

# Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology<sup>1</sup>

Brenda Winkel-Shirley\*

Department of Biology, Virginia Tech, Blacksburg, Virginia 24061-0406

The role of flavonoids as the major red, blue, and purple pigments in plants has gained these secondary products a great deal of attention over the years. From the first description of acid and base effects on plant pigments by Robert Boyle in 1664 to the characterization of structural and regulatory genes in the late 20th century, a wealth of information has been collected on the structures, chemical activities, and biosynthesis of these compounds. Flavonoids constitute a relatively diverse family of aromatic molecules that are derived from Phe and malonyl-coenzyme A (CoA; via the fatty acid pathway). These compounds include six major subgroups that are found in most higher plants: the chalcones, flavones, flavonols, flavandiols, anthocyanins, and condensed tannins (or proanthocyanidins); a seventh group, the aurones, is widespread, but not ubiquitous (Fig. 1). Some plant species also synthesize specialized forms of flavonoids, such as the isoflavonoids that are found in legumes and a small number of nonlegume plants. Similarly, sorghum (*Sorghum bicolor*), maize (*Zea mays*), and gloxinia (*Sinningia cardinalis*) are among the few species known to synthesize 3-deoxyanthocyanins (or phlobaphenes in the polymerized form). The stilbenes, which are closely related to flavonoids, are synthesized by yet another group of unrelated species that includes grape (*Vitis vinifera*), peanut (*Arachis hypogaea*), and pine (*Pinus sylvestris*). Thus, it appears that branches in this pathway have evolved multiple times or been lost from specific plant lineages over the course of evolution.

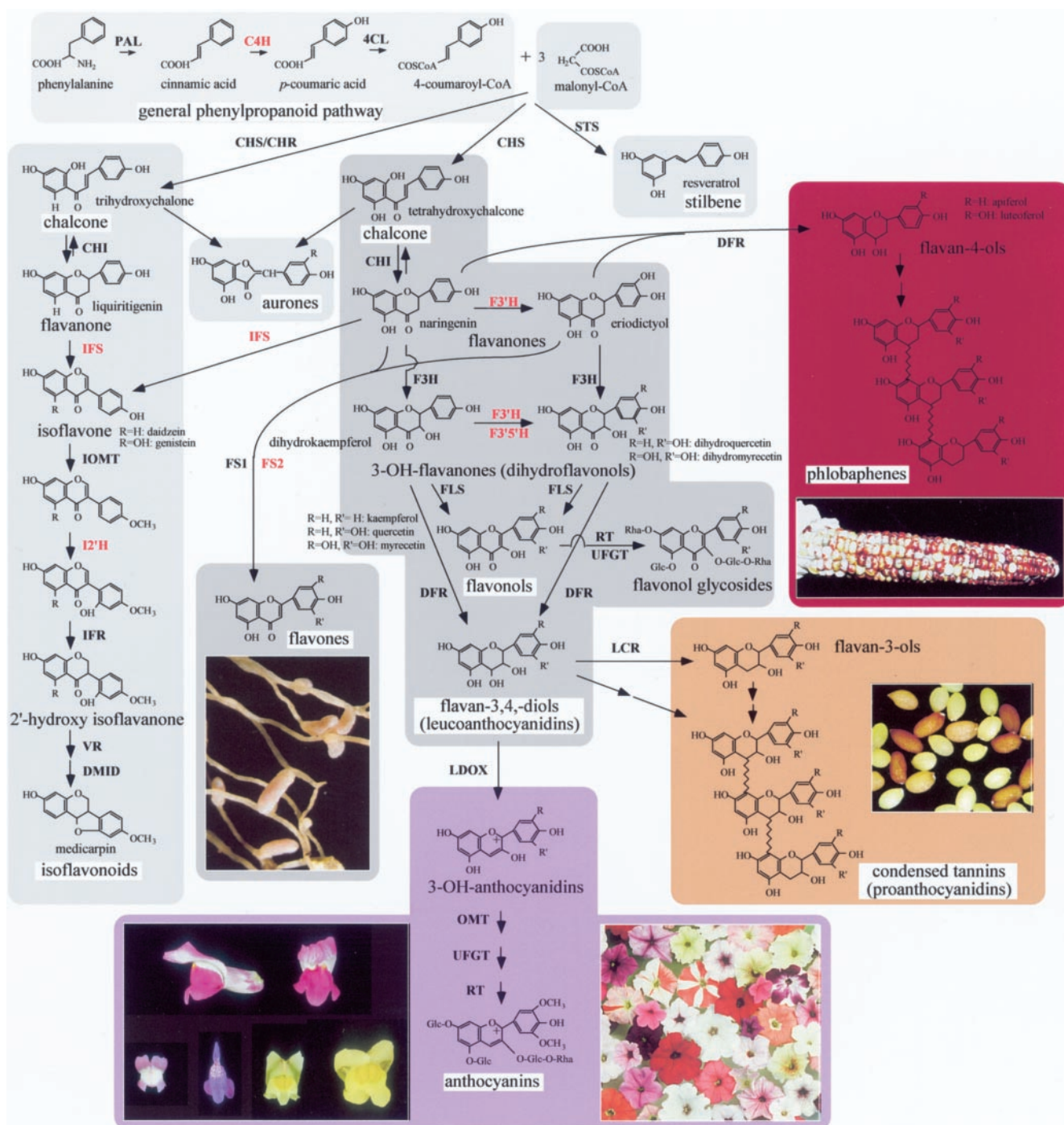
A well-known physiological function of the anthocyanin pigments and flavonol copigments is the recruitment of pollinators and seed dispersers. These compounds also have figured into some of the major scientific breakthroughs of the past 150 years, including Mendel's elucidation of genetics, seed coat color being one of the major characters followed in his experiments with peas (*Pisum sativum*), and McClintock's discovery of transposable elements, which moved in and out of flavonoid biosynthetic genes expressed in maize kernels. Anthocyanins more recently have aided in understanding the phenomenon of cosuppression, particularly in petunia (*Petunia hybrida*). But besides providing beautiful pigmentation in flowers, fruits, seeds, and leaves, flavonoids also

have key roles in signaling between plants and microbes, in male fertility of some species, in defense as antimicrobial agents and feeding deterrents, and in UV protection. The "early" steps in the pathway are found even in the bryophytes (mosses) and it has been suggested that synthesis of flavones, flavanones, and flavonols may have evolved first to provide chemical messengers and then UV sunscreens (Stafford, 1991). Flavonoids also have significant activities when ingested by animals, and there is great interest in their potential health benefits, particularly for compounds such as isoflavonoids, which have been linked to the anticancer benefits of soy-based foods, and the stilbenes in red wine that are believed to contribute to reduced heart disease.

In recent years, much effort has been directed at elucidating the flavonoid biosynthetic pathway from a molecular genetic point of view. Mutants affecting flavonoid synthesis have been isolated in a variety of plant species based on alterations in flower and seed pigmentation. Maize, snapdragon (*Antirrhinum majus*), and petunia were established as the first major experimental models in this system, and work in these species led to the isolation of many flavonoid structural and regulatory genes (for review, see Holton and Cornish, 1995; Mol et al., 1998). Arabidopsis more recently has helped facilitate analysis of the regulation and subcellular organization of the flavonoid pathway. One unique aspect of using Arabidopsis for studying flavonoid biosynthesis is that all but one of the enzymes of central flavonoid metabolism (leading to flavonols and anthocyanins) are encoded by single-copy genes. The exception is flavonol synthase (FLS), which appears to be encoded by six genes, two of which may not be expressed (A. Bandara, D. Owens, and B. Winkel-Shirley, unpublished data). Genetic loci for both structural (Table I) and regulatory genes are scattered across the Arabidopsis genome and have been identified largely on the basis of mutations that abolish or reduce pigmentation in the seed coat. As a result, the loci were named *transparent testa* by Maarten Koornneef (Wageningen Agricultural University, The Netherlands), who isolated many of the first mutants in this class in the 1980s (for review, see Koornneef, 1990). The initial collection of 12 *tt* mutants has been expanded to include 21 members (*tt1-19* plus *ttg1* and *ttg2*), largely through directed searches in the Koornneef lab for new lines producing yellow or pale

<sup>1</sup> This work was supported by the National Science Foundation.

\* E-mail [winkel@vt.edu](mailto:winkel@vt.edu); fax 540-231-9307.



**Figure 1.** Schematic of the major branch pathways of flavonoid biosynthesis, starting with general phenylpropanoid metabolism and leading to the nine major subgroups: the colorless chalcones, aurones, isoflavonoids, flavones, flavonols, and flavandiols (gray boxes), and the anthocyanins, condensed tannins, and phlobaphene pigments (colored boxes). The first committed step is catalyzed by chalcone synthase (CHS), which uses malonyl CoA and 4-coumaroyl CoA as substrates. Only a few examples are shown of the enormous variety of end products that arise through terminal modification by the addition of sugars as well as methyl, ferulate, and other groups. P450 hydroxylases that may function as membrane anchors for multienzyme assemblies are indicated in red. The photographs illustrate the three major classes of pigments in the model plants, snapdragon, *Arabidopsis*, maize, and petunia. Root nodulation by rhizobia, which involves flavone as well as flavanone and isoflavone signal molecules, is also shown, in this case for sweet clover (*Melilotus alba*). Enzyme names are abbreviated as follows: cinnamate-4-hydroxylase (C4H), chalcone isomerase (CHI), chalcone reductase (CHR), chalcone synthase (CHS), 4-coumaroyl:CoA-ligase (4CL), dihydroflavanol 4-reductase (DFR), 7,2'-dihydroxy, 4'-methoxyisoflavanol dehydratase (DMID), flavanone 3-hydroxylase (F3H), flavonoid 3' hydroxylase (F3'H) or (Legend continues on facing page.)

**Table 1.** Genetic loci for cloned flavonoid enzymes in *Arabidopsis*<sup>a</sup>

Enzyme	Locus	Chromosome	Map Position <sup>b</sup>
CHS	<i>tt4</i>	5	7,050 kb (MAC12)
CHI	<i>tt5</i>	3	21,000 kb (T15C9)
F3H	<i>tt6</i>	3	19,600 kb (F24M12)
F3'H	<i>tt7</i>	5	4,400 kb (F13G24)
FLS	<i>fls1::En</i> <sup>c</sup>	5	FLS1: 4,700 kb (MAH20) FLS2-5: 32,150 kb (MBK5) FLS6: 24,350 kb (MRH10)
DFR	<i>tt3</i>	5	23,800 kb (MJB21)
LDOX	<i>tt19</i> <sup>d</sup>	4	16,900 kb (F7H19)
LCR	<i>ban</i> , <i>ast</i> <sup>d</sup>	1	26,800 kb (T13M11)

<sup>a</sup> Similar information for maize, petunia, and snapdragon is described by Holton and Cornish (1995). <sup>b</sup> Based on the AGI map, 11/12/00; numbers in parentheses refer to P1 or bacterial artificial chromosome clones on which these sequences reside. <sup>c</sup> Transposon-tagged mutant for FLS1 (Wisman et al., 1998). <sup>d</sup> A. Tanaka (personal communication).

brown seeds, and indirectly, by Loic Lepiniec and coworkers at the Institut National de la Recherche Agronomique (Versailles, France), in screens for plants exhibiting reduced seed dormancy. Transposon and activator tagging have been used to isolate additional mutations in genes either directly or indirectly involved in flavonoid biosynthesis (Wisman et al., 1998; Kubo et al., 1999; Borevitz et al., 2000). As a result most of the structural genes, as well as a number of regulatory genes, have now been correlated with specific mutant loci in *Arabidopsis*. Although *Arabidopsis* does not appear to use flavonoids in all of the same ways as some other species (for example, in defense or for male fertility), the *Arabidopsis* mutants are helping to define a role for these compounds in essential processes such as UV protection (Li et al., 1993; Landry et al., 1995) and the regulation of auxin transport (Murphy et al., 2000; Brown et al., 2001).

## NEW GENES FOR ENGINEERING FLAVONOID METABOLISM

Several important new genes required for flavonoid biosynthesis have been characterized in a variety of plant species over the past few years, including some with direct practical applications. One of these is the *Arabidopsis* *BANYULS* gene, which encodes a DFR-like protein that may be an LCR that catalyzes an early step in condensed tannin biosyn-

thesis (Fig. 1; Devic et al., 1999). A locus identified independently as *anthocyanin spotted testa* (*ast*; Tanaka et al., 1997), with a very similar mutant phenotype, now appears to be identical to *BAN* (A. Tanaka, personal communication). Controlling condensed tannin levels in forage crops has long been of interest, both to improve nutritional value either by increasing or decreasing endogenous levels and to provide amounts sufficient to control pasture bloat. Some success has been achieved by modulating late steps in the central flavonoid pathway, such as the DFR reaction, for which cloned genes were previously available (Morris and Robbins, 1997). The LCR gene may provide an opportunity to direct metabolic engineering efforts more specifically to the proanthocyanidin branch pathway.

Several breakthroughs have been made in the isoflavonoid pathway, including the isolation of the first IFS genes. Biochemical and genetic data have long suggested that this enzyme is a member of the cytochrome P450 oxygenase family of enzymes. This was confirmed by Shin-ichi Ayabe's laboratory (Nihon University, Fujisawa, Kanagawa, Japan) with the isolation of IFS from a licorice (*Glycyrrhiza echinata*) cell line that produces isoflavonoids upon elicitation (Akashi et al., 1999). At the same time, Richard Dixon's group (The Noble Foundation, Ardmore, OK) identified an IFS gene from soybean (*Glycine max*) by functional screening of candidate P450 cDNAs in insect cells (Steele et al., 1999), whereas a group at DuPont Wilmington, DE identified the same gene as well as a second IFS using a similar screen in yeast (*Saccharomyces cerevisiae*; Jung et al., 2000). Both isoforms of soybean IFS appear to be able to use both liquiritigenin and naringenin as substrates to produce genistein or daidzein, respectively (Fig. 1), although naringenin is used less efficiently. The DuPont group showed that soybean IFS1 can function to convert naringenin to genistein in transgenic *Arabidopsis* and, more recently, in tobacco (*Nicotiana tabacum*) and maize (Yu et al., 2000). They have also shown that introduction of IFS1 together with chalcone reductase, which provides the additional substrate, liquiritigenin, results in the synthesis of daidzein in maize. Nancy Paiva's laboratory (The Noble Foundation) is attempting to express VR in tobacco plants; this enzyme is one of several that will be required to engineer production of the isoflavonoid, medicarpin, the major phytoalexin produced by alfalfa (*Medicago sativa*) in response to fungal pathogens (Fig. 1; Watson and Paiva, 2000). A cDNA en-

**Figure 1.** (Legend continued from facing page.)

flavonoid 3'5' hydroxylase (F3'5'H), isoflavone O-methyltransferase (IOMT), isoflavone reductase (IFR), isoflavone 2'-hydroxylase (I2'H), isoflavone synthase (IFS), leucoanthocyanidin dioxygenase (LDOX), leucoanthocyanidin reductase (LCR), O-methyltransferase (OMT), Phe ammonia-lyase (PAL), rhamnosyl transferase (RT), stilbene synthase (STS), UDPG-flavonoid glucosyl transferase (UFGT), and vestitone reductase (VR). Photographs are courtesy of Cathie Martin (John Innes Centre, Norwich, UK; *Antirrhinum*), Francesca Quattrocchio (Free University, Amsterdam; petunia), Erich Grotewold (Ohio State University, Columbus; maize), and Yimei Lin and Ann Hirsch (University of California, Los Angeles; sweet clover).

coding the cytochrome P450 protein, I2'H, another enzyme required for medicarpin biosynthesis (Fig. 1), has also been isolated from licorice (Akashi et al., 1998). Together, these advances are laying the foundation for engineering isoflavonoid biosynthesis for agronomic and nutritional enhancement of a wide variety of crop plants that do not normally synthesize these compounds. There may also be more immediately feasible applications in the engineering of legumes, for example, to improve the palatability of soy milk by down-regulating isoflavonoid synthesis in soybean seeds.

Efforts to engineer flower color have also led to some interesting developments in the last few years. The hydroxylation pattern of the B ring of anthocyanins is a major determinant of the color of these pigments. All flavonoids carry a hydroxyl group at the 4' position, including the pink-to-red cyanidin-based pigments. Hydroxylation at two variable positions is controlled by the P450 enzymes, F3'H, which leads to brick-red to orange pelargonidin-based pigments, and F3'5'H, which is required for synthesis of purple and blue delphinidin-based pigments (Fig. 1). Several years ago, workers at Florigene isolated two F3'5'H genes from petunia based on sequence homology to other p450s, a pattern of high-level expression in flowers, and correlation with the *Hf1* and *Hf2* loci (Holton et al., 1993). Cloning of the first F3'H gene took a bit longer, but a petunia gene was eventually isolated using a similar approach (Brugliera et al., 1999). Together with Chris Cobbett (The University of Melbourne, Parkville, Victoria, Australia), this group also identified the F3'H gene in Arabidopsis by chromosome walking to the *tt7* locus (C. Cobbett, personal communication); the same gene has been identified independently based on information from the Arabidopsis Genome Project (Schoenbohm et al., 2000; Saslowsky and Winkel-Shirley, 2001). It is unfortunate that these genes are not, by themselves, sufficient for engineering altered flower color in horticulturally important species, for example, by overexpression in roses (*Rosa* spp.) and carnations (*Dianthus caryophyllus*) that normally lack F3'5'H activity and therefore do not produce blue or purple pigments (Brugliera et al., 2000). However, it has been shown that a specific cytochrome b5 is required for maximal activity of the petunia F3'5'H enzyme, presumably acting as an alternative electron donor to the NADPH:cytochrome P450 reductase that is generally associated with cytochrome P450 proteins (de Vetten et al., 1999). It is remarkable that the group at Florigene recently reported that when the petunia F3'5'H and *cyt b5* genes are introduced together into carnations, flower color is transformed from red to a deep purple (Brugliera et al., 2000). It appears that the long sought-after blue rose may be close at hand.

Advances are also being made in understanding the regulation of flavonoid biosynthesis, particularly as a result of molecular genetic approaches such as

transposon tagging and positional cloning. This has led to the identification of a number of novel regulatory proteins that are beginning to fill in the void between signals that induce the pathway and well-known flavonoid regulators such as the myb domain and basic helix-loop-helix transcription factors (R, B, C1, and P) of maize. In particular, progress has been made in characterizing genes that regulate expression of "late" pathway enzymes, which are specific to proanthocyanidin and anthocyanin biosynthesis. Examples include a new class of flavonoid regulatory proteins, defined by AN11 in petunia (de Vetten et al., 1997) and TTG1 in Arabidopsis (Walker et al., 1999), that contain WD40 repeats and are related to the  $\beta$ -subunits of heterotrimeric G proteins. AN11 appears to be a cytoplasmic protein that regulates expression of the recently cloned myb domain-containing protein, AN2 (Quattrocchio et al., 1999). TTG1 differs from AN11 in several ways, including its role not only in flavonoid synthesis but in trichome development, root epidermal cell patterning, and the production of seed mucilage. A related gene, MP1, has recently been isolated from maize (E. Grotewold, personal communication), which may shed some light on the evolution and function of this new class of regulatory proteins. In addition, a MYB regulator of phenylpropanoid metabolism, *PAP1*, has been cloned in Arabidopsis, which, when overexpressed, results in intensely purple flowers (Borevitz et al., 2000). In addition to providing insights into coordinate regulation of phenylpropanoid metabolism, the purple-flowered lines may provide a new inroad into identifying Arabidopsis genes that function beyond the branch to condensed tannins, a limitation of cloning genes based on the *tt* phenotype.

The Arabidopsis *ttg2* locus, which has a similar pleiotropic phenotype to *ttg1*, has identified yet another novel class of flavonoid regulatory factors, a member of the WRKY family that contains two SPF1 zinc finger-like domains (Johnson and Smyth, 1998). Thus, the overlapping regulatory pathways controlling flavonoid synthesis and trichome development involves at least two components. This type of overlap has been observed so far in only one other species, *Matthiola incana*, a close relative of Arabidopsis (Heller et al., 1985). It will be interesting to discover whether homologs of *ttg2* are involved in regulating flavonoid gene expression in species like petunia and maize. Two additional transcription factors, identified by the petunia regulatory locus, *an1* (Spelt et al., 2000), and the Arabidopsis *tt8* locus (Nesi et al., 2000), encode new members of the basic helix-loop-helix family of transcription factors and have similarity to R1 in maize and DELILAH in snapdragon. AN1 directly activates expression of a DFR gene as well as an as-yet-uncharacterized myb-domain protein, whereas TT8 is required for expression of DFR and BAN, suggesting that these may be homologous proteins, although there is reason from both se-

quence and experimental data to believe that they are not orthologs of *r1* and *del* (Spelt et al., 2000).

Further characterization of genes identified by regulatory loci such as petunia *an4* and *jaf13*, maize *pac1*, and Arabidopsis *tt1*, *tt16*, *an11*, and *icx1* (Table I), as well as continued analysis of the differences in flavonoid regulatory mechanisms among plant species, should help fill in some of the many missing pieces that still remain in the flavonoid gene regulation puzzle. The ongoing characterization of the activity of regulatory molecules within species is another important area of investigation. For example, Grote-wold et al. (2000) have recently shown that the differential interaction of maize P and C1 with R is mediated by a small number of specific residues in the P and C1 myb domains, thereby explaining an essential aspect of combinatorial transcriptional regulation in this system.

Genomics approaches are yielding promise for helping complete the cast of players, both structural and regulatory, involved in flavonoid biosynthesis and for making connections within the larger context of plant metabolism. Two excellent examples have already been published. In one case, researchers at Pioneer Hi-Bred (Johnston, IA) and Curagen (New Haven, CT) showed that genes exhibiting altered expression in maize cell lines overproducing flavonoid pathway activators included both known maize flavonoid genes as well as numerous novel sequences (Bruce et al., 2000). In the second example, Steve Kay and his colleagues (Scripps Research Institute, La Jolla, CA) have demonstrated coordinate circadian-controlled expression of both known and candidate phenylpropanoid pathway genes and implicated PAP1 as the master regulator of this system (Harmer et al., 2000). Together with genetic and biochemical approaches, these efforts are rapidly adding to our understanding of how flavonoid biosynthesis is controlled and how this information may be used to engineer flavonoid metabolism in diverse plant species.

#### A MODEL FOR SUBCELLULAR ORGANIZATION OF METABOLISM

The question of the intracellular localization and organization of flavonoid enzymes was first raised by Helen Stafford more than 25 years ago, together with the suggestion that the enzymes of general phenylpropanoid, sinapate, lignin, and flavonoid biosynthesis were likely to function as multienzyme complexes (Stafford, 1974). The concept that these pathways may be organized as enzyme complexes that facilitate the direct transfer, or channeling, of intermediates between active sites is compelling for a number of reasons. For example, there is competition for substrates at the numerous branch points within these pathways, the intermediates are highly reactive and potentially toxic, and the overall concentrations

of these compounds appears to be extremely low. Moreover, there is the need for these pathways to respond quickly to external and internal signals to change the amounts and/or types of end products that are synthesized. The first evidence to support this hypothesis came largely from Geza Hrazdina's group (Cornell University, Geneva, NY), who used cell fractionation, ultracentrifugation, and gel-filtration experiments to show that PAL, CHS, and UFGT were located in the cytosol, loosely associated with the cytoplasmic face of the endoplasmic reticulum (ER; for review, see Hrazdina, 1992). This group also published data from immunolocalization experiments showing an association of CHS with the cytoplasmic face of the rough ER (rER), but not with nuclei, plastids, mitochondria, Golgi, or tonoplasts, in buckwheat (*Fagopyrum esculentum*) hypocotyls. The results of this work gave rise to a model for phenylpropanoid and flavonoid synthesis involving a complex of membrane-associated, linear assembly of enzymes that is organized through weak interactions with membrane proteins that include C4H and F3'H.

Additional evidence has now been obtained in support of a membrane-associated flavonoid enzyme complex. Co-immunoprecipitation, affinity chromatography, and two-hybrid experiments indicate that there are direct associations between CHS, CHI, F3H, and DFR in Arabidopsis (Burbulis and Winkel-Shirley, 1999). However, the data do not necessarily point to a linear assembly of enzymes, but to a globular arrangement in which CHS contacts not only CHI, the next enzyme in the pathway, and also F3H and DFR. In addition, the association of CHS with the rER, first described by Hrazdina's group, has now also been observed in Arabidopsis, together with colocalization of this enzyme with CHI, despite the fact that both are typical "soluble" enzymes (Saslowky and Winkel-Shirley, 2001). Moreover, a mutation in the Arabidopsis F3'H gene that deletes most of the cytoplasmic domain of this P450 enzyme results in altered localization of CHS and CHI, indicating that F3'H may function as part of a membrane anchor for other enzymes of the flavonoid pathway, as first suggested by Hrazdina.

There have also been indications that the isoflavonoid branch pathway exists as an enzyme complex. This pathway actually contains two cytochrome P450 oxygenases that could function as membrane anchors, IFS and I2'H (Fig. 1). Hrazdina (1992) first reported evidence for the association of an enzyme from this pathway, isoflavone reductase, with the ER in cell fractionation experiments. Feeding experiments with radiolabeled precursors in elicitor-treated alfalfa seedlings, together with evidence for in vitro and in vivo differences in the site specificities for methylation, have provided evidence that isoflavone O-methyltransferase is part of a metabolic channel (Dixon et al., 1998; He and Dixon, 2000). It is interesting that this enzyme functions between the

two P450 proteins, IFS and I2'H, in isoflavonoid biosynthesis.

These recent findings, together with evidence for channeling between PAL and C4H in the general phenylpropanoid pathway (Czichi and Kindl, 1977; Hrazdina and Wagner, 1985; Rasmussen and Dixon, 1999), indicate that the organization of these systems may become an important consideration in understanding how plant metabolism is regulated. Although the flavonoid pathway is likely to be an example of a dynamic, rather than a stable, enzyme complex, it is also possible that this organization could complicate metabolic engineering by limiting the access of intermediates to introduced enzymes (Dixon and Steele, 1999), while at the same time providing opportunities for the efficient redirection of flux into existing or introduced branch pathways. A great deal of work is clearly still needed in this area, including efforts to define protein interaction domains, to determine whether plant cells contain single branched complexes or a variety of complexes dedicated to the production of specific products, and to characterize changes in the organization of enzyme systems in response to developmental and environmental cues.

New information is also emerging regarding the transport of flavonoids from the site of synthesis in the cytoplasm to final destinations in the vacuole or cell wall. Transport of pigments to the vacuole in the maize aleurone and petunia flowers and perhaps also in soybean, requires both a glutathione *S*-transferase (GST) and a glutathione pump belonging to the ATP-binding cassette (ABC) family of transporters (Marrs et al., 1995; Alfenito et al., 1998). The GST enzymes, AN9 in petunia and BZ2 in maize, are widely divergent and yet are able to complement deficiencies in vacuolar sequestration of flavonoids in a variety of plant species. Mueller et al. (2000) have recently suggested that AN9 and BZ2 may function as "escort" proteins without actually catalyzing GSH conjugate formation and that transport is accomplished by a cotransport mechanism with reduced GSH, analogous to the transport of vincristine in the liver. A somewhat different story is emerging in Arabidopsis, where cloning and characterization of the *TT12* gene has provided evidence that transport of proanthocyanidin precursors into the vacuole in the seed coat endothelium involves a protein belonging to the multidrug and toxic compound extrusion family of secondary transporters in this species (Debeaujon et al., 2001). Moreover, Erich Grotewold's group has new evidence from fluorescent microscopic analysis of maize Black Mexican Sweet cells transformed with *P*, a regulator of 3-deoxy flavonoid synthesis (Fig. 1), for the transport of ER-derived vesicles containing yellow or green fluorescent compounds to the vacuole and cell wall, respectively (for review, see Grotewold, 2001). These are reminiscent of the subcellular inclusions containing 3-deoxyanthocyanidins that accumu-

late in infected sorghum cells (Snyder and Nicholson, 1990). Evidence for colocalization of flavonoid enzymes in electron-dense regions in Arabidopsis root cortex cells (Saslowsky and Winkel-Shirley, 2001) may be related to this transport phenomenon. Therefore, between the identification of genes that are essential for this process and the continued analysis of the cell biology of the system, elucidation of the molecular mechanisms underlying the deposition of different flavonoid compounds in various parts of the cell may be on the horizon. This is an issue of central importance in understanding how flavonoid biosynthesis is controlled and may provide additional insights into engineering this metabolic pathway.

### STRUCTURAL ANALYSIS OF FLAVONOID ENZYMES

The elucidation of the structures of CHS and CHI from alfalfa by Joseph Noel (The Salk Institute, La Jolla, CA) and Richard Dixon's groups is one of the most exciting recent developments in understanding the flavonoid pathway in three dimensions (Ferrer et al., 1999; Jez et al., 2000b). In addition to providing new information about the evolution of flavonoid synthesis, as discussed above, analyses of the structures of CHS and the related 2-pyrone synthase (Jez et al., 2000a) are providing a great deal of information regarding the reaction mechanisms of plant polyketide synthesis. This is opening possibilities for engineering these enzymes to produce new products, similar to what has been done with the bacterial polyketide synthases. It is remarkable that homology modeling of the active sites can be used to predict possible catalytic activities of CHS-like enzymes in at least some cases (for review, see Dixon and Steele, 1999; Schröder, 1999). Modeling of the Arabidopsis CHS enzyme has also provided insights into how specific mutations reduce activity, not only by disrupting the active site, but also by destabilizing the protein or interfering with dimerization (Saslowsky et al., 2000). CHI, on the other hand, has a three-dimensional structure and enzyme activity that are unique to plants. Solving the crystal structure has given the first information on how the alfalfa enzyme recognizes its substrates, naringenin chalcone and 6'-deoxychalcone, and catalyzes the stereospecific cyclization of chalcones. The structure also suggests residues that may be important in determining the different substrate specificities of CHI enzymes in legumes and other plant species. Efforts are under way to solve the structures of Arabidopsis flavonoid enzymes to facilitate the design of experiments to define the interaction interfaces of these proteins in the apparent enzyme complex. It is clear that the structures that have already been solved are providing long-awaited information for studying a variety of different aspects of flavonoid biosynthesis, from evolution to enzymology and subcellular organization.

## EVOLUTION OF FLAVONOID GENES

The ubiquity of flavonoid biosynthesis among plants has long raised questions about the evolution of the pathway and its various structural and regulatory components. Biochemical data first prompted the speculation that this pathway had been derived from primary metabolism. In fact, the majority of the enzymes of flavonoid biosynthesis are members of three classes of enzymes found in all organisms: the oxoglutarate-dependent dioxygenases (F3H, flavonol synthase, FSI, and LDOX), NADPH-dependent reductases (DFR, LCR, IFR, and VR), and cytochrome P450 hydroxylases (F3'H, F3'5'H, FSII, IFS, and I2'H; Fig. 1; Dixon and Steele, 1999). CHS and CHI, on the other hand, appear to have a more limited ancestry. CHI, in particular, appears to be unique to plants in both sequence and three-dimensional structure (Jez et al., 2000b). CHS is a member of the plant polyketide synthase superfamily, which also includes STS, acridone synthase, pyrone synthase, bibenzyl synthase, and *p*-coumaroyltriacytic acid synthase. This family of enzymes uses similar reaction mechanisms and similar or related substrates to produce a wide variety of secondary products. Although these enzymes are not related to the bacterial or fungal polyketide synthases, other related proteins are present in bacteria. In addition, recent work has uncovered a sequence similarities with plant  $\beta$ -ketoacyl-CoA synthases, including the enzyme identified by *FIDDLEHEAD* in Arabidopsis, which catalyze the first step in long-chain fatty acid biosynthesis (Yephremov et al., 1999; Pruitt et al., 2000). Studies on the CHS gene family in morning glory (*Ipomoea purpurea*) point to recurrent gene duplication and specialization of this enzyme over the course of evolution (Durbin et al., 2000). It has also been suggested that STS has evolved from CHS multiple independent times in different plant species, based on the results of mutational and structural analyses and the fact that STS is found in a limited number of unrelated plant species (for review, see Schröder, 1997). Somewhat different evidence for recurrent evolution of specific enzyme activities comes from flavone synthase, which is present as a dioxygenase in parsley (FSI) and a P450 monooxygenase in snapdragon (FSII; Stafford, 1990). In addition, some of the proteins of the flavonoid pathway may have evolved significantly different functions from the ancestral enzymes, as suggested for GSTs, which may act in flavonoid transport as cytoplasmic "escort" proteins rather than directly conjugating these compounds to glutathione (Mueller et al., 2000). It is interesting to note that the upstream genes in the flavonoid pathway appear to have evolved more slowly than the downstream genes (Rausher et al., 1999). The rapid accumulation of genome sequence and protein structure information should shed additional light on the origins of the flavonoid pathway and should also provide insights into structure-function relation-

ships, such as the co-evolution of domains in enzymes that interact to control flux at major branch points.

## PERSPECTIVES

The flavonoid biosynthetic pathway has been one of the most intensively studied metabolic systems in plants. As with any good model, each new piece of information appears to raise a number of unanticipated and intriguing questions. At the same time, new tools are providing the opportunity to consider flavonoid biosynthesis, not as an assemblage of independent components, but as part of a large, complex, and tightly orchestrated metabolic network. The ability to now consider flavonoid enzymes, for the very first time, in three dimensions and to examine the interdependence of the pathways of secondary metabolism using genomic, proteomic, and metabolic profiling methods are likely to move us much more rapidly toward this end. The development of genomics datasets for organisms that include Arabidopsis, the moss *Physcomitrella patens* (<http://www.moss.leeds.ac.uk/>), and the model legume *Medicago truncatula* (<http://www.noble.org/medicago/>), are also offering opportunities to examine this metabolic model system from entirely new perspectives. It is quite clear that, even for much-studied "old" pathways like flavonoid biosynthesis, these are exciting times.

## ACKNOWLEDGMENTS

The author gratefully acknowledges the members of her laboratory, Dr. Jim Westwood, and two anonymous reviewers for helpful comments on the manuscript. Many thanks are also extended to Chris Cobbett, Erich Grotewold, Atsushi Tanaka, and Gloria Muday for sharing information prior to publication and to Douglas Cook, Erich Grotewold, Ann Hirsch, Cathie Martin, and Francesca Quattrocchio for providing photographs.

Received November 20, 2000; accepted March 19, 2001.

## LITERATURE CITED

- Akashi T, Aoki T, Ayabe S (1998) CYP81E1, a cytochrome P450 cDNA of licorice (*Glycyrrhiza echinata* L.), encodes isoflavone 2'-hydroxylase. *Biochem Biophys Res Comm* **251**: 67–70
- Akashi T, Aoki T, Ayabe S (1999) Cloning and functional expression of a cytochrome P450 cDNA encoding 2-hydroxyisoflavanone synthase involved in biosynthesis of the isoflavonoid skeleton in licorice. *Plant Physiol* **121**: 821–828
- Alfenito MR, Souer E, Goodman CD, Buell R, Mol J, Koes R, Walbot V (1998) Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases. *Plant Cell* **10**: 1135–1149

- Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C** (2000) Activation tagging identifies a conserved MYB regulatory of phenylpropanoid biosynthesis. *Plant Cell* **12**: 2383–2393
- Brown DE, Rashotte AM, Murphy AS, Normanly J, Tague BW, Peer WA, Taiz L, Muday GK** (2001) Flavonoids act as negative regulators of auxin transport in vivo in *Arabidopsis*. *Plant Physiol* **126**: 524–535
- Bruce W, Folkerts O, Garnaat C, Crasta O, Roth B, Bowen B** (2000) Expression profiling of the maize flavonoid pathway genes controlled by estradiol-inducible transcription factors CRC and P. *Plant Cell* **12**: 65–79
- Brugliera F, Barri-Rewell G, Holton TA, Mason JG** (1999) Isolation and characterization of a flavonoid 3'-hydroxylase cDNA clone corresponding to the H1 locus of *Petunia hybrida*. *Plant J* **19**: 441–451
- Brugliera F, Tull D, Holton TA, Karan M, Treloar N, Simpson K, Skurczynska J, Mason JG** (2000) Introduction of a cytochrome b5 enhances the activity of flavonoid 3'5' hydroxylase (a cytochrome P450) in transgenic carnation. Sixth International Congress of Plant Molecular Biology. University of Laval, Quebec, pp S6–S8
- Burbulis IE, Winkel-Shirley B** (1999) Interactions among enzymes of the *Arabidopsis* flavonoid biosynthetic pathway. *Proc Natl Acad Sci USA* **96**: 12929–12934
- Czichi U, Kindl H** (1977) Phenylalanine ammonia-lyase and cinnamic acid hydroxylase as assembled consecutive enzymes on microsomal membranes of cucumber cotyledons: cooperation and subcellular distribution. *Planta* **134**: 133–143
- de Vetten N, Quattrocchio F, Mol J, Koes R** (1997) The *an11* locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. *Genes Dev* **11**: 1422–1434
- Debeaujon I, Peeters AJM, Léon-Kloosterziel KM, Koornneef M** (2001) The *TRANSPARENT TESTA12* gene of *Arabidopsis* encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium. *Plant Cell* **13**: 853–872
- de Vetten N, ter Horst J, van Schaik H-P, de Boer A, Mol J, Koes R** (1999) A cytochrome *b5* is required for full activity of flavonoid 3',5'-hydroxylase, a cytochrome P450 involved in the formation of blue flowers. *Proc Natl Acad Sci USA* **96**: 778–783
- Devic M, Guillemot J, Debeaujon I, Bechtold N, Bensaude E, Koornneef M, Pelletier G, Delseny M** (1999) The *BANYULS* gene encodes a DFR-like protein and is a marker of early seed coat development. *Plant J* **19**: 387–398
- Dixon RA, Howles PA, Lamb C, He X-Z, Reddy JT** (1998) Prospects for the metabolic engineering of bioactive flavonoids and related phenylpropanoid compounds. In JA Manthey, BS Buslig, eds, *Flavonoids in the Living System*. Plenum Press, New York, pp 55–66
- Dixon RA, Steele CL** (1999) Flavonoids and isoflavonoids: a gold mine for metabolic engineering. *Trends Plant Sci* **4**: 394–400
- Durbin ML, McCaig B, Clegg MT** (2000) Molecular evolution of the chalcone synthase multigene family in the morning glory genome. *Plant Mol Biol* **42**: 79–92
- Ferrer J-L, Jez JM, Bowman ME, Dixon RA, Noel JP** (1999) Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat Struct Biol* **6**: 775–784
- Grotewold E** (2001) Subcellular trafficking of phytochemicals. *Rec Res Dev Plant Physiol* **2**: 31–48
- Grotewold E, Sainz MB, Tagliani L, Hernandez JM, Bowen B, Chandler VL** (2000) Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. *Proc Natl Acad Sci USA* **97**: 13579–13584
- Harmer SL, Hogenesch JB, Straume M, Chang H-S, Han B, Zhu T, Wang X, Kreps JA, Kay SA** (2000) Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* **290**: 2110–2113
- He X-Z, Dixon RA** (2000) Genetic manipulation of isoflavone 7-O-methyltransferase enhances biosynthesis of 4'-O-methylated isoflavonoid phytoalexins and disease resistance in alfalfa. *Plant Cell* **12**: 1689–1702
- Heller W, Forkmann G, Britsch L, Griseback H** (1985) Enzymatic reduction of (+)-dihydroflavonols to flavan-3,4-cis-diols with flower extracts from *Matthiola incana* and its role in anthocyanin biosynthesis. *Planta* **165**: 284–287
- Holton TA, Brugliera F, Lester DR, Tanaka Y, Hyland CD, Menting JGT, Lu C-Y, Farcy E, Stevenson TW, Cornish EC** (1993) Cloning and expression of cytochrome P450 genes controlling flower color. *Nature* **366**: 276–279
- Holton TA, Cornish EC** (1995) Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* **7**: 1071–1083
- Hrazdina G** (1992) Compartmentation in aromatic metabolism. In HA Stafford, RK Ibrahim, eds, *Recent Advances in Phytochemistry*. Plenum Press, New York, pp 1–23
- Hrazdina G, Wagner GJ** (1985) Metabolic pathways as enzyme complexes: evidence for the synthesis of phenylpropanoids and flavonoids on membrane associated enzyme complexes. *Arch Biochem Biophys* **237**: 88–100
- Jez JM, Austin MB, Ferrer J-L, Bowman ME, Schröder J, Noel JP** (2000a) Structural control of polyketide formation in plant-specific polyketide synthases. *Chem Biol* **7**: 919–930
- Jez JM, Bowman ME, Dixon RA, Noel JP** (2000b) Structure and mechanism of chalcone isomerase: an evolutionarily unique enzyme in plants. *Nat Struct Biol* **7**: 786–791
- Johnson CS, Smyth DR** (1998) The *TTG2* gene of *Arabidopsis* encodes a WRKY family transcription factor that regulates trichome development and the production of pigment and mucilage in seed coats. 9th International Conference on *Arabidopsis* Research. University of Wisconsin, Madison, p 186
- Jung W, Yu O, Sze-Mei CL, O'Keefe DP, Odell J, Fader G, McGonigle B** (2000) Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes. *Nat Biotechnol* **18**: 208–213
- Koornneef M** (1990) Mutations affecting the testa color in *Arabidopsis*. *Arabidopsis Inf Serv* **28**: 1–4
- Kubo H, Peeters AJ, Aarts MG, Pereira A, Koornneef M** (1999) *ANTHOCYANINLESS2*, a homeobox gene affect-



- ing anthocyanin distribution and root development in *Arabidopsis*. *Plant Cell* **11**: 1217–1226
- Landry LG, Chapple CCS, Last RL** (1995) *Arabidopsis* mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage. *Plant Physiol* **109**: 1159–1166
- Li J, Ou-Lee T-M, Raba R, Amundson RG, Last RL** (1993) *Arabidopsis* flavonoid mutants are hypersensitive to UV-B irradiation. *Plant Cell* **5**: 171–179
- Marrs KA, Alfenito MR, Lloyd AM, Walbot V** (1995) A glutathione *S*-transferase involved in vacuolar transfer encoded by the maize gene *Bronze-2*. *Nature* **375**: 397–400
- Mol J, Grotewold E, Koes R** (1998) How genes paint flowers and seeds. *Trends Plant Sci* **3**: 212–217
- Morris P, Robbins MP** (1997) Manipulating condensed tannins in forage legumes. In BD McKersie, DCW Brown, eds, *Biotechnology and the Improvement of Forage Legumes*. CAB International, Wallingford, CT, pp 147–173
- Mueller L, Goodman CD, Silady RA, Walbot V** (2000) AN9, a petunia glutathione *S*-transferase required for anthocyanin sequestration is a flavonoid-binding protein. *Plant Physiol* **123**: 1561–1570
- Murphy A, Peer WA, Taiz L** (2000) Regulation of auxin transport by aminopeptidases and endogenous flavonoids. *Planta* **211**: 315–324
- Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, Lepiniec L** (2000) The *TT8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* genes in *Arabidopsis* siliques. *Plant Cell* **12**: 1863–1878
- Pruitt RE, Vielle-Calzada J-P, Ploense SE, Grossnicklaus U, Lolle SJ** (2000) *FIDDLEHEAD*, a gene required to suppress epidermal cell interactions in *Arabidopsis*, encodes a putative lipid biosynthetic enzyme. *Proc Natl Acad Sci USA* **97**: 1311–1316
- Quattrocchio F, Wing J, van der Woude K, Souer E, de Vetten N, Mol J, Koes R** (1999) Molecular analysis of the *anthocyanin2* gene of petunia and its role in the evolution of flower color. *Plant Cell* **11**: 1433–1444
- Rasmussen S, Dixon RA** (1999) Transgene-mediated and elicitor-induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway. *Plant Cell* **11**: 1537–1551
- Rausher M, Miller RE, Tiffin P** (1999) Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol Biol Evol* **16**: 266–274
- Saslowsky DE, Dana CD, Winkel-Shirley B** (2000) An allelic series for the chalcone synthase locus in *Arabidopsis*. *Gene* **255**: 127–138
- Saslowsky DE, Winkel-Shirley B** (2001) Localization of flavonoid enzymes in *Arabidopsis* roots. *Plant J* (in press)
- Schoenbohm C, Martens S, Eder C, Forkmann G, Weishaar B** (2000) Identification of the *Arabidopsis thaliana* flavonoid 3'-hydroxylase gene and functional expression of the encoded P450 enzyme. *Biol Chem* **381**: 749–753
- Schröder J** (1997) A family of plant-specific polyketide synthases: facts and predictions. *Trends Plant Sci* **2**: 373–378
- Schröder J** (1999) Probing plant polyketide biosynthesis. *Nat Struct Biol* **6**: 714–716
- Snyder BA, Nicholson RL** (1990) Synthesis of phytoalexins in sorghum as a site-specific response to fungal ingress. *Science* **248**: 1637–1639
- Spelt C, Quattrocchio F, Mol JNM, Koes R** (2000) *anthocyanin1* of petunia encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. *Plant Cell* **12**: 1619–1631
- Stafford HA** (1974) Possible multi-enzyme complexes regulating the formation of C<sub>6</sub>-C<sub>3</sub> phenolic compounds and lignins in higher plants. *Rec Adv Phytochem* **8**: 53–79
- Stafford HA** (1990) *Flavonoid Metabolism*. CRC Press, Boca Raton, pp 298
- Stafford HA** (1991) Flavonoid evolution: an enzymic approach. *Plant Physiol* **96**: 680–685
- Steele CL, Gijzen M, Qutob D, Dixon RA** (1999) Molecular characterization of the enzyme catalyzing the aryl migration reaction of isoflavonoid biosynthesis in soybean. *Arch Biochem Biophys* **367**: 146–150
- Tanaka A, Shigemitsu T, Yokota Y, Shika N** (1997) A new *Arabidopsis* mutant induced synthesis with spotted pigmentation. *Genes Genet Syst* **72**: 141–148
- Walker AR, Davison PA, Bolognesi-Winfield AC, James CM, Srinivasan N, Blundell TL, Esch JJ, Marks MD, Gray JC** (1999) The *TRANSPARENT TESTA GLABRA1* locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein. *Plant Cell* **11**: 1337–1350
- Watson BS, Paiva NL** (2000) Heterologous expression of vestitone reductase, an isoflavonoid biosynthetic gene, in tobacco. Sixth International Congress of Plant Molecular Biology. University of Laval, Quebec, pp S10–S21
- Wisman E, Hartmann U, Sagasser M, Baumann E, Palme K, Hahlbrock K, Saedler H, Weisshaar B** (1998) Knock-out mutants from an En-1 mutagenized *Arabidopsis thaliana* population generate phenylpropanoid biosynthesis phenotypes. *Proc Natl Acad Sci USA* **95**: 12432–12437
- Yephremov A, Wisman E, Huijser P, Huiser C, Wellesen K, Saedler H** (1999) Characterization of the *FIDDLEHEAD* gene of *Arabidopsis* reveals a link between adhesion response and cell differentiation in the epidermis. *Plant Cell* **11**: 2187–2201
- Yu O, Jung W, Shi J, Croes RA, Fader GM, McGonigle B, Odell JT** (2000) Production of the isoflavones genistein and daidzein in non-legume dicot and monocot tissues. *Plant Physiol* **124**: 781–793