a caffeine synthase enzyme isolated from tea leaves catalyzes both the N-methylation of *N*-methylxanthine and theobromine.[213] The substrate specificity of the methyltransferases can be changed by site-directed mutagenesis.[214]

Coffee beans with low caffeine levels could be valuable commercially, given the demand for decaffeinated coffee. Because of the discovery of these *N*-methyltransferase genes, genetically engineered coffee plants with reduced caffeine content now can be constructed.[212,215] For example, a 70% reduction in caffeine content in *Coffea* was obtained by downregulating MXMT (theobromine synthase) using RNAi.[216] Additionally, the promoter of one of the *N*-methyltransferases has been discovered recently, which may allow transcriptional gene silencing.[217]

1.25.9 Conclusions and Outlook

Alkaloids constitute a structurally diverse array of natural products, and these compounds have a wide range of biological activities. Many have important pharmaceutical uses. Plants are regarded as the oldest source of alkaloids, and some of the most widely recognized alkaloids, such as morphine, quinine, strychnine, and cocaine, are derived from plants. However, rapid advances in molecular biology and sequencing of bacterial and fungal genomes have fostered the discovery of new alkaloids in these simpler microbial organisms, and a

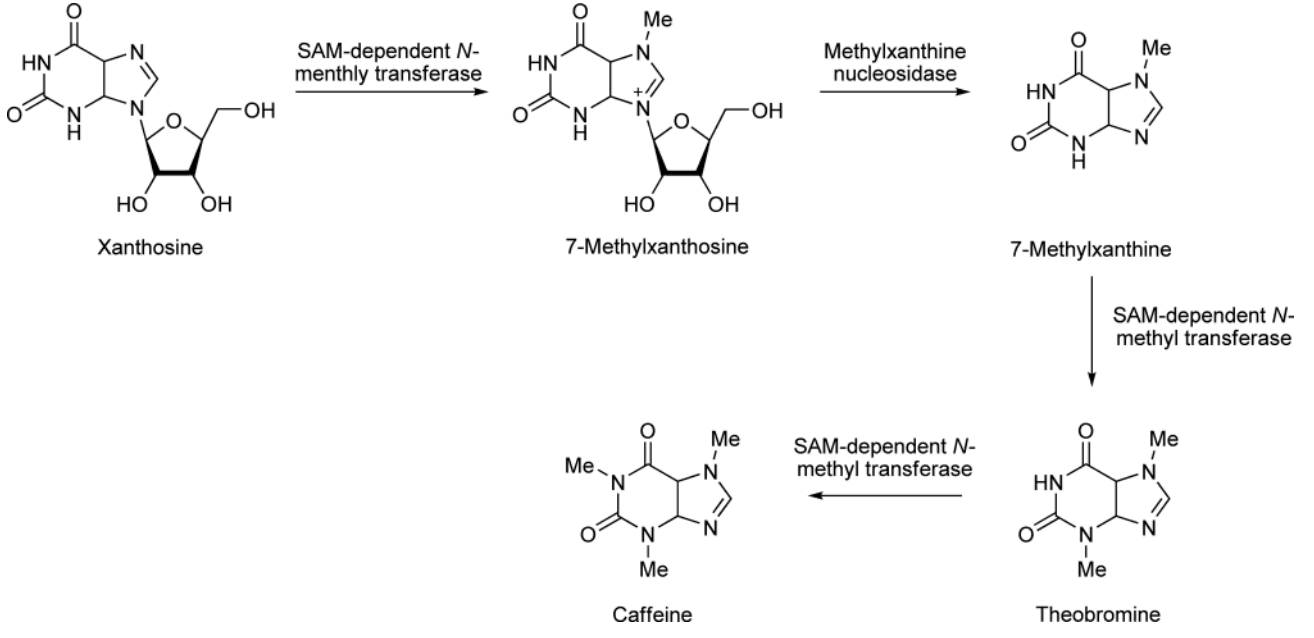


Figure 19 Biosynthesis of caffeine.

wealth of biosynthetic information for these compounds has been obtained in a number of recent mechanistic studies. Undoubtedly, many more microbe-derived alkaloids remain to be discovered. Given the medicinal importance of many alkaloids, metabolic engineering efforts serve as an important application of biosynthetic studies. Metabolic engineering has been used to increase the production levels of alkaloids, reconstitute alkaloid biosynthesis in simpler host organisms, and rationally modify the structure of alkaloids by engineering of the biosynthetic enzymes. The past several years have resulted in major advances in alkaloid discovery, biosynthetic pathway elucidation, and metabolic engineering.

Glossary

acridine-type alkaloids Anthranilic acid-derived natural products.
alkaloid Nitrogen-containing organic substances of natural origin with basic character.
indolizidine alkaloids Lysine-derived natural products.
monoterpene indole alkaloids Tryptophan- and secologanin-derived natural products.
piperdine alkaloids Lysine-derived natural products.
pyrrolizidine alkaloids Ornithine-derived natural products.
quinazoline alkaloids Anthranilic acid-derived natural products.
quinolizidine alkaloids Lysine-derived natural products.
tetrahydroisoquinoline alkaloids Tyrosine-derived natural products.
tropane alkaloids Ornithine-derived natural products.

References

1. J. Ziegler; P. J. Facchini, *Annu. Rev. Plant. Biol.* **2008**, *59*, 735–769.
2. M. A. Fischbach; C. T. Walsh, *Chem. Rev.* **2006**, *106*, 3468–3496.
3. M. Hesse, *Alkaloids: Nature's Curse or Blessing?* Wiley-VCH: Weinheim, 2002.
4. Y. S. Kim; H. Uefuji; S. Ogita; H. Sano, *Transgenic Res.* **2006**, *15*, 667–672.
5. M. A. Fischbach; J. Clardy, *Nat. Chem. Biol.* **2007**, *3*, 353–355.
6. R. D. Firn; C. G. Jones, *Mol. Microbiol.* **2000**, *37*, 989–994.
7. W. Eisenreich; A. Bacher, *Phytochemistry* **2007**, *68*, 2799–2815.
8. X. Qi; S. Bakht; M. Leggett; C. Maxwell; R. Melton; A. Osbourn, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8233–8238.
9. B. Field; A. E. Osbourn, *Science* **2008**, *320*, 543–547.
10. T. Hashimoto; Y. Yamada, *Curr. Opin. Biotech.* **2003**, *14*, 163–168.
11. T. M. Kutchan; J. Schroder, *Meth. Enzymol.* **2002**, *36*, 370–381.
12. K. J. Martin; A. B. Pardee, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3789–3791.
13. R. Li; D. W. Reed; E. Liu; J. Nowak; L. E. Pelcher; J. E. Page; P. S. Covello, *Chem. Biol.* **2006**, *13*, 513–520.
14. L. van der Fits; J. Memelink, *Science* **2000**, *289*, 295–297.
15. P. J. Facchini; D. A. Bird; R. Bourgault; J. M. Hagel; D. K. Liscombe; B. P. MacLeod; K. G. Zulak, *Can. J. Bot.* **2005**, *83*, 1189–1206.
16. S. E. O'Connor; J. Maresh, *Nat. Prod. Rep.* **2006**, *23*, 532–547.
17. A. J. Humphrey; D. O'Hagan, *Nat. Prod. Rep.* **2001**, *18*, 494–502.
18. T. Hemscheidt, *Top. Curr. Chem.* **2000**, *209*, 175–206.
19. H. Ashihara; A. Crozier, *Trends Plant Sci.* **2001**, *6*, 407–413.
20. L. T. Tan, *Phytochemistry* **2007**, *68*, 954–979.
21. B. H. Long; W. C. Rose; D. M. Vyas; J. A. Matson; S. Forenza, *Curr. Med. Chem. Anticancer Agents* **2002**, *2*, 255–266.
22. H. B. Bode; R. Mueller, *Angew. Chem. Intl. Ed. Engl.* **2005**, *44*, 6828–6846.
23. S. Mo; P. K. Sydor; C. Corre; M. M. Alhamadsheh; A. E. Stanley; S. W. Haynes; L. Song; K. A. Reynolds; G. L. Challis, *Chem. Biol.* **2008**, *15*, 137–148.
24. A. R. Howard-Jones; C. T. Walsh, *J. Am. Chem. Soc.* **2006**, *128*, 12289–12298.
25. A. R. Howard-Jones; C. T. Walsh, *J. Am. Chem. Soc.* **2007**, *129*, 11016–11017.
26. K. Shinoda; T. Hasegawa; H. Sato; M. Shinozaki; H. Kuramoto; Y. Takamiya; T. Sato; N. Nikaidou; T. Watanabe; T. Hoshino, *Chem. Commun.* **2007**, 4140–4142.
27. Y. Hu; V. Phelan; I. Ntai; C. M. Farnet; E. Zazopoulos; B. O. Bachmann, *Chem. Biol.* **2007**, *14*, 691–701.
28. L. Li; W. Deng; J. Song; W. Ding; Q. F. Zhao; C. Peng; W. W. Song; G. L. Tang; W. Liu, *J. Bacteriol.* **2008**, *190*, 251–263.
29. J. T. Nelson; J. Lee; J. W. Sims; E. W. Schmidt, *Appl. Environ. Microbiol.* **2007**, *73*, 3575–3580.
30. J. D. Scott; R. M. Williams, *Chem. Rev.* **2002**, *102*, 1669–1730.
31. C. L. Schardl; D. G. Panaccione; P. Tudzynski, *Alkaloids Chem. Biol.* **2006**, *63*, 45–86.
32. S. Saikia; M. J. Nicholson; C. Young; E. J. Parker; B. Scott, *Mycol. Res.* **2008**, *112*, 184–199.
33. J. W. Blunt; B. P. Copp; W. P. Hu; M. H. Munro; P. T. Northcote; M. R. Prinsep, *Nat. Prod. Rep.* **2008**, *25*, 35–94.

34. D. D. Baker; K. A. Alvi, *Curr. Opin. Biotechnol.* **2004**, *15*, 576–583.
35. R. S. Allen; A. G. Millgate; J. A. Chitty; J. Thisleton; J. A. C. Miller; A. J. Fist; W. L. Gerlach; P. J. Larkin, *Nat. Biotechnol.* **2004**, *22*, 1559–1566.
36. M. Rueffer; M. H. Zenk, *Z. Naturforsch.* **1987**, *42c*, 319–332.
37. N. Samanani; D. K. Liscombe; P. J. Facchini, *Plant J.* **2004**, *40*, 302–313.
38. H. Minami; E. Dubouzet; K. Iwasa; F. Sato, *J. Biol. Chem.* **2007**, *282*, 6274–6282.
39. F. Sato; T. Tsujita; Y. Katagiri; S. Yoshida; Y. Yamada, *Eur. J. Biochem.* **1994**, *225*, 125–131.
40. A. Ounaroon; G. Decker; J. Schmidt; F. Lottspeich; T. M. Kutchan, *Plant J.* **2003**, *36*, 808–819.
41. K.-B. Choi; T. Morishige; N. Shitan; K. Yazaki; F. Sato, *J. Biol. Chem.* **2002**, *277*, 830–835.
42. K.-B. Choi; T. Morishige; F. Sato, *Phytochemistry* **2001**, *56*, 649–655.
43. H. H. Pauli; T. M. Kutchan, *Plant J.* **1998**, *13*, 793–801.
44. T. Morishige; T. Tsujita; Y. Yamada; F. Sato, *J. Biol. Chem.* **2000**, *275*, 23398–23405.
45. N. Samanani; S.-U. Park; P. J. Facchini, *Plant Cell* **2005**, *17*, 915–926.
46. J. Ziegler; M. L. Diaz-Chavez; R. Kramell; C. Ammer; T. M. Kutchan, *Planta* **2005**, *222*, 458–471.
47. H. Dittrich; T. M. Kutchan, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 9969–9973.
48. P. J. Facchini; C. Penzes; A. G. Johnson; D. Bull, *Plant Physiol.* **1996**, *112*, 1669–1677.
49. K. Hauschild; H. H. Pauli; T. M. Kutchan, *Plant Mol. Biol.* **1998**, *36*, 473–478.
50. T. M. Kutchan; H. Dittrich, *J. Biol. Chem.* **1995**, *270*, 24475–24481.
51. A. Winkler; F. Hartner; T. M. Kutchan; A. Glieder; P. Macheroux, *J. Biol. Chem.* **2006**, *281*, 21276–21285.
52. A. Winkler; A. Lyskowski; S. Riedl; M. Puhl; T. M. Kutchan; P. Macheroux; K. Gruber, *Nat. Chem. Biol.* **2008**, *4*, 739–741.
53. N. Takeshita; H. Fujiwara; H. Mimura; J. H. Fitchen; Y. Yamada; F. Sato, *Plant Cell Physiol.* **1995**, *36*, 29–36.
54. N. Ikezawa; M. Tanaka; M. Nagayoshi; R. Shinkyo; T. Sakaki; K. Inouye; F. Sato, *J. Biol. Chem.* **2003**, *278*, 38557–38565.
55. M. Amann; N. Nagakura; M. H. Zenk, *Eur. J. Biochem.* **1988**, *175*, 17–25.
56. F. Sato; Y. Yamada, *Phytochemistry* **1984**, *23*, 281–285.
57. K. Matsubara; S. Kitani; T. Yoshioka; T. Morimoto; Y. Fujita, *J. Chem. Technol. Biotechnol.* **1989**, *46*, 61–69.
58. M. H. Zenk, *Pure Appl. Chem.* **1994**, *66*, 2023–2028.
59. W. Bauer; M. H. Zenk, *Phytochemistry* **1991**, *30*, 2953–2961.
60. N. Ikezawa; K. Iwasa; F. Sato, *FEBS J.* **2007**, *274*, 1019–1035.
61. D. K. Liscombe; P. J. Facchini, *J. Biol. Chem.* **2007**, *282*, 14741–14751.
62. T. Tanahashi; M. H. Zenk, *Phytochemistry* **1990**, *29*, 1113–1122.
63. H. Arakawa; W. G. Clark; M. Psenak; C. J. Coscia, *Arch. Biochem. Biophys.* **1992**, *299*, 1–7.
64. H. M. Schumacher; M. H. Zenk, *Plant Cell Rep.* **1988**, *7*, 43–46.
65. K. Hirata; C. Poeaknapo; J. Schmidt; M. H. Zenk, *Phytochemistry* **2004**, *65*, 1039–1046.
66. W. De-Eknamkul; M. H. Zenk, *Phytochemistry* **1992**, *31*, 813–821.
67. R. Gerardy; M. H. Zenk, *Phytochemistry* **1992**, *32*, 79–86.
68. J. Ziegler; S. Voigtlaender; J. Schmidt; R. Kramell; O. Miersch; C. Ammer; A. Gesell; T. M. Kutchan, *Plant J.* **2006**, *48*, 177–192.
69. T. Grothe; R. Lenz; T. M. Kutchan, *J. Biol. Chem.* **2001**, *276*, 30717–30723.
70. R. Lenz; M. H. Zenk, *Tetrahedron Lett.* **1995**, *36*, 2449–2452.
71. B. Unterlinner; R. Lenz; T. M. Kutchan, *Plant J.* **1999**, *18*, 465–475.
72. C. Poeaknapo; J. Schmidt; M. Brandsch; B. Drager; M. H. Zenk, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14091–14096.
73. N. Samanani; J. Alcantara; R. Bourgault; K. G. Zulak; P. J. Facchini, *Plant J.* **2006**, *47*, 547–563.
74. D. A. Bird; V. R. Franceschi; P. J. Facchini, *Plant Cell* **2003**, *15*, 2626–2635.
75. M. Weid; J. Ziegler; T. M. Kutchan, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13957–13962.
76. J. E. Page, *Trends Biotechnol.* **2005**, *23*, 331–333.
77. S. Frick; R. Kramell; T. M. Kutchan, *Metab. Eng.* **2007**, *9*, 169–176.
78. K. M. Hawkins; C. D. Smolke, *Nat. Chem. Biol.* **2008**, *4*, 564–573.
79. H. Minami; J.-S. Kim; N. Ikezawa; T. Takemura; T. Katayama; H. Kumagai; F. Sato, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 7393–7398.
80. J. Leonard, *Nat. Prod. Rep.* **1999**, *16*, 319–338.
81. R. Van der Heijden; D. I. Jacobs; W. Snoeijer; D. V. R. Hallard, *Curr. Med. Chem.* **2004**, *11*, 607–628.
82. H. Rischer; M. Oresic; T. Seppanen-Laakso; M. Katajamaa; F. Lammertyn; W. Ardiles-Diaz; M. C. E. Van Montagu; D. Inze; K.-M. Oksman-Caldentey; A. Goossens, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 5614–5619.
83. J. Murata; D. Bienzle; J. E. Brandle; C. W. Sensen; V. De Luca, *FEBS Lett.* **2006**, *580*, 4501–4507.
84. J. Murata; J. Roepke; H. Gordon; De V. Luca, *Plant Cell* **2008**, *20*, 524–542.
85. A. R. Battersby; A. R. Burnett; P. G. Parsons, *Chem. Comm.* **1968**, 1280–1281.
86. A. R. Battersby; A. R. Burnett; P. G. Parsons, *J. Chem. Soc. C* **1969**, 1187–1192.
87. A. R. Battersby; J. C. Byrne; R. S. Kapil; J. A. Martin; T. G. Payne; D. Arigoni; P. Loew, *Chem. Commun.* **1968**, 951–953.
88. A. Contin; R. Van der Heijden; A. W. M. Lefeber; R. Verpoorte, *FEBS Lett.* **1998**, *434*, 413–416.
89. K. Chahed; A. Oudin; N. Guivarc'h; S. Hamdi; J.-C. Chenieux; M. Rideau; M. Clastre, *Plant Physiol. Biochem.* **2000**, *38*, 559–566.
90. B. Veau; M. Courtois; A. Oudin; J.-C. Chenieux; M. Rideau; M. Clastre, *Biochim. Biophys. Acta* **2000**, *1517*, 159–163.
91. G. Collu; N. Unver; A. M. G. Peltenburg-Looman; R. van der Heijden; R. Verpoorte; J. Memelink, *FEBS Lett.* **2001**, *508*, 215–220.
92. G. Collu; A. Alonso Garcia; R. van der Heijden; R. Verpoorte, *Plant Sci.* **2002**, *162*, 165–172.
93. H. P. Vetter; U. Mangold; G. Schroeder; F. J. Marner; D. Werck-Reichhart; J. Schroeder, *Plant Physiol.* **1992**, *100*, 998–1007.
94. S. Irmler; G. Schroder; B. St-Pierre; N. P. Crouch; M. Hotze; J. Schmidt; D. Strack; U. Matern; J. Schroder, *Plant J.* **2000**, *24*, 797–804.
95. V. de Luca; C. Marineau; N. Brisson, *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 2582–2586.
96. P. J. Facchini; K. L. Huber-Allanach; L. W. Tari, *Phytochemistry* **2000**, *54*, 121–138.
97. T. M. Kutchan, *Phytochemistry* **1993**, *32*, 493–506.
98. T. D. McKnight; C. A. Roessner; R. Devagupta; A. I. Scott; C. Nessler, *Nucleic Acids Res.* **1990**, *18*, 4939.

99. T. M. Kutchan; N. Hampp; F. Lottspeich; K. Beyreuther; M. H. Zenk, *FEBS Lett.* **1988**, *237*, 40–44.
100. Y. Yamazaki; H. Sudo; M. Yamazaki; N. Aimi; K. Saito, *Plant Cell Physiol.* **2003**, *44*, 395–403.
101. X. Ma; S. Panjikar; J. Koepke; E. Loris; J. Stockigt, *Plant Cell* **2006**, *18*, 907–920.
102. P. Bernhardt; E. McCoy; S. E. O'Connor, *Chem. Biol.* **2007**, *14*, 888–897.
103. S. Chen; M. C. Galan; C. Coltharp; S. E. O'Connor, *Chem. Biol.* **2006**, *13*, 1137–1141.
104. E. A. Loris; S. Panjikar; M. Ruppert; L. Barleben; M. Unger; H. Schubel; J. Stockigt, *Chem. Biol.* **2007**, *14*, 979–985.
105. V. Brandt; A. Geerlings; M. Tits; C. Delaude; R. Van der Heijden; R. Verpoorte; L. Angenot, *Plant Physiol. Biochem.* **2000**, *38*, 187–192.
106. A. Geerlings; M. M.-L. Ibanez; J. Memelink; R. Van der Heijden; R. Verpoorte, *J. Biol. Chem.* **2000**, *275*, 3051–3056.
107. C. Kan-Fan; H. P. Husson, *J. Chem. Soc. Chem. Commun.* **1979**, 1015–1018.
108. P. Heinstein; G. Hofle; J. Stockigt, *Planta Med.* **1979**, *37*, 349–357.
109. C. Kan-Fan; H. P. Husson, *J. Chem. Soc. Chem. Commun.* **1978**, 618–619.
110. C. Kan-Fan; H. P. Husson, *Tetrahedron Lett.* **1980**, *21*, 1460–1463.
111. M. Lounasmaa; P. Hanhinen, *Heterocycles* **1998**, *48*, 1483–1492.
112. M. Ruppert; X. Ma; J. Stoeckigt, *Curr. Org. Chem.* **2005**, *9*, 1431–1444.
113. J. Brugada; P. Brugada; R. Brugada, *Eur. Heart J.* **2003**, *24*, 1085–1086.
114. D. Schmidt; J. Stockigt, *Planta Med.* **1995**, *61*, 254–258.
115. J. Stockigt, The Alkaloids. In *Biosynthesis in Rauwolfia serpentina*; G. A. Cordell, Ed.; Academic Press: San Diego, CA, 1995; Vol. 47, p 115.
116. A. Pfitzner; J. Stoeckigt, *Phytochemistry* **1982**, *21*, 1585–1588.
117. E. Mattern-Dogru; X. Ma; J. Hartmann; H. Decker; J. Stockigt, *Eur. J. Biochem.* **2002**, *269*, 2889–2896.
118. A. Pfitzner; J. Stoeckigt, *Planta Med.* **1983**, *48*, 221–227.
119. E. Dogru; H. Warzecha; F. Seibel; S. Haebel; F. Lottspeich; J. Stockigt, *Eur. J. Biochem.* **2000**, *267*, 1397–1406.
120. A. Pfitzner; J. Stoeckigt, *Tetrahedron Lett.* **1983**, *24*, 5197–5200.
121. A. Bayer; X. Ma; J. Stockigt, *Bioorg. Med. Chem.* **2004**, *12*, 2787–2795.
122. I. Gerasimenko; X. Ma; Y. Sheludko; R. Mentele; F. Lottspeich; J. Stockigt, *Bioorg. Med. Chem.* **2004**, *12*, 2781–2786.
123. X. Ma; J. Koepke; S. Panjikar; G. Fritzsch; J. Stoeckigt, *J. Biol. Chem.* **2005**, *280*, 13576–13583.
124. H. Falkenhagen; L. Polz; H. Takayama; M. Kitajima; S. Sakai; N. Aimi; J. Stockigt, *Heterocycles* **1995**, *41*, 2683–2690.
125. S. Gao; G. von Schumann; J. Stockigt, *Planta Med.* **2002**, *68*, 906–911.
126. G. von Schumann; S. Gao; J. Stockigt, *Bioorg. Med. Chem.* **2002**, *10*, 1913–1918.
127. M. Ruppert; J. Woll; A. Giritch; E. Genady; X. Ma; J. Stoeckigt, *Planta* **2005**, *222*, 888–898.
128. J. Stockigt; A. Pfitzner; P. I. Keller, *Tetrahedron Lett.* **1983**, *244*, 2485.
129. S. Li; J. Long; Z. Ma; Z. Xu; J. Li; Z. Zhang, *Curr. Med. Res. Opin.* **2004**, *20*, 409–415.
130. L. Costa-Campos; M. Iwu; E. Elisabetsky, *J. Ethnopharmacol.* **2004**, *93*, 307–310.
131. T. Hemscheidt; M. H. Zenk, *Plant Cell Rep.* **1985**, *4*, 216–219.
132. J. Stockigt; T. Hemscheidt; G. Hofle; P. Heinstein; V. Formacek, *Biochemistry* **1983**, *22*, 3448–3452.
133. A. R. Battersby; A. R. Burnett; E. S. Hall; P. G. Parsons, *Chem. Commun.* **1968**, 1582–1583.
134. A. I. Scott; P. C. Cherry; A. A. Qureshi, *J. Am. Chem. Soc.* **1969**, *91*, 4932–4933.
135. A. R. Battersby; E. S. Hall, *Chem. Commun.* **1969**, 793–794.
136. A. R. Battersby; A. R. Burnett; P. G. Parsons, *J. Chem. Soc. C* **1969**, 1193–1200.
137. A. R. Battersby; A. K. Bhatnagar, *Chem. Commun.* **1970**, 193–195.
138. A. R. Battersby; K. H. Gibson, *Chem. Commun.* **1971**, 902–903.
139. A. I. Scott; P. B. Reichardt; M. B. Slaytor; J. G. Sweeny, *Bioorg. Chem.* **1971**, *1*, 157–173.
140. A. I. Scott, *J. Am. Chem. Soc.* **1972**, *94*, 8262.
141. A. I. Scott; C. C. Wei, *J. Am. Chem. Soc.* **1972**, *94*, 8263–8264.
142. A. I. Scott; C. C. Wei, *J. Am. Chem. Soc.* **1972**, *94*, 8264–8265.
143. A. I. Scott; C. C. Wei, *J. Am. Chem. Soc.* **1972**, *94*, 8266–8267.
144. V. De Luca, *Rec. Adv. Phytochem.* **2003**, *37*, 181–202.
145. B. St. Pierre; V. De Luca, *Plant Physiol.* **1995**, *109*, 131–139.
146. G. Schroder; E. Unterbusch; M. Kaltenbach; J. Schmidt; D. Strack; V. de Luca; J. Schroder, *FEBS Lett.* **1999**, *458*, 97–102.
147. D. Levac; J. Murata; W. S. Kim; V. De Luca, *Plant J.* **2008**, *53*, 225–236.
148. M. Dethier; V. de Luca, *Phytochemistry* **1993**, *32*, 673–678.
149. F. Vazquez-Flota; E. De Carolis; A. Alarco; V. De Luca, *Plant Mol. Biol.* **1997**, *34*, 935–948.
150. B. St. Pierre; P. Laflamme; A. Alarco; V. De Luca, *Plant J.* **1998**, *14*, 703–713.
151. M. N. Islam; M. N. Iskander, *Mini Rev. Med. Chem.* **2004**, *4*, 1077–1104.
152. T. Beckers; S. Mahboobi, *Drugs Future* **2003**, *28*, 767–785.
153. A. I. Scott; F. Gueritte; S. L. Lee, *J. Am. Chem. Soc.* **1978**, *100*, 6253–6255.
154. T. Endo; A. Goodbody; J. Vukovic; M. Misawa, *Phytochemistry* **1988**, *27*, 2147–2149.
155. J. L. Smith; E. Amouzou; A. Yamagushi; S. McLean; F. DiCosmo, *Biotechnol. Appl. Bioeng.* **1988**, *10*, 568–575.
156. F. Hillou; M. Costa; I. Almeida; I. Lopes Cardoso; M. Leech; A. Ros Barcelo; M. Sottomayor, Cloning of a Peroxidase Enzyme Involved in the Biosynthesis of Pharmaceutically Aciveterpenoid Indole Alkaloids in *Catharanthus roseus.* In *Proceedings of the VI International Plant Peroxidase Symposium*In *Proceedings of the VI International Plant Peroxidase Symposium*; M. Acosta, J. N. Rodriguez-Lopez, M. A. Pedreno, Eds.; University of Murcia, University of a Coruña: Murcia, Spain, 2002; 152–158.
157. M. Sottomayor; M. Lopez-Serrano; F. DiCosmo; A. Ros Barcelo, *FEBS Lett.* **1998**, *428*, 299–303.
158. M. Sottomayor; F. DiCosmo; A. Ros Barcelo, *Enzyme Microb. Technol.* **1997**, *21*, 543–549.
159. V. De Luca; B. St. Pierre, *Trends Plant Sci.* **2000**, *5*, 168–173.
160. L. H. Stevens; T. J. M. Blom; R. Verpoorte, *Plant Cell Rep.* **1993**, *12*, 573–576.
161. V. de Luca; A. J. Cutler, *Plant Physiol.* **1987**, *85*, 1099–1102.
162. E. de Carolis; F. Chan; J. Balsevich; V. de Luca, *Plant Physiol.* **1990**, *94*, 1323–1329.

163. V. Burlat; A. Oudin; M. Courtois; M. Rideau; B. St-Pierre, *Plant J.* **2004**, *38*, 131–141.
164. J. Murata; V. de Luca, *Plant J.* **2005**, *44*, 581–594.
165. B. St-Pierre; F. Vazquez-Flota; V. de Luca, *Plant Cell* **1999**, *11*, 887–900.
166. C. Canel; M. I. Lopes-Cardoso; S. Whitmer; L. van der Fits; G. Pasquali; R. van der Heijden; J. H. C. Hoge; R. Verpoorte, *Planta* **1998**, *205*, 414–419.
167. S. Di Fiore; V. Hoppmann; R. Fischer; S. Schillberg, *Plant Mol. Biol.* **2004**, *22*, 15–22.
168. T. Ayora-Talavera; J. Chappell; E. Lozoya-Gloria; V. M. Loyola-Vargas, *Appl. Biochem. Biotechnol.* **2002**, *97*, 135–145.
169. C. A. M. Peebles; S.-B. Hong; S. I. Gibson; J. V. Shanks; K.-Y. San, *Biotech. Prog.* **2005**, *21*, 1572–1576.
170. S. Hedhili; V. Courdavault; N. Gigioli-Guivarc'h; P. Gantet, *Phytochem. Rev.* **2007**, *6*, 341–351.
171. E. McCoy; S. E. O'Connor, *J. Am. Chem. Soc.* **2006**, *128*, 14276–14277.
172. W. Runguphan; S. E. O'Connor, *Nat. Chem. Biol.* **2009**, *5*, 151–153.
173. A. Geerlings; F. J. Redondo; A. Contin; J. Memelink; R. van der Heijden; R. Verpoorte, *Appl. Microbiol. Biotechnol.* **2001**, *56*, 420–424.
174. F. L. H. Menke; A. Champion; J. W. Kijne; J. Memelink, *EMBO J.* **1999**, *18*, 4455–4463.
175. B. Pauw; F. A. O. Hilliou; V. S. Martin; G. Chatel; C. J. F. de Wolf; A. Champion; M. Pre; B. van Duijn; J. W. Kijne; L. van der Fits; J. Memelink, *J. Biol. Chem.* **2004**, *279*, 52940–52948.
176. G. Montiel; C. Breton; M. Thiersault; V. Burlat; C. Jay-Allemand; P. Gantet, *Metab. Eng.* **2007**, *9*, 125–132.
177. K. M. Oksman-Caldenty; R. Arroo, Regulation of Tropane Alkaloid Metabolism in Plants and Plant Cell Cultures. In *Metabolic Engineering of Plant Secondary Metabolism*; R. Verpoorte, A. W. Alfermann, Eds.; Kluwer Academic Publishers: The Netherlands, 2000; pp 253–281.
178. N. Hibi; T. Fujita; M. Hatano; T. Hashimoto; Y. Yamada, *Plant Physiol.* **1992**, *100*, 826–835.
179. N. Hibi; S. Higashiguchi; T. Hashimoto; Y. Yamada, *Plant Cell* **1994**, *6*, 723–735.
180. T. Hashimoto; T. Shoji; T. Mihara; H. Oguri; K. Tamaki; K. I. Suzuki; Y. Yamada, *Plant Mol. Biol.* **1998**, *37*, 25–37.
181. K. Suzuki; Y. Yamada; T. Hashimoto, *Plant Cell Physiol.* **1999**, *40*, 289–297.
182. M. Teuber; M. E. Azemi; F. Namjoyan; A.-C. Meier; A. Wodak; W. Brandt; B. Drager, *Plant Mol. Biol.* **2007**, *63*, 787–801.
183. S. C. Haslam; T. W. Young, *Phytochemistry* **1992**, *31*, 4075–4079.
184. H. Kaiser; U. Richter; R. Keiner; A. Brabant; B. Hause; B. Drager, *Planta* **2006**, *225*, 127–137.
185. W. R. McLauchlan; R. A. McKee; D. M. Evans, *Planta* **1993**, *191*, 440–445.
186. W. G. Heim; K. A. Sykes; S. B. Hldreth; J. Sun; R. H. Lu; J. G. Jelesko, *Phytochemistry* **2007**, *68*, 454–463.
187. W. G. Heim; J. G. Jelesko, *Plant Mol. Biol.* **2004**, *56*, 299–308.
188. T. Hashimoto; K. Nakajima; G. Ongena; Y. Yamada, *Plant Physiol.* **1992**, *100*, 836–845.
189. P. Rocha; O. Stenzel; A. Parr; N. J. Walton; P. Christou; B. Drager; M. J. Leech, *Plant Sci.* **2002**, *162*, 905–913.
190. K. Nakajima; A. Yamashita; H. Akama; T. Nakatsu; H. Kato; T. Hashimoto; J. Oda; Y. Yamada, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4876–4881.
191. B. Dräger, *Nat. Prod. Rep.* **2004**, *21*, 211–223.
192. K. Nakajima; H. Kato; J. Oda; Y. Yamada; T. Hashimoto, *J. Biol. Chem.* **1999**, *274*, 16563–16568.
193. A. Yamashita; H. Kato; S. Wakatsuki; T. Tomizaki; T. Nakatsu; K. Nakajima; T. Hashimoto; Y. Yamada; J. Oda, *Biochemistry* **1999**, *38*, 7630–7637.
194. R. Duran-Patron; D. O'Hagan; J. T. G. Hamilton; C. W. Wong, *Phytochemistry* **2000**, *53*, 777–784.
195. A. Lanoue; M. Boitel-Conti; J. C. Portais; J. C. Laberche; J. N. Barbotin; P. Christen; B. Sangwan-Norreel, *J. Nat. Prod.* **2002**, *65*, 1131–1135.
196. S. Oliagnier; E. Kervio; J. Retey, *FEBS Lett.* **1998**, *437*, 309–312.
197. S. Patterson; D. O'Hagan, *Phytochemistry* **2002**, *61*, 323–329.
198. T. Hashimoto; J. Kohno; Y. Yamada, *Phytochemistry* **1989**, *28*, 1077–1082.
199. Y. Yamada; T. Hashimoto, *Proc. Jpn. Acad. B* **1989**, *65*, 156–159.
200. J. Matsuda; S. Okabe; T. Hashimoto; Y. Yamada, *J. Biol. Chem.* **1991**, *266*, 9460–9464.
201. T. Hashimoto; Y. Yamada, *Planta Med.* **1983**, *47*, 195–199.
202. D. J. Yun; T. Hashimoto; Y. Yamada, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 11799–11803.
203. F. Sato; T. Hashimoto; A. Hachiya; K. Tamura; K.-B. Choi; T. Morishige; H. Fujimoto; Y. Yamada, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 367–372.
204. D. O'Hagan; R. J. Robins; M. Wilson; C. W. Wong; M. Berry, *J. Chem. Soc. Perkin Trans. 1* ***1999,*** 2117–2120.
205. H. D. Boswell; B. Drager; W. R. McLauchlan; A. Portseffen; D. J. Robins; R. J. Robins; N. J. Walton, *Phytochemistry* **1999**, *52*, 871–878.
206. A. Crozier; H. Ashihara, *Biochemist* **2006**, 23–26.
207. C. Koshiishi; A. Kato; S. Yama; A. Crozier; H. Ashihara, *FEBS Lett.* **2001**, *499*, 50–54.
208. K. Mizuno; M. Kato; F. Irino; N. Yoneyama; T. Fujimura; H. Ashihara, *FEBS Lett.* **2003**, *547*, 56–60.
209. H. Uefuji; S. Ogita; Y. Yamaguchi; N. Koizumi; H. Sano, *Plant Physiol.* **2003**, *132*, 372–380.
210. S. S. M. Waldhauser; F. M. Gillies; A. Crozier; T. W. Baumann, *Phytochemistry* **1997**, *45*, 1407–1414.
211. G. Stoychev; B. Kierdaszuk; D. Shugar, *Eur. J. Biochem.* **2002**, *269*, 4048–4057.
212. S. Ogita; H. Uefuji; H. Morimoto; H. Sano, *Plant Mol. Biol.* **2004**, *54*, 931–941.
213. M. Kato; K. Mizuno; A. Crozier; T. Fujimura; H. Ashihara, *Nature* **2000**, *406*, 956–957.
214. N. Yoneyama; H. Morimoto; C. X. Ye; H. Ashira; K. Mizuno; M. Kato, *Mol. Gen. Genomics* **2006**, *275*, 125–135.
215. H. Ashira; X. Q. Zheng; R. Katahira; M. Morimoto; S. Ogita; H. Sano, *Phytochemistry* **2006**, *67*, 882–886.
216. S. Ogita; H. Uefuji; Y. Yamaguchi; N. Koizumi; H. Sano, *Nature* **2003**, *423*, 823.
217. K. V. Satyanarayana; V. Kumar; A. Chandrasheker; G. A. Ravishankar, *J. Biotechnol.* **2005**, *119*, 20–25.

Biographical Sketch

Sarah E. O'Connor received a bachelor of science degree in chemistry from the University of Chicago. She received her Ph.D. in chemistry under the direction of Barbara Imperiali at Caltech and MIT, where her thesis work focused on the synthesis and structural analysis of N-linked glycopeptides. She did her postdoctoral research with Chris Walsh at Harvard Medical School where she studied the biosynthesis of epothilone and several other polyketide and peptide natural products. She began her independent research program at MIT where her group is investigating the biosynthesis of plant-derived alkaloid natural products.