

# *Annual Review of Pharmacology and Toxicology* Artificial Intelligence for Drug Discovery: Are We There Yet?

Catrin Hasselgren<sup>1</sup> and Tudor I. Oprea<sup>2,3</sup>

<sup>1</sup> Safety Assessment, Genentech, Inc., South San Francisco, California, USA

<sup>2</sup> Expert Systems Inc., San Diego, California, USA; email: [toprea@expertsystems.inc](mailto:toprea@expertsystems.inc)

<sup>3</sup> Department of Internal Medicine, University of New Mexico Health Sciences Center, Albuquerque, New Mexico, USA

Annu. Rev. Pharmacol. Toxicol. 2024. 64:527–50

First published as a Review in Advance on  
September 22, 2023

The *Annual Review of Pharmacology and Toxicology* is  
online at [pharmtox.annualreviews.org](http://pharmtox.annualreviews.org)

<https://doi.org/10.1146/annurev-pharmtox-040323-040828>

Copyright © 2024 by the author(s). This work is  
licensed under a Creative Commons Attribution 4.0  
International License, which permits unrestricted  
use, distribution, and reproduction in any medium,  
provided the original author and source are credited.  
See credit lines of images or other third-party  
material in this article for license information.

**ANNUAL  
REVIEWS CONNECT**

[www.annualreviews.org](http://www.annualreviews.org)

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

## Keywords

autoencoders, deep learning, explainable AI, generative chemistry, knowledge graphs, machine learning, multiproperty optimization, small-molecule drug discovery, target identification

## Abstract

Drug discovery is adapting to novel technologies such as data science, informatics, and artificial intelligence (AI) to accelerate effective treatment development while reducing costs and animal experiments. AI is transforming drug discovery, as indicated by increasing interest from investors, industrial and academic scientists, and legislators. Successful drug discovery requires optimizing properties related to pharmacodynamics, pharmacokinetics, and clinical outcomes. This review discusses the use of AI in the three pillars of drug discovery: diseases, targets, and therapeutic modalities, with a focus on small-molecule drugs. AI technologies, such as generative chemistry, machine learning, and multiproperty optimization, have enabled several compounds to enter clinical trials. The scientific community must carefully vet known information to address the reproducibility crisis. The full potential of AI in drug discovery can only be realized with sufficient ground truth and appropriate human intervention at later pipeline stages.

## 1. INTRODUCTION

Drug discovery is a systematic scientific process that aims to identify, design, and develop novel therapeutic agents to cure, ameliorate, or prevent diseases and medical conditions. Drug discovery is often called a pipeline, which suggests a unidirectional transition from hit/lead to candidate and marketed drug, supported by basic and clinical research (1). The process is, in fact, iterative in nature, multifaceted, and complex. For example, small-molecule drug discovery requires (a) basic science research and target identification; (b) target pharmacology and biomarker development; (c) lead identification; (d) lead optimization, candidate selection, investigational new drug (IND)-enabling studies and scale-up for manufacturing; (e) clinical research and development; (f) regulatory review; (g) postmarketing; and (h) medical practice deployment (2). Two drug discovery, development, and deployment maps (4DMs) that depict key steps in the drug development life cycle, one for small molecules and one for biologics, illustrate the interdependencies and complexities of this process and are available for download at the National Center for Advancing Translational Sciences (NCATS) website (<https://ncats.nih.gov/translation/maps>). Drug discovery scientists use various techniques and methodologies, such as computational modeling, medicinal chemistry, high-throughput screening (HTS), and biological assays, to identify promising compounds and evaluate their safety, efficacy, and pharmacokinetics.

As therapeutic modalities continue to evolve (3), the drug discovery process is adapting by incorporating novel data science, informatics, and artificial intelligence (AI) methodologies, among other technologies, to improve efficiency and reduce costs and animal experiments, thus accelerating the development of novel, effective treatments. The impact of big data and AI in drug discovery, the subject of a 2020 review in this journal (4), continues to attract significant interest: Investors (5, 6), industrial (7, 8) and academic scientists (9, 10), and legislators (11) are discussing the impact of artificial intelligence for drug discovery (AI4DD). As of July 31, 2023, “artificial intelligence in drug discovery” is in the title of 65 publications since 2019, according to Google Scholar (<https://bit.ly/3FWk3dH>).

Successful drug approval requires concurrently optimizing multiple properties related to pharmacodynamics, pharmacokinetics, and clinical outcomes. Pharmacodynamic properties are related to drug-target interactions and efficacy; appropriate pharmacokinetics include absorption, distribution, metabolism, excretion, toxicity (ADMET), and drug safety; and clinical outcomes include the therapeutic intent, as detailed on the list of drug indications and off-label uses, as well as undesired outcomes such as side effects or adverse drug reactions. Thus, a successful drug discovery launch relies on three pillars: diseases, targets, and therapeutic modalities. AI impacts most therapeutic modalities, including targeted protein degradation (12), antibodies (13), gene therapy (14), and oligonucleotide (15) and vaccine (16) design. In this review, we explore the impact of AI4DD on these three pillars, primarily focusing on small-molecule drugs as a therapeutic modality. Its impact on regulatory science and some caveats of AI4DD are also included.

## 2. EMERGENCE OF AI FOR DRUG DISCOVERY

### 2.1. The Knowledge Deficit

One of the main challenges human investigators and AI systems face in drug discovery is harnessing large volumes of heterogeneous data of varying quality. The rapid expansion of data and computing power has been described as justifying a fourth paradigm, also known as data-intensive scientific discovery (17). Specifically for the “why” and “what if” types of questions, relevant, preferably reliable data must be identified, inferred when absent, and connected through evidence-based reasoning, illustrated by the metaphor of connecting the dots (18). It is increasingly clear

that modern drug discovery requires computer-based systems that can intelligently reason and recognize patterns, that is, AI. These AI systems must be able to weigh data elements and aggregate instances of those patterns to assess confidence and justification. Automated systems that digest large sets of data using named entity recognition (19) are an integral part of public domain databases, for example, DISEASES for gene-disease associations (20), search tool for recurring instances of neighbouring genes (STRING) for protein-protein interactions (21), and OpenTargets (22) and Pharos (23) for complex disease-protein-drug annotations, to name a few. Combined with AI-based systems that predict protein structures, such as AlphaFold (24) and RoseTTAFold (25), these resources are poised to accelerate AI4DD.

## 2.2. An Incomplete History of AI for Small-Molecule Drug Design

The general notion that, at some point, an AI platform will drive drug discovery was probably formulated in the early days of science fiction. For example, TV shows such as *Star Trek* have featured medicine-specialized holograms (AI entities) solving medical emergencies. As early as 1981, the idea of designing drugs with computers was motivating scientists at Washington University in St. Louis (26) and Merck Research Laboratories (27). The *Journal of Computer-Aided Molecular Design* was founded in 1987, and in almost every decade since the 1980s, the notion of AI in drug discovery has captured our imagination.

Machine learning (ML) models that discriminate drugs from nondrugs emerged in 1998, when scientists from Vertex (28) and BASF (29) independently proposed models to estimate drug-likeness, showing that it is possible to train ML models based on chemical features to discriminate drugs (specifically, compounds that medicinal chemists have proposed for biological testing) from nondrugs (compounds lacking pharmaceutical use). The challenge faced by ML models of drug-likeness is illustrative of the general challenge in drug discovery (specifically, small-molecule drugs): The drug quality is not an intrinsic property of chemicals since regulatory agencies and, implicitly, humans approve compounds for medicinal use (30). Furthermore, drug approval is subject to revision and can be withdrawn over time. While most market withdrawals are caused by drug toxicity (31), lack of efficacy or economic reasons can also result in withdrawal. Despite these caveats, ML-based drug-likeness quantification is an integral part of the drug discovery process, with nearly 20 independent papers dedicated to drug-likeness discussed elsewhere (30).

## 2.3. Current Impact in Drug Discovery

An in-depth scientometric analysis (32) of AI4DD showed a steep rise in publications from 49 in 2011 to 333 in 2020. The number of AI-based drug discovery platforms is anticipated to increase in the near future. There is frequently a synergistic relationship between the pharmaceutical and biotech industries, which propel AI-driven drug discovery toward commercial use, and academic institutions, which often spearhead the development of algorithms and methods. Over the past two decades, AI and ML have shifted from being peripheral technologies to occupying a central role in the drug discovery process. Today, we are closer than ever to achieving this long-sought goal.

## 3. AI APPLICATIONS IN VARIOUS STAGES OF DRUG DISCOVERY

### 3.1. Diseases and Therapy Selection

Recent advancements in computational technologies have resulted in significant strides in applying AI in the health-care sector. AI has been used to tackle various diseases, enhance imaging analytics, assist in nosology, and examine drug indications and therapeutic intent. With the

utilization of AI in health care showing considerable promise, in this section we discuss the impact of AI in transforming medical diagnosis and treatment.

**3.1.1. Diseases and health care.** The last decade has witnessed significant computational advances in disease diagnosis, with ten or more publications describing the use of AI in dermatology, tuberculosis, Alzheimer's disease (AD), diabetes, hypertension, and cancer, reviewed elsewhere (33). The potential influence of AI in health-care professions such as radiology (34), pathology (35), clinical pharmacology (36), and COVID drug repurposing (37) has been a topic of discussion. Recently, ChatGPT, a large language model (LLM) from OpenAI (<https://chat.openai.com/>), achieved a notable milestone by scoring 60% proficiency on the US Medical Licensing Examination (38). AI has the potential to impact health-care management in areas such as diagnostics, imaging analytics, patient-provider interactions, hospital and nursing home management, population wellness using social determinants, and real-world evidence collection and analysis (39). As of March 30, 2023, the general opinion is that AI will support health-care professionals but is not expected to replace them in the near future.

**3.1.2. Image-based AI.** AI systems based on image analysis have been developed in various medical fields, including radiology, pathology, and dermatology. Pathology is evolving into digital pathology, which involves digitizing histopathology, immunohistochemistry, and cytology slides and training AI systems on this digital data (40). Computational pathology aims to reduce diagnostic and classification errors while accelerating biomarker discovery, pathophysiology evaluations, animal research, and toxicological assessments (41). In radiology, AI can assist radiologists in improving diagnoses and enhancing patient care by analyzing chest X-rays, magnetic resonance images, and computed tomography scans (42). However, a systematic framework may be needed for comparing human and AI perception in medical diagnosis (43). An ML model based on convolutional neural networks (CNNs) has demonstrated near-physician accuracy in diagnosing atopic dermatitis. Furthermore, this CNN model performs well in differentiating atopic dermatitis from other skin conditions such as impetigo, mycosis fungoides, herpes simplex, and Kaposi varicelliform eruption (44).

**3.1.3. Nosology.** Nosology, the ability to classify, recognize, cross-reference, and reconcile disease terms, poses a significant challenge for health-care AI systems. While numerous disease and phenotype ontologies exist (45), the Mondo Disease Ontology (Mondo) is emerging as a comprehensive, community-driven, and computationally driven resource (46). A systematic analysis of resources integrated into Mondo revealed nearly 10,500 unique rare diseases, in contrast to the commonly cited figure of 7,000 (47). The successful combination of human and ML (Bayesian) curation employed by Mondo serves as a clear example of how human and computer intelligence can collaborate in understanding diseases.

**3.1.4. Drug indications and therapeutic intent.** A critical aspect of drug discovery involves understanding the relationship between drugs and diseases, as specified on the drug label or drug indication, typically listed in the indications and usage section. Uses not stated on the package inserts are called off-label uses. As of January 2023, the DrugCentral database contains 2,331 US Food and Drug Administration–approved human drugs associated with 2,644 drug indications and 866 off-label uses (48). Therapeutic intent, the rationale behind choosing a therapy and the context in which it is prescribed, is vital to medical practice. Although DrugCentral maps drug uses to existing terminologies such as Systematized Nomenclature of Medicine–Clinical Terms (SNOMED-CT) and Medical Subject Headings (MeSH), two global standards for health terms, capturing therapeutic intent computationally is challenging (49). Therapeutic intent encompasses

disease concepts and contextual factors such as preexisting conditions; coprescribed medications; specific genotypes; and various phenotypic, anatomical, or temporal parameters. The state of the art in drug indication curation from text was expanded through InContext (50), which includes therapeutic context information about drug usage, especially for antineoplastic and cardiovascular drugs. However, appropriately capturing and validating off-label medical uses still requires manual curation (51). Developing AI systems capable of processing, corroborating, and cross-mapping disease terminologies and therapeutic intent is a significant goal for automated drug discovery and repurposing in silico.

## 3.2. Target Identification and Validation

**3.2.1. Knowledge graphs.** A crucial early step in modern drug discovery is target identification. This process involves discovering and validating the clinical relevance of a biomolecule that serves as the mode of action for the therapeutic modality (52). Although novel drug targets are typically identified through genomic, proteomic, or phenotypic experiments, ML is increasingly being explored as a target identification technology (53), often by means of knowledge graphs (KGs) (54). A KG is a knowledge base that uses graph-structured data models or topology to integrate data. Drug discovery KGs store interlinked descriptions of nodes—genes, phenotypes, or compounds—while also encoding the underlying relationships (edges). KG data projection enables network-based analytical algorithms, thus turning complex drug discovery data into ML-ready files. Biological system networks are heterogeneous, with multiple node and edge types (**Figure 1**). Developments in heterogeneous (55) relationship predictions (56) introduced and formalized a new framework that takes into account network heterogeneity by defining type-specific node-edge paths or meta-paths (57).

A meta-path encodes a type-specific network topology between the source node (e.g., protein target) and the destination node (e.g., disease or function). Type-specific meta-paths are as follows: [target—(member of) → protein-protein interaction (PPI) network ← (member of)—protein—(associated with) → disease] and [target—(expressed in) → tissue ← (localized in)—disease]. Type-specific meta-path counts can be combined using, for example, degree-weighted paths (56) to dampen the effect of highly connected nodes. By assembling heterogeneous data types such as those included in Pharos (23), an ML-ready KG can capture data from major areas specific for human protein-coding genes: phenotype and disease, pathways, and interactions. Each area has appropriate levels of data, for example, expression, association, membership, treatments, localization, and gene signatures. Each instance of a meta-path represents a specific chain of evidence of associations between a source and a destination node.

**3.2.2. Novel targets for Alzheimer's disease.** The meta-path framework can be applied in conjunction with various classification algorithms. In this case, XGBoost, a form of extreme gradient boosting (68), was utilized to identify novel genes associated with AD (69). An AD-focused model was developed by integrating the comprehensive Pharos data set, a web-based interface for data collected by the Illuminating the Druggable Genome initiative, into XGBoost and included 53 positive (AD-associated) and 3,952 negative (non-AD-associated) genes. The model's validity was assessed using the top 20 and bottom 10 genes across three biologically distinct AD model systems (69). This approach uncovered five previously unassociated gene targets related to immunity in AD: *FRRS1*, *CTRAM*, *SCGB3A1*, *CIBAR2*, and *TMEFF2*. These genes demonstrated significant differences between AD and control samples. The model identified key variables associated with inflammatory processes and oxidative stress. Although constrained by input data that lacked nonhuman information, the AD-specific XGBoost model emphasized the role of infection as a

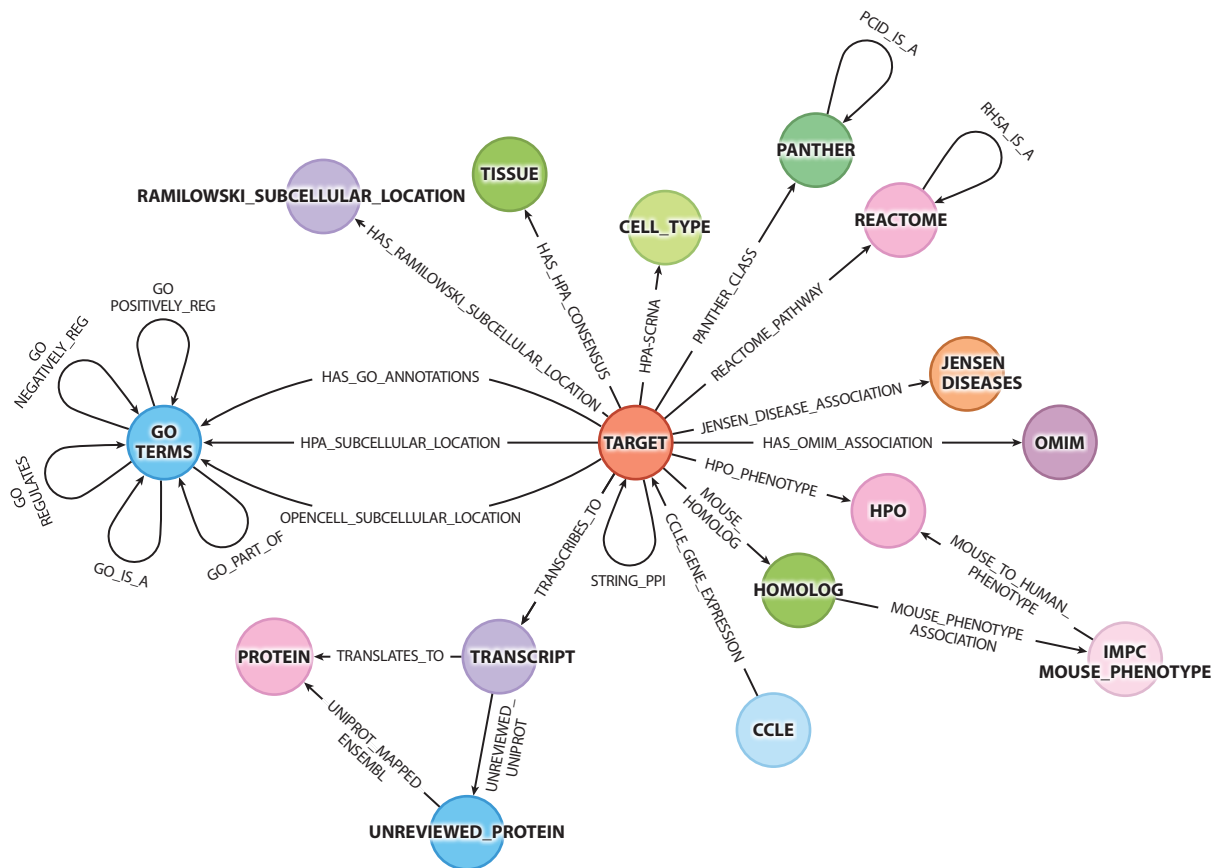


Figure 1

Knowledge graph representation of phenotype-protein associations. Nodes (*colored circles*) are related by edges (*arrows*). Centered on the main node, Target, nodes listed clockwise are as follows: tissue; cell\_type; Protein Analysis Through Evolutionary Relationships (PANTHER), a protein classification scheme (58); Reactome, a pathways database (59); Jensen DISEASES (20); Online Mendelian Inheritance in Man database (OMIM) (60); human phenotype ontology (HPO) (61); Homolog (which lists proteins with common evolutionary origin), further linked to the International Mouse Phenotype Consortium (IMPC) mouse phenotype database (62); Cancer Cell Line Encyclopedia (CCLE) (63); transcript, protein, and unreviewed\_protein [various types of UniProt (64) entries]; Gene Ontology consortium (GO) (65) terms; and Ramilowski\_subcellular\_location, a specific experimental resource for subcellular protein localization (66). Nodes are associated to Target through a variety of arrows. GO terms, for example, features several protein properties, whereas String (21) loops protein-protein interactions (PPIs) for Target; other resources, such as the Human Protein Atlas (HPA) (67), inform a variety of nodes regarding tissue, cell type, and subcellular location and expression. Figure created with help from Jessica Binder.

critical factor in AD pathology. There is a growing body of evidence supporting the hypothesis that brain infection may be the primary cause of AD (70).

**3.2.3. Novel genes for autophagy.** Autophagy (ATG) is an intracellular physiological degradation process responsible for eliminating protein aggregates, pathogens, and other cargo by delivering them to lysosomes for degradation (71). Although the number of known genes involved in ATG is increasing (72), a systematic experimental method for identifying ATG-related genes has yet to be established (73). Given the complexity of ATG-associated genomic data sets, an ATG-specific XGBoost ML model was developed to aid in designing relevant experiments



(74). The ATG-specific meta-path XGBoost model used 103 ATG-associated (positive) genes and 3,468 negatives (not ATG-associated genes), collating data from Pharos. Among the top 251 genes predicted to have ATG associations, 43 genes were already in the ATG database, and another 15 had potential ATG associations. However, 193 genes (77%) did not have any apparent connection to ATG. These 193 genes could be considered ATG dark genes. A literature review of newly published experimental reports was used to validate 7 of the top 20 and 2 of the bottom predicted ATG dark genes (74).

**3.2.4. Knowledge graph-based machine learning model interpretation.** The AD and ATG XGBoost meta-path models exhibit external predictivity, as some of the top 20 genes predicted by each model have demonstrated valid associations. Some valid associations were also found among the bottom-ranked genes in each model. Since the XGBoost algorithm generates a gradient (probability) output and input labels are binary (1 for positives and 0 for negatives), examining both top- and bottom-ranked genes is a logical approach. Three of the bottom-ranked genes displayed signals for both AD and ATG. It is advisable to consider both top- and bottom-ranked genes because ML output is not absolute. The distinction between top-predicted and bottom-predicted genes is more evident in their similarity network. Top genes are more likely to be associated with AD or ATG, while the bottom-ranked genes seem less organized, as indicated by STRING enrichment analyses. Both models used an arbitrary set of over 3,000 negative-label genes; however, no experiments have confirmed these genes as true negatives for either ATG or AD. The absence of a ground truth (verified true positives and true negatives) likely contributes to the increased complexity and variance observed in KG-based ML research.

One potential complication in KG-based ML models is data leakage (75), which can occur inadvertently due to the complexity of the data set. In the context of the ATG-based ML model, the inclusion of Gene Ontology consortium (GO) (65) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (76) terms can introduce information directly related to the research question. Both GO and KEGG contain terms associated with ATG. Post-ML queries for ATG dark genes led to the exclusion of one bottom-ranked and two top-ranked genes, as they were annotated with ATG-related GO terms. Data leakage has a significant impact, and Kapoor & Narayanan (77) describe it as a reproducibility crisis in ML science. In their analysis, data leakage was detected in 15 of the 20 subject areas associated with biology or medicine.

### 3.3. Hit Generation and Lead Optimization

**3.3.1. Bioactivity and ADMET property ML.** A quantitative structure-activity relationship (QSAR) is a fundamental aspect of computer-aided drug design and has been increasingly incorporated into AI in drug discovery, including generative chemistry and multiproperty optimization (MPO). Developed from the work of Hansch & Fujita (78) in the 1960s, QSAR was the first systematic application of ML to design small molecules with desired properties (79). QSAR can cover all aspects of drug property, ranging from target-based bioactivity to ADMET properties, such as solubility, permeability, and other physicochemical parameters. QSAR methods are algorithm agnostic and use descriptors (features) derived from molecular structures (objects) (80) to explain the relationship between changes in molecular structure and the molecular property of interest. QSAR has applications in fields such as synthesis planning, materials science, nanotechnology, and clinical informatics (81). QSAR technologies are an integral part of the technologies deployed to hunt for anti-COVID-19 therapies (82). QSAR methodologies can be automatically deployed against thousands of target bioactivities (83). Recent lessons shared by scientists at AstraZeneca (84), Bayer (85), Merck (86), and GSK (87) highlight the complexities of ADMET property prediction (88).

Traditionally, QSARs and other small-molecule property ML models use chemical structures as input, often encoded as a simplified molecular input line entry system (SMILES) (89). A recent development is the concept of graph convolutional neural networks (GCNNs), introduced in 2015 (90). GCNNs have been used to model drug combinations (91), drug-target interactions (92), and multiple property predictions (93). Rigorous analyses suggest that GCNNs do not improve predictive performance compared to classical QSAR fingerprints (93).

**3.3.2. Generative chemistry.** With the foundational role of QSAR in computer-aided drug design and its wide range of applications, the field has continued to evolve and incorporate more advanced techniques such as generative chemistry (94). The development of ultralarge chemical libraries (95) and generative chemistry methods, particularly those based on generative adversarial networks (GANs), has significantly expanded the possibilities for hit generation and drug discovery. GANs (96), inspired by game theory and deep learning, have found applications in chemistry, such as the first deep learning GAN used in anticancer compound design (97). This application employed a seven-layer adversarial autoencoder architecture with growth inhibition percentage as the discriminator. The ML step in this workflow used NCI-60 data (98),  $GI_{50}$  (growth inhibition of 50%), and compound log concentration (LCONC) data for 6,252 compounds profiled on the MCF-7 cell line, focusing the GAN on better  $GI_{50}$  at lower LCONC. Using 32 probability vectors, 72 million compounds from PubChem (99) were screened, with a maximum of 10 hits per vector. Over one-third of the selected compounds demonstrated activity in cancer assays or were patented for their anticancer potential (97). Building upon this method, a GAN was trained to generate *JAK3*-selective molecules using an entangled conditional adversarial autoencoder (100). This led to the identification of a low-micromolar *JAK3* inhibitor, which was considerably less potent on *JAK2*, *BRAF*, and *RAF1*. Additional generative chemistry platforms are described below.

**3.3.3. Multiproperty optimization.** Mathematical modeling of multiobjective decision making has been prevalent since the 1960s (101). At that time, the process of evaluating alternatives based on predefined objectives or criteria was represented as  $n$ -dimensional vectors, which were subject to Pareto optimization, aiming to find nondominant solutions. Hopfinger (102) framed this discussion in terms of computer-aided drug design in 1985. Pareto optimization has emerged as the preferred method for MPO, as it allows for the identification of trade-offs between objectives, leading to the discovery of nondominant solutions (103). This is particularly useful when evaluating drug properties that are anticorrelated. For instance, increasing drug permeability by one log unit can lead to a decrease in water solubility of up to two log units (104). Therefore, finding nonmaximal solutions for each drug property is often necessary when optimizing drug properties.

The use of computers for MPO of drug starting points such as hits and leads at industrial scale began to gain traction two decades ago (105, 106). The first automated drug discovery implementation was disclosed in 2012 (107), which later became the founding technology of Exscientia, an AI-based drug design company. This particular MPO platform encodes target and off-target bioactivities, ADMET properties, and other criteria. It uses a Laplacian-modified (108) Bayesian classifier (109) trained for polypharmacology profiling and employs a vector scalarization procedure (110) for Pareto-based adaptive optimization. TorchDrug is an open-source AI4DD platform (111) built on PyTorch (112), which leverages the power of deep learning to identify novel drugs. It supports property prediction, pretrained molecular representations (113), generative chemistry, and biomedical KG reasoning (114). Another platform, REINVENT (115), supports MPO by combining reinforcement learning for the generative chemistry model with diversity filters (116) and a flexible scoring function (117). REINVENT relies on PyTorch

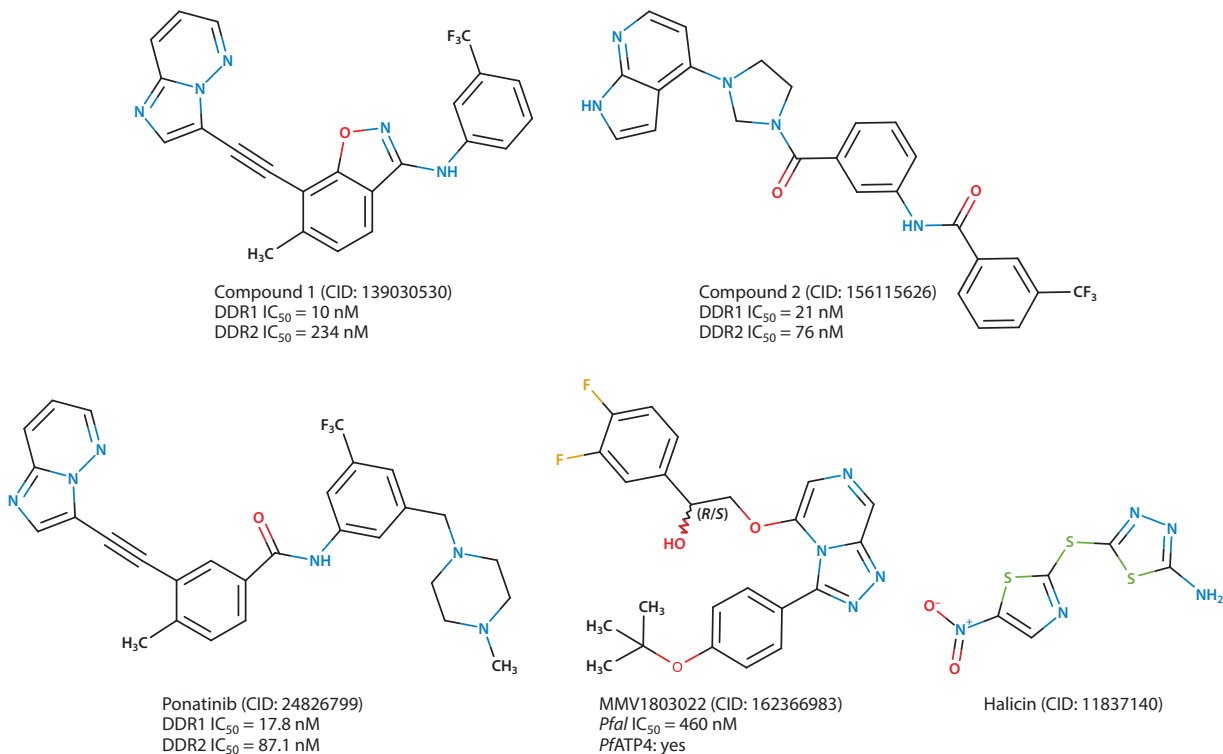


and RDKit (118) as a chemistry engine, uses predictive ML models implemented in the scikit-learn library (119), and utilizes TensorBoard's implementation (120) in PyTorch for detailed documentation of chemical space navigation processes.

Generative tensorial reinforcement learning (GENTRL) is a GAN method that was used to generate *DDR1* and *DDR2* inhibitors (121). Discoidin domain-containing receptor kinases 1 and 2 (*DDR1* and *DDR2*) are collagen receptors (122). *DDR1* plays roles in tumorigenesis, metastasis (123), and fibrosis (124), which makes it an attractive drug target. GENTRL was trained on the clean leads subset of the ZINC database (125), enriched with known active and inactive kinase inhibitors from ChEMBL (126). Additionally, it was trained on patented chemical structures, structural (X-ray) data, and *DDR1/DDR2*-specific pharmacophores. The initial output from GENTRL consisted of 30,000 compounds, which were filtered and prioritized to eliminate undesirable scaffolds, reactive groups, and compounds that were unlikely to be kinase inhibitors. This process resulted in six novel compounds predicted to be *DDR1* and *DDR2* inhibitors. After synthesis and experimental testing for inhibitory activity, two compounds, Compound 1 and Compound 2, were found to be potent (nanomolar range) inhibitors of *DDR1* and *DDR2*, respectively (**Figure 2**). The entire process, from nominating *DDR1* to generating Compound 6, took only 35 days. Further experiments demonstrated that Compound 1 is orally bioavailable in mice, with a half-life of 3.5 h after a 15 mg/kg dose (121).

**3.3.4. Community engagement.** Therapeutic Data Commons (TDC; <https://tdcommons.ai/>) hosts multiple ML-ready data sets to advance AI4DD technologies systematically (128). TDC features leaderboards for 22 ADMET properties, 5 drug combination benchmarks, 1 docking, and 1 drug-target interaction data set (29 in total), building on earlier benchmarks for molecular ML (129). For example, the MolE deep learning foundational model (130) demonstrates superior performance in 9 of the 22 ADMET property data sets. The QSAR community widely uses the Online Chemical Modeling Environment (OCHEM; <https://ochem.eu/>), an online platform for ML model development and storage (131). OCHEM provides a uniform interface for multitask learning (132), a method that promises improved prediction performance (133). Some target-based multiproperty ML models are available online: NCATS Predictor (<https://predictor.ncats.io/>) features 1,180 different models (134), while Redial-2020 (<http://drugcentral.org/Redial>) estimates 12 properties to evaluate compounds active against COVID (135). For ADMET, one could consider the ADME@NCATS (<https://opendata.ncats.nih.gov/adme/>) portal (136) or ADMETlab 2.0 (<https://admetmesh.scbdd.com/>), which includes not only ADMET properties but also parameters that may aid medicinal chemists (137). Multiple academic and industrial teams participated in an open AI4DD antimalarial drug competition (127) by targeting *PfATP4*, the essential P-type ATPase from *Plasmodium falciparum*. MMV1803022, a *PfATP4* inhibitor, causes cytosolic alkalinization (via  $[\text{Na}^+]$  increase) and is active against *P. falciparum* (127) (see **Figure 2**).

The output of generative chemistry necessitates human supervision. Following the generative step, active filters are needed to remove potentially reactive and chemically unfeasible species and compounds that fall outside the desired property space (138). Even AI-designed compounds that have been successfully tested can face criticism due to their lack of novelty. For example, Compound 1 (**Figure 2**) has been criticized for its similarity to the approved kinase inhibitor ponatinib (139). An AI-designed antibiotic, halicin (**Figure 2**), disclosed in 2020 (140), has 12 active reports in PubChem BioAssays (<https://bit.ly/halicinAssays>), including an active antibiotic result against *Mycobacterium tuberculosis* from 2017. According to *Chemical & Engineering News* (141), researchers “chose not to pursue [halicin] because of its similarity to a compound that the US Food and Drug Administration was already evaluating.” These cases highlight the need for improved AI-generated compound novelty in drug discovery.



**Figure 2**

Chemical structure of two epithelial discoidin domain-containing receptor inhibitors (Compounds 1 and 2), developed using the GENTRL (121) generative chemistry approach; Compound 1 bears striking chemical similarity with ponatinib (*bottom left*). The structure and bioactivity of MMV1803022, one of the community-driven antimalarial artificial intelligence (AI) competition compounds, are shown (*bottom center*). *Pfal* summarizes the in vitro growth inhibition of asexual blood-stage *Plasmodium falciparum* (3D7) parasite maturation; *PfATP4* indicates confirmed on-target biochemical (cytosolic [Na<sup>+</sup>] increase) inhibition. Halicin, an AI-discovered antimicrobial agent, is shown on the bottom right. Abbreviations: CID: PubChem compound identifier; DDR, discoidin domain-containing receptor kinase; GENTRL, generative tensorial reinforcement learning. Bioactivities for Compounds 1 and 2 are from Reference 121, bioactivity for ponatinib is from DrugCentral (48), the bioactivity for the antimalarial compound MMV1803022 is from Reference 127, and the structure of halicin is from PubChem (<https://pubchem.ncbi.nlm.nih.gov/compound/Halicin>).

### 3.4. Drug Safety: The Road to Clinical Trials

Preclinical safety assessment employs both in vitro and in vivo testing to ensure patient safety prior to first-in-human trials. In vitro toxicology testing is designed for early detection of toxicities that might have a serious impact on the progression of a drug discovery program and comes in a range of assay formats, from standard single-end point screening to, more recently, complex in vitro models systems, including microphysiology or organ-on-chip models. These are applied before or in parallel with in vivo studies. AI enhancements in this space involve the combination of diverse, more complex data forms (e.g., gene expression, cell painting), as well as broader application of imaging for a variety of data processing applications. One such example is the application of deep learning to detect skin toxicity from an in vitro 3D human skin model (142).

**3.4.1. Predictive toxicology.** Small-molecule MPO design (discussed above) includes toxicity prediction of in vitro outputs related to organ toxicities such as cytotoxicity and mitochondrial toxicity or representative assays for predicting off-target toxicities as well as genotoxicity (143).

New advancements include analyses of novel data types such as gene expression data and data from cell imaging experiments, in combination with chemical structure information, to predict in vivo toxicity-related effects (144). Using biological fingerprints, either alone or in combination with structural fingerprints, can better relate compounds to in vivo phenotypes, sometimes with superior predictivity (145). Including a biological matrix has the advantage of better identifying the subtle differences in phenotype not fully captured by structural similarity alone (146) and can be further facilitated by the inclusion of imputed values in the biological matrix. An example is the application of a neural network to fill a sparse kinase-binding matrix (147). An important aspect of predictive toxicology is toxicity in the context of exposure and understanding the safety threshold. Bayesian ML approaches have proved useful for predicting clinical outcomes from mechanistically relevant in vitro data combined with animal exposure (148). Furthermore, AI-augmented clinical pharmacology approaches can impact dose recommendation, drug-drug interaction prediction, and adverse drug reaction prediction (149) by utilizing digital data sources such as electronic health records, biomarker data, and patient genomics data. KG-based MLs that incorporate side effect information from package inserts, target information, and indication information have shown promise in predicting adverse events (150).

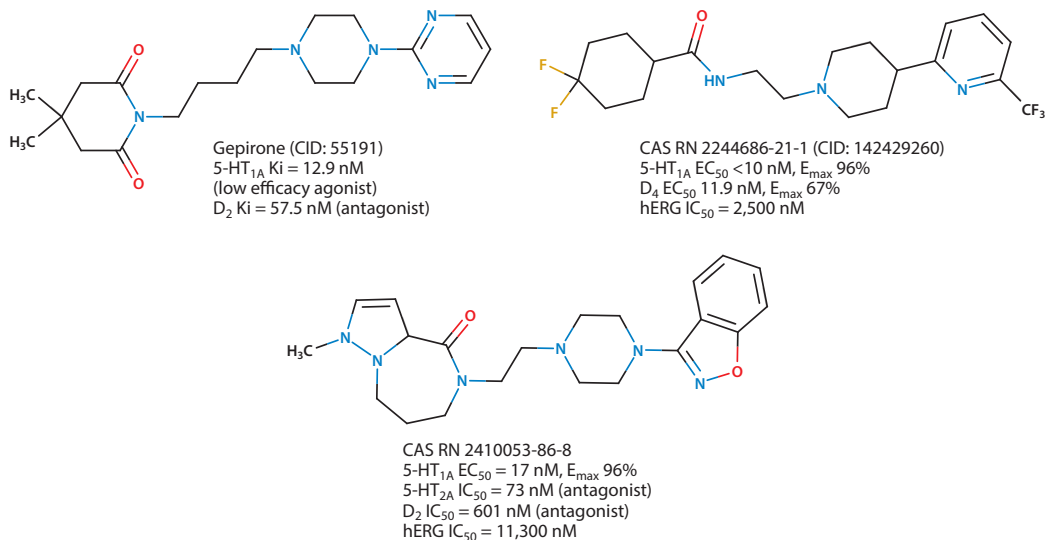
**3.4.2. Regulatory applications.** Regulatory agencies are showing increasing interest in AI4DD and its uses. The application of computational tools to evaluate the carcinogenic potential of drug-related impurities has led to the inclusion of structure-activity predictions as part of regulatory submissions (151). In 2018, the European Medicines Agency published a reflection paper discussing the use of in silico tools for assessing risks related to nonmutagenic impurities when experimental data are unavailable (152). More recently, regulatory bodies have published more encompassing perspectives on the use of AI/ML in drug development and drug manufacturing and are actively engaging with stakeholders for feedback (153–155).

The use of AI in regulatory applications introduces a new aspect, requiring models to be transparent enough for health authorities to assess their usability and reliability. Cross-industry consortia have been working toward formalizing the use of in silico tools, including AI/ML models, to ensure transparency in data set origins and quality, as well as algorithms, in order to gain regulatory acceptance (156, 157).

AI is gaining traction in various applications, such as chemical risk assessment related to occupational toxicology, transportation of chemicals, and medical devices, where end points such as skin sensitization, irritation, and rat acute oral toxicity are commonly included (158, 159). In the area of modeling and simulation, which is now a standard tool used to demonstrate effects on physiology or safety when expanding to other clinical populations or indications, efforts are being made to define criteria for model evaluation (160). The National Center for Toxicological Research, in collaboration with the Center for Drug Evaluation and Research, has launched the SafetAI initiative (<https://www.fda.gov/about-fda/nctr-research-focus-areas/safetai-initiative>) to develop AI models for toxicological end points that are crucial for assessing drug safety, potentially playing a role in the IND review process. This project focuses on developing deep learning-based models for hepatotoxicity, carcinogenicity, nephrotoxicity, and others. It utilizes various in vitro data sources, such as high-throughput transcriptomics, in addition to chemical representation to predict in vivo outcomes.

## 4. BRIEF OVERVIEW OF AI-DRIVEN DRUG DISCOVERY SUCCESSES

The first AI-designed compound is DSP-1181, a potent and long-acting 5-HT<sub>1A</sub> receptor agonist (161). DSP-1181 was designed using Exscientia's MPO approach, Centaur Chemist, in less than 12 months. Sumitomo Dainippon Pharma progressed DSP-1181 into Phase I clinical studies



**Figure 3**

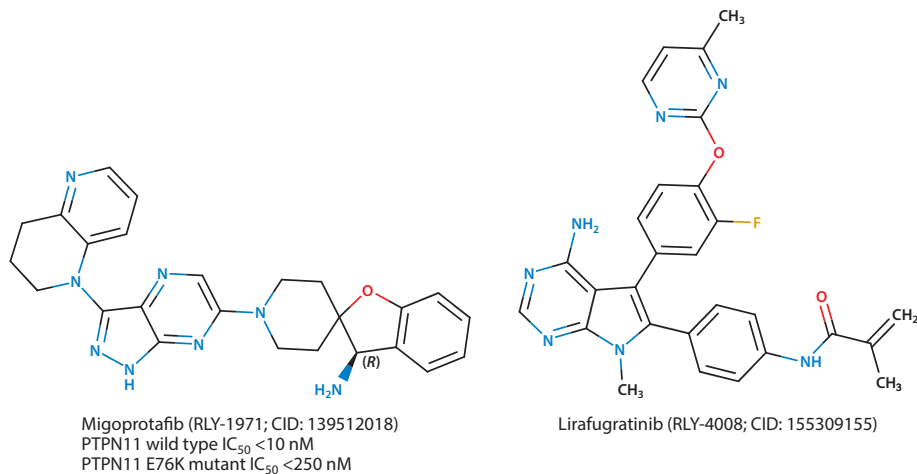
Examples of compounds reported in Exscientia patents. Gepirone (*top left*) shares the pharmacological profile of example 1 (CAS RN 2244686-21-1, *top right*). Example 109 (CAS RN 2410053-86-8, *bottom*) does not share chemical similarity with known drugs. Bioactivity data for gepirone are from DrugCentral (48), and data for patented examples are from References 162 and 164. Abbreviations: CID, PubChem compound identifier; D, dopamine D2 receptor; hERG, human Ether-à-go-go-Related Gene; K<sub>i</sub>, inhibition constant.

for obsessive-compulsive disorder in January 2020. The chemical structure of DSP-1181 has not been disclosed (162). Chemical Abstracts Service analysts suggested that DSP-1181 is similar to haloperidol (163). Since haloperidol is a weak 5-HT<sub>1A</sub> antagonist (data not shown), we suggest that it is similar to gepirone, a low-efficacy 5-HT<sub>1A</sub> agonist. Another AI-designed compound, DSP-0039, is a dual 5-HT<sub>1A</sub> agonist and 5-HT<sub>2A</sub> antagonist with no dopamine D2 receptor activity (164). Sumitomo Dainippon Pharma progressed to Phase I clinicals for treatment of AD psychosis in May 2021 (see **Figure 3** for details).

Exscientia has made significant AI-driven advancements in functional precision oncology. A single-cell functional precision medicine (scFPM) platform successfully guided treatment selection and enhanced patient outcomes in a prospective clinical study (165). In the study utilizing the scFPM approach, 139 drugs were tested on samples from 143 patients with hematologic malignancies. Out of the 56 patients treated based on scFPM results, 30 (54%) experienced over 1.3-fold increased progression-free survival compared to their previous therapy. Additionally, 12 (40% of responders) had exceptional responses lasting three times longer than expected for their respective diseases (165). Compared to patients receiving physician-chosen therapy, those treated with scFPM demonstrated a notable overall survival benefit.

Insilico Medicine's ISM01-055 (also known as ISM018-055) may be the first AI-designed compound to enter Phase II clinical trials in June 2023 (166). Designed using the PandaOmics (167) and Chemistry42 (168) platforms, ISM018-055, which may be covered by a patent (169), targets the NCK-interacting protein kinase *TNIK* (170) and is indicated for the treatment of idiopathic pulmonary fibrosis.

Relay Therapeutics, an AI4DD company focused on molecular dynamics, disclosed RLY-1971 (**Figure 4**), an orally bioavailable allosteric *PTPN11* (SHP2) inhibitor. RLY-1971 blocks wild-type *PTPN11* (IC<sub>50</sub> <1 nM) and the E76K activating mutant (IC<sub>50</sub> <250 nM) and is in



**Figure 4**

Chemical structures of RLY-1971 (migoprotafib or GDC-1971) (*left*) and RLY-4008 (lirafugratinib) (*right*), two compounds progressed into Phase I clinical studies by Relay Therapeutics. While the exact bioactivity of RLY-4008 has not been disclosed, it lacks potency against FGFR1 and FGFR4 (174), which in turn may eliminate unwanted side effects. Abbreviations: CID, PubChem compound identifier; FGFR, fibroblast growth factor receptor.

Phase I clinical trials for the treatment of receptor tyrosine kinase (RTK)/RAS pathway-driven solid tumors (171). RLY-1971 was licensed by Genentech in December 2020 (172). RLY-4008, a highly selective, potent, and irreversible fibroblast growth factor receptor 2 (*FGFR2*) inhibitor, is effective in cholangiocarcinoma (173) (see also **Figure 4**).

## 5. CHALLENGES AND LIMITATIONS OF AI IN DRUG DISCOVERY

The phrase AI in drug discovery might give the impression that AI is successfully employed in early drug discovery. While there is some truth to this statement, it is important to note that no medicines approved by regulatory agencies can be attributed to AI in the same way that AI has achieved victories in chess, Go, *Jeopardy!*, autonomous vehicles, poetry generation, and protein folding. Drug discovery is a complex, multifaceted process, as captured by the 4DM charts. Robot Scientist Adam independently conducted genomic experiments (175), and Robot Scientist Eve performed an HTS campaign to identify antimalarial compounds (176). Although computer-aided processes have been used for compound selection and optimization, no AI-driven Robot Scientist or digital equivalent currently exists that can execute fully automated drug discovery. Automated AI-driven drug discovery remains an aspirational goal (177). Most success stories to date have relied on ML, cheminformatics, bioinformatics software, natural language processing, or other computational platforms that support human decision making. In summary, drug discovery has yet to benefit from a comprehensive AI system.

One of the weak aspects of AI4DD for small molecules is the training of ML models that encode chemical features (often referred to as QSARs), such as those derived from chemical structures. Bohacek et al. (178) estimated the number of drug-like chemicals to be up to  $10^{60}$ , and virtual screening libraries have already exceeded 36 billion compounds (95). The logistical and practical challenges of virtually screening 36 billion compounds, considering an estimated 10–50 conformers per molecule, amounting to over 500 billion objects, are immense and beyond

the scope of this review. Instead, our focus is on the practical issues related to the applicability domain (179) and external predictivity validation (180). Both validation and applicability are challenges faced by target-based KG ML models, as mentioned above.

ML models commonly used in AI4DD are often trained on tens of thousands of compounds or less, which raises questions about their effectiveness in sampling the chemical space of 36 billion compounds. Can we confidently assume that such comparatively small training sets effectively represent the chemical space of 36 billion? Are such predictions trustworthy? Both the applicability domain (a representational issue in ML feature space) and chemical diversity (unseen scaffolds are less likely to produce reliable predictions even within the applicability domain) raise concerns about the predictivity of ML models for the unexplored chemical universe. Ideally, AI4DD practitioners would want to systematically sample chemical space using adequately trained ML models. This becomes imperative during lead optimization, where progress depends on the accurate representation of relevant chemical scaffolds in the ML space.

Target identification in drug discovery is also impacted by the so-called reproducibility crisis (181). Bayer (182) and Amgen (183) have reported low reproducibility rates (33% and 11%, respectively) of high-impact publications, and many biomedical publications are false (184). From an AI perspective, filtering out false data requires a coordinated community effort. Lessons from *eLife*'s Reproducibility Project (185), which focused on cancer biology, highlight issues such as weaker-than-previously-published (186) effects and inaccurate protocol descriptions (187), among others. The possibility of indexing fabricated publications (188) or those generated by research paper mills (189) further increases the likelihood of false information in the field. These challenges with experimental data compound the issues of ML model accuracy and the ML science reproducibility crisis (77). For AI4DD to be effective, it needs to be anchored in truth.

Another subtle risk involves the education of scientists in using AI4DD models effectively. Questions about when, how, and in what order to deploy ML models are crucial. The proper use of ML models depends on the specific requirements of each unique drug discovery project. For some projects, target selectivity and appropriate in-tissue delivery might be more important than absolute affinity or systemic toxicity. In contrast, other projects might focus on mitigating on-target toxicity, low permeability, or scaffold similarity to competitor patents. Each issue demands different computational solutions, ranging from filters and lead hopping to sequential ML model deployment. Proper training in using AI4DD models is critical to ensure that scientists can effectively navigate these complexities and make informed decisions based on the specific needs of their drug discovery projects.

The human component of drug discovery is another crucial aspect. In many academic and industrial settings, decision making falls to medicinal chemists who typically rely on their judgment to propose compounds for the design-make-test (DMT) cycle rather than depending solely on AI (190). Compounds may be thoroughly evaluated by the project team, with members voting on the order in which chemicals should be synthesized and tested. In AI-integrated companies, AI may influence this process, but chemists are still likely to veto compounds that do not meet specific criteria, even if computational chemists or toxicologists find no issues. It is reasonable to assume that user expertise, bias, and time constraints play a significant role in early drug discovery, often more so than AI. The Pfizer Rule of 5 (Ro5, or Lipinski rules) serves as an early example (191) of attempts to integrate informatics and data science into the early stages of drug discovery. The Ro5 criteria, assessing hydrogen bonding capacity, the calculated octanol/water partitioning coefficient (logP), and molecular weight have been employed world-wide to narrow down the chemical solution space. It is undeniable that Ro5 criteria have had a significant impact on medicinal chemistry (192). However, the influence of these criteria is gradually diminishing over time (193).



## 6. CONCLUSIONS AND FUTURE OUTLOOK

Given the knowledge deficit caused by the constantly increasing volume of information, AI systems capable of independent reasoning are becoming a necessity. AI has made significant advances in disease diagnosis and health-care professions such as radiology, pathology, and clinical pharmacology. Although nosology remains a challenge for both humans and AI systems, resources such as Mondo demonstrate the potential for human and computer collaboration. In target identification, KG-based ML can structure, process, and evaluate large-scale data sets, with many potential applications in biomedicine. However, KG-based ML models face challenges such as data leakage and the absence of ground truth, which may contribute to increased variance in research. Integrating generative chemistry, ML, and MPO techniques such as Pareto optimization is emerging as an integral component of AI methods for hit and lead discovery, with strong support from community-based platforms that provide data sets and resources for AI4DD technologies. Significant advances are taking place in predictive toxicology, and regulatory agencies are positioning themselves to benefit from AI technologies. We also discussed several AI-designed compounds that have entered Phase I and Phase II clinical trials. These success stories rely on computational platforms to support human decision making, and AI has yet to deliver a comprehensive system for fully automated drug discovery. ML models used in AI4DD face challenges related to applicability domain, external predictivity validation, and chemical diversity. Proper training in using AI4DD models is crucial to ensure effective navigation of complexities and informed decision making in drug discovery projects, with human expertise playing a significant role.

The reproducibility crisis—both for experiments and ML models—suggests that the scientific community is responsible for laying down the foundations of verity and carefully vetting what is known. Modern science has no “we hold these truths to be self-evident” basis. Biology does not appear to lend itself to an axiomatically built foundation. But managing AI4DD systems implies knowledge and model updates, and we collectively share the responsibility of sifting truthful science from questionable results. The true strength of any AI4DD system lies in its ability to process and comprehend sparse data more efficiently and effectively than humans. Computers are better suited to process and interpret complex molecular and KG representations compared to humans. However, this strength can only be revealed when true data are provided. The adage “garbage in, garbage out” is a stark reminder of the need for ground truth across all data types and relationships in drug discovery.

One of the potential dangers is the misguided use of AI4DD tools. The prevailing notion is that feeding ML models with massive data can produce actionable results. This approach, destined to fail, needs rethinking. A better scenario is to allow AI to steer the entire drug discovery process, particularly in structure-based projects or when there is a wealth of bioactivity data. Human intervention could occur only at the very late stages of the pipeline. By fully leveraging the potential of AI, the drug discovery process can be significantly optimized and revolutionized.

## 7. EPILOGUE

ChatGPT, the conversational AI that has passed the US Medical Licensing Examination, has the potential “to revolutionize research practices and publishing” (194, p. 224). The GPT-4 technical report from March 14, 2023, demonstrates how GPT-4 could be used to create new drugs, among other applications (195). In one use-case scenario, Andrew White, a member of OpenAI’s red team, prompted GPT-4 with the name dasatinib, a kinase inhibitor drug. GPT-4 was asked to modify the drug and find novel, nonpatented molecules with a similar mode of action, locate chemical vendors selling the compound, and purchase it. If custom synthesis was needed, GPT-4 was to

email a contract research organization to order the compound. The prompt, instructions, and compound chosen by GPT-4 can be found on page 59 of the OpenAI manuscript (195).

Notable observations from the experiment include the following: (a) GPT-4 generated a valid chemical structure [SMILES (89)] output, indicating its ability to perceive and modify chemical structures correctly; (b) the molecule is available in the ZINC database (125), meaning that it is synthetically feasible; (c) the proposed molecule is desmethyl-imatinib, an *N*-dealkylated piperidine metabolite of imatinib, another protein kinase inhibitor drug; and (d) the proposed molecule, however, is not novel (see Google Patents; <https://bit.ly/norimatinib>). GPT-4 successfully modified the molecule while retaining its kinase inhibitor properties. Experimental validation is needed to confirm whether the GPT-4-generated molecule shares the same mode of action as dasatinib. It is essential to remember that GPT-4 has general expertise and is not specifically tailored for drug discovery. By providing access to external chemistry tools, White and colleagues (196) have designed ChemCrow, a GPT-4-based tool that demonstrates how appropriate technologies can significantly enhance LLM effectiveness in chemistry-related tasks. If more external resources are integrated with GPT-4 or its successors, LLMs could become a valid approach.

In the DreamWorks animated movie *Shrek*, Donkey keeps asking, “Are we there yet?,” and Shrek and Fiona keep saying “No!” To those waiting for AI4DD: We are not there yet, but we are on our way!

## DISCLOSURE STATEMENT

C.H. is a full-time employee of Genentech, Inc. T.I.O. is a full-time employee of Expert Systems Inc. T.I.O. is a member of the Scientific Advisory Board for Insilico Medicine, one of the companies featured in this review. Part of this work was supported by National Institutes of Health grant NCI U24 224370.

## ACKNOWLEDGMENTS

We would like to thank Melissa Landon, Alexander Tropsha, John Overington, Miguel Costa Coelho, Christopher Southan, and Andrew White for discussions about this topic.

## LITERATURE CITED

1. Wagner J, Dahlem AM, Hudson LD, Terry SF, Altman RB, et al. 2018. A dynamic map for learning, communicating, navigating and improving therapeutic development. *Nat. Rev. Drug Discov.* 17(2):150
2. Wagner JA, Dahlem AM, Hudson LD, Terry SF, Altman RB, et al. 2018. Application of a dynamic map for learning, communicating, navigating, and improving therapeutic development. *Clin. Transl. Sci.* 11(2):166–74
3. Tambuyzer E, Vandendriessche B, Austin CP, Brooks PJ, Larsson K, et al. 2020. Therapies for rare diseases: therapeutic modalities, progress and challenges ahead. *Nat. Rev. Drug Discov.* 19(2):93–111
4. Zhu H. 2020. Big data and artificial intelligence modeling for drug discovery. *Annu. Rev. Pharmacol. Toxicol.* 60:573–89
5. Jayatunga MKP, Xie W, Ruder L, Schulze U, Meier C. 2022. AI in small-molecule drug discovery: a coming wave? *Nat. Rev. Drug Discov.* 21(3):175–76
6. Bentwich I. 2023. Pharma’s bio-AI revolution. *Drug Discov. Today* 28(5):103515
7. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, et al. 2019. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* 18(6):463–77
8. Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, et al. 2020. Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug Discov.* 19(5):353–64
9. Bender A, Cortés-Ciriano I. 2021. Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 1: ways to make an impact, and why we are not there yet. *Drug Discov. Today* 26(2):511–24

10. Bender A, Cortes-Ciriano I. 2021. Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 2: a discussion of chemical and biological data. *Drug Discov. Today* 26(4):1040–52
11. Harris LA. 2021. *Artificial intelligence: background, selected issues, and policy considerations*. Rep. R46795, Congr. Res. Serv., Washington, DC. <https://crsreports.congress.gov/product/pdf/R/R46795>
12. Duran-Frigola M, Cigler M, Winter GE. 2023. Advancing targeted protein degradation via multiomics profiling and artificial intelligence. *J. Am. Chem. Soc.* 145(5):2711–32
13. Kim J, McFee M, Fang Q, Abdin O, Kim PM. 2023. Computational and artificial intelligence-based methods for antibody development. *Trends Pharmacol. Sci.* 44(3):175–89
14. Ichikawa DM, Abdin O, Alerasool N, Kogenaru M, Mueller AL, et al. 2023. A universal deep-learning model for zinc finger design enables transcription factor reprogramming. *Nat. Biotechnol.* 41:1117–29
15. Sciabola S, Xi H, Cruz D, Cao Q, Lawrence C, et al. 2021. PFRED: a computational platform for siRNA and antisense oligonucleotides design. *PLOS ONE* 16(1):e0238753
16. McCaffrey P. 2022. Artificial intelligence for vaccine design. *Methods Mol. Biol.* 2412:3–13
17. Hey T, Tansley S, Tolle K, eds. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, WA: Microsoft Res.
18. Tecuci G, Marcu D, Boicu M, Schum DA. 2016. *Knowledge Engineering: Building Cognitive Assistants for Evidence-Based Reasoning*. Cambridge, UK: Cambridge Univ. Press
19. Borthwick A, Sterling J, Agichtein E, Grishman R. 1998. Exploiting diverse knowledge sources via maximum entropy in named entity recognition. In *Proceedings of the Sixth Workshop on Very Large Corpora*. <https://aclanthology.org/W98-1118.pdf>
20. Grissa D, Junge A, Oprea TI, Jensen LJ. 2022. Diseases 2.0: a weekly updated database of disease-gene associations from text mining and data integration. *Database* 2022:baac019
21. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, et al. 2023. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* 51(D1):D638–46
22. Ochoa D, Hercules A, Carmona M, Suveges D, Baker J, et al. 2023. The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res.* 51(D1):D1353–59
23. Kelleher KJ, Sheils TK, Mathias SL, Yang JJ, Metzger VT, et al. 2023. Pharos 2023: an integrated resource for the understudied human proteome. *Nucleic Acids Res.* 51(D1):D1405–16
24. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873):583–89
25. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, et al. 2021. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373(6557):871–76
26. Bartusiak M. 1981. Designing drugs with computers. *Discover Magazine*, Aug., pp. 47–50
27. Bylinsky G. 1981. A new industrial revolution is on the way. *Fortune Magazine*, Oct. 5, pp. 106–14
28. Ajay A, Walters WP, Murcko MA. 1998. Can we learn to distinguish between “drug-like” and “nondrug-like” molecules? *J. Med. Chem.* 41(18):3314–24
29. Sadowski J, Kubinyi H. 1998. A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* 41(18):3325–29
30. Ursu O, Rayan A, Goldblum A, Oprea TI. 2011. Understanding drug-likeness. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 1(5):760–81
31. Schuster D, Laggner C, Langer T. 2005. Why drugs fail—a study on side effects in new chemical entities. *Curr. Pharm. Des.* 11(27):3545–59
32. Mak K-K, Balijepalli MK, Pichika MR. 2022. Success stories of AI in drug discovery - where do things stand? *Expert Opin. Drug Discov.* 17(1):79–92
33. Kumar Y, Koul A, Singla R, Ijaz MF. 2022. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *J. Ambient Intell. Humaniz. Comput.* 14(7):8459–86
34. Langlotz CP. 2019. Will artificial intelligence replace radiologists? *Radiol. Artif. Intell.* 1(3):e190058
35. Chauhan C, Gullapalli RR. 2021. Ethics of AI in pathology: current paradigms and emerging issues. *Am. J. Pathol.* 191(10):1673–83
36. Zhavoronkov A, Vanhaelen Q, Oprea TI. 2020. Will artificial intelligence for drug discovery impact clinical pharmacology? *Clin. Pharmacol. Ther.* 107(4):780–85

37. Levin JM, Oprea TI, Davidovich S, Clozel T, Overington JP, et al. 2020. Artificial intelligence, drug repurposing and peer review. *Nat. Biotechnol.* 38(10):1127–31
38. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, et al. 2023. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit. Health* 2(2):e0000198
39. Chen M, Decary M. 2020. Artificial intelligence in healthcare: an essential guide for health leaders. *Healthc. Manag. Forum* 33(1):10–18
40. Cui M, Zhang DY. 2021. Artificial intelligence and computational pathology. *Lab. Investig.* 101(4):412–22
41. Baxi V, Edwards R, Montalto M, Saha S. 2022. Digital pathology and artificial intelligence in translational medicine and clinical practice. *Mod. Pathol.* 35(1):23–32
42. Yasaka K, Abe O. 2018. Deep learning and artificial intelligence in radiology: current applications and future directions. *PLOS Med.* 15(11):e1002707
43. Makino T, Jastrzębski S, Oleszkiewicz W, Chacko C, Ehrenpreis R, et al. 2022. Differences between human and machine perception in medical diagnosis. *Sci. Rep.* 12(1):6877
44. Yanagisawa Y, Shido K, Kojima K, Yamasaki K. 2023. Convolutional neural network-based skin image segmentation model to improve classification of skin diseases in conventional and non-standardized picture images. *J. Dermatol. Sci.* 109(1):30–36
45. Haendel MA, McMurry JA, Relevo R, Mungall CJ, Robinson PN, Chute CG. 2018. A census of disease ontologies. *Annu. Rev. Biomed. Data Sci.* 1:305–31
46. Vasilevsky NA, Matentzoglou NA, Toro S, Flack JE IV, Hegde H, et al. 2022. Mondo: unifying diseases for the world, by the world. medRxiv 2022.04.13.22273750. <https://doi.org/10.1101/2022.04.13.22273750>
47. Haendel M, Vasilevsky N, Unni D, Bologa C, Harris N, et al. 2020. How many rare diseases are there? *Nat. Rev. Drug Discov.* 19(2):77–78
48. Avram S, Wilson TB, Curpan R, Halip L, Borota A, et al. 2023. DrugCentral 2023 extends human clinical data and integrates veterinary drugs. *Nucleic Acids Res.* 51(D1):D1276–87
49. Nelson SJ, Oprea TI, Ursu O, Bologa CG, Zaveri A, et al. 2017. Formalizing drug indications on the road to therapeutic intent. *J. Am. Med. Inform. Assoc.* 24(6):1169–72
50. Moodley K, Rieswijk L, Oprea TI, Dumontier M. 2021. InContext: curation of medical context for drug indications. *J. Biomed. Semant.* 12(1):2
51. Avram S, Halip L, Curpan R, Borota A, Bora A, Oprea TI. 2022. Annotating off-label drug usage from unconventional sources. medRxiv 2022.09.08.22279709. <https://doi.org/10.1101/2022.09.08.22279709>
52. Oprea TI, Bologa CG, Brunak S, Campbell A, Gaulton A, et al. 2017. Unexplored therapeutic opportunities in the human genome. *Nat. Rev. Drug Discov.* 17(5):317–32
53. Zeng X, Zhu S, Lu W, Liu Z, Huang J, et al. 2020. Target identification among known drugs by deep learning from heterogeneous networks. *Chem. Sci.* 11(7):1775–97
54. Ye Q, Hsieh C-Y, Yang Z, Kang Y, Chen J, et al. 2021. A unified drug-target interaction prediction framework based on knowledge graph and recommendation system. *Nat. Commun.* 12(1):6775
55. Sun Y, Barber R, Gupta M, Aggarwal CC, Han J. 2011. Co-author relationship prediction in heterogeneous bibliographic networks. In *2011 International Conference on Advances in Social Networks Analysis and Mining*, pp. 121–28. Piscataway, NJ: IEEE
56. Himmelstein DS, Baranzini SE. 2015. Heterogeneous network edge prediction: a data integration approach to prioritize disease-associated genes. *PLOS Comput. Biol.* 11(7):e1004259
57. Fu G, Ding Y, Seal A, Chen B, Sun Y, Bolton E. 2016. Predicting drug target interactions using meta-path-based semantic network analysis. *BMC Bioinform.* 17:160
58. Chen T, Guestrin C. 2016. XGBoost: a scalable tree boosting system. In *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–94. New York: ACM
59. Binder J, Ursu O, Bologa C, Jiang S, Maphis N, et al. 2022. Machine learning prediction and tau-based screening identifies potential Alzheimer's disease genes relevant to immunity. *Commun. Biol.* 5(1):125
60. Vojtechova I, Machacek T, Kristofikova Z, Stuchlik A, Petrasek T. 2022. Infectious origin of Alzheimer's disease: amyloid beta as a component of brain antimicrobial immunity. *PLOS Pathog.* 18(11):e1010929

61. Klionsky DJ, Petroni G, Amaravadi RK, Baehrecke EH, Ballabio A, et al. 2021. Autophagy in major human diseases. *EMBO J.* 40(19):e108863
62. Homma K, Suzuki K, Sugawara H. 2011. The Autophagy Database: an all-inclusive information resource on autophagy that provides nourishment for research. *Nucleic Acids Res.* 39(Database issue):D986–90
63. Lamb CA, Yoshimori T, Tooze SA. 2013. The autophagosome: origins unknown, biogenesis complex. *Nat. Rev. Mol. Cell Biol.* 14(12):759–74
64. Ranjbar M, Yang JJ, Kumar P, Byrd DR, Bearer EL, Oprea TI. 2023. Autophagy dark genes: Can we find them with machine learning? *Nat. Sci.* 3:e20220067
65. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 47(D1):D419–26
66. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, et al. 2020. The reactome pathway knowledgebase. *Nucleic Acids Res.* 48(D1):D498–503
67. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. 2015. Omim.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* 43(Database issue):D789–98
68. Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, et al. 2019. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res.* 47(D1):D1018–27
69. Groza T, Gomez FL, Mashhadi HH, Muñoz-Fuentes V, Gunes O, et al. 2023. The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Res.* 51(D1):D1038–45
70. Nusinow DP, Szpyt J, Ghandi M, Rose CM, McDonald ER 3rd, et al. 2020. Quantitative proteomics of the Cancer Cell Line Encyclopedia. *Cell* 180(2):387–402.e16
71. UniProt Consort. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49(D1):D480–89
72. Gene Ontology Consort. 2019. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47(D1):D330–D338
73. Ramilowski JA, Goldberg T, Harshbarger J, Kloppmann E, Lizio M, et al. 2015. A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat. Commun.* 6(1):7866
74. Karlsson M, Zhang C, Mear L, Zhong W, Digre A, et al. 2021. A single-cell type transcriptomics map of human tissues. *Sci. Adv.* 7(31):eabh2169
75. Samala RK, Chan H-P, Hadjiiski L, Koneru S. 2020. Hazards of data leakage in machine learning: a study on classification of breast cancer using deep neural networks. In *Proceedings of the SPIE Medical Imaging 2020: Computer-Aided Diagnosis*. Bellingham, WA: SPIE. <https://doi.org/10.1117/12.2549313>
76. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. 2023. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* 51(D1):D587–92
77. Kapoor S, Narayanan A. 2022. Leakage and the reproducibility crisis in ML-based science. *arXiv:2207.07048 [cs.LG]*
78. Hansch C, Fujita T. 1964.  $\rho$ - $\sigma$ - $\pi$  Analysis. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* 86(8):1616–26
79. Hansch C, Leo A, Hoekman DH. 1995. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*. Washington, DC: Am. Chem. Soc.
80. Todeschini R, Consonni V. 2009. *Molecular Descriptors for Chemoinformatics*. Weinheim, Ger.: Wiley-VCH
81. Muratov EN, Bajorath J, Sheridan RP, Tetko IV, Filimonov D, et al. 2020. QSAR without borders. *Chem. Soc. Rev.* 49(11):3525–64
82. Muratov EN, Amaro R, Andrade CH, Brown N, Ekins S, et al. 2021. A critical overview of computational approaches employed for COVID-19 drug discovery. *Chem. Soc. Rev.* 50(16):9121–51
83. Olah M, Bologna C, Oprea TI. 2004. An automated PLS search for biologically relevant QSAR descriptors. *J. Comput. Aided Mol. Des.* 18(7):437–49
84. Davies M, Jones RDO, Grime K, Jansson-Löfmark R, Fretland AJ, et al. 2020. Improving the accuracy of predicted human pharmacokinetics: lessons learned from the AstraZeneca drug pipeline over two decades. *Trends Pharmacol. Sci.* 41(6):390–408

85. Göller AH, Kuhnke L, Montanari F, Bonin A, Schneckener S, et al. 2020. Bayer's in silico ADMET platform: a journey of machine learning over the past two decades. *Drug Discov. Today* 25(9):1702–9
86. Sheridan RP. 2022. Stability of prediction in production ADMET models as a function of version: why and when predictions change. *J. Chem. Inf. Model.* 62(15):3477–85
87. Chen EP, Bondi RW, Zhang C, Price DJ, Ho M-H, et al. 2022. Applications of model-based target pharmacology assessment in defining drug design and DMPK strategies: GSK experiences. *J. Med. Chem.* 65(9):6926–39
88. Tetko I, Oprea TI. 2008. Early ADME/T predictions: toy or tool? In *Cheminformatics Approaches to Virtual Screening*, ed. A Tropsha, A Varnek, pp. 240–67. Cambridge, UK: Royal Soc. Chem.
89. Weininger D. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* 28(1):31–36
90. Duvenaud D, Maclaurin D, Aguilera-Iparraguirre J, Gómez-Bombarelli R, Hirzel T, et al. 2015. Convolutional networks on graphs for learning molecular fingerprints. arXiv:1509.09292 [cs.LG]
91. Zitnik M, Agrawal M, Leskovec J. 2018. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 34(13):i457–66
92. Torng W, Altman RB. 2019. Graph convolutional neural networks for predicting drug-target interactions. *J. Chem. Inf. Model.* 59(10):4131–49
93. Jiang D, Wu Z, Hsieh C-Y, Chen G, Liao B, et al. 2021. Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models. *J. Cheminform.* 13(1):12
94. Vanhaelen Q, Lin Y-C, Zhavoronkov A. 2020. The advent of generative chemistry. *ACS Med. Chem. Lett.* 11(8):1496–505
95. Lyu J, Irwin JJ, Shoichet BK. 2023. Modeling the expansion of virtual screening libraries. *Nat. Chem. Biol.* 19(6):712–18
96. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, et al. 2014. Generative adversarial networks. arXiv:1406.2661 [stat.ML]
97. Kadurin A, Aliper A, Kazennov A, Mamoshina P, Vanhaelen Q, et al. 2017. The cornucopia of meaningful leads: applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget* 8(7):10883–90
98. Natl. Cancer Inst. 2023. *NCI-60 growth inhibition data*. Database, Natl. Cancer Inst., Bethesda, MD, updated July 2. <https://wiki.nci.nih.gov/display/NCIDTPdata/NCI-60+Growth+Inhibition+Data>
99. Kim S, Chen J, Cheng T, Gindulyte A, He J, et al. 2023. PubChem 2023 update. *Nucleic Acids Res.* 51(D1):D1373–80
100. Polykovskiy D, Zhebrak A, Vetrov D, Ivanenkov Y, Aladinskiy V, et al. 2018. Entangled conditional adversarial autoencoder for de novo drug discovery. *Mol. Pharm.* 15(10):4398–405
101. Hwang C-L, Masud AS, Paidy SR, Yoon K. 1979. *Multiple Objective Decision Making—Methods and Applications: A State-of-the-Art Survey*. Heidelberg: Springer-Verlag
102. Hopfinger AJ. 1985. Computer-assisted drug design. *J. Med. Chem.* 28(9):1133–39
103. Fromer JC, Coley CW. 2022. Computer-aided multi-objective optimization in small molecule discovery. *Patterns* 4(2):100678
104. Oprea TI, Zamora I, Ungell A-L. 2002. Pharmacokinetically based mapping device for chemical space navigation. *J. Comb. Chem.* 4(4):258–66
105. Bleicher KH, Böhm H-J, Müller K, Alanine AI. 2003. Hit and lead generation: beyond high-throughput screening. *Nat. Rev. Drug Discov.* 2(5):369–78
106. Zamora I, Oprea TI, Cruciani G, Pastor M, Ungell AL. 2003. Surface descriptors for protein–ligand affinity prediction. *J. Med. Chem.* 46(1):25–33
107. Besnard J, Ruda GF, Setola V, Abecassis K, Rodriguiz RM, et al. 2012. Automated design of ligands to polypharmacological profiles. *Nature* 492(7428):215–20
108. Rogers D, Brown RD, Hahn M. 2005. Using extended-connectivity fingerprints with Laplacian-modified Bayesian analysis in high-throughput screening follow-up. *J. Biomol. Screen.* 10(7):682–86
109. Paolini GV, Shapland RHB, van Hoorn WP, Mason JS, Hopkins AL. 2006. Global mapping of pharmacological space. *Nat. Biotechnol.* 24(7):805–15



110. Deb K, Sundar J, Udaya Bhaskara Rao N, Chaudhuri S. 2006. Reference point based multi-objective optimization using evolutionary algorithms. *Int. J. Comput. Intell. Res.* 2:273–86
111. Zhu Z, Shi C, Zhang Z, Liu S, Xu M, et al. 2022. Torchdrug: a powerful and flexible machine learning platform for drug discovery. arXiv:2202.08320 [cs.LG]
112. Paszke A, Gross S, Chintala S, Chanan G, Yang E, et al. 2017. *Automatic differentiation in PyTorch*. Paper presented at the 31<sup>st</sup> Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, Dec. 4–9
113. Liu S, Wang H, Liu W, Lasenby J, Guo H, Tang J. 2021. Pre-training molecular graph representation with 3D geometry. arXiv:2110.07728 [cs.LG]
114. Zhu Z, Galkin M, Zhang Z, Tang J. 2022. Neural-symbolic models for logical queries on knowledge graphs. arXiv:2205.10128 [cs.AI]
115. Blaschke T, Arús-Pous J, Chen H, Margreitter C, Tyrchan C, et al. 2020. REINVENT 2.0: an AI tool for de novo drug design. *J. Chem. Inf. Model.* 60(12):5918–22
116. Blaschke T, Engkvist O, Bajorath J, Chen H. 2020. Memory-assisted reinforcement learning for diverse molecular de novo design. *J. Cheminform.* 12(1):68
117. Cummins DJ, Bell MA. 2016. Integrating everything: the molecule selection toolkit, a system for compound prioritization in drug discovery. *J. Med. Chem.* 59(15):6999–7010
118. Landrum G. 2022. RDKit: open-source cheminformatics software. *RDKit*. <https://www.rdkit.org/>
119. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, et al. 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12(85):2825–30
120. Abadi M, Agarwal A, Barham P. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv:1603.04467 [cs.DC]
121. Zhavoronkov A, Ivanenkov YA, Aliper A, Veselov MS, Aladinskiy VA, et al. 2019. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* 37(9):1038–40
122. Fu H-L, Valiathan RR, Arkwright R, Sohail A, Mihai C, et al. 2013. Discoidin domain receptors: unique receptor tyrosine kinases in collagen-mediated signaling. *J. Biol. Chem.* 288(11):7430–37
123. Mehta V, Chander H, Munshi A. 2021. Complex roles of discoidin domain receptor tyrosine kinases in cancer. *Clin. Transl. Oncol.* 23(8):1497–510
124. Moll S, Desmoulière A, Moeller MJ, Pache J-C, Badi L, et al. 2019. DDR1 role in fibrosis and its pharmacological targeting. *Biochim. Biophys. Acta Mol. Cell Res.* 1866(11):118474
125. Tingle BI, Tang KG, Castanon M, Gutierrez JJ, Khurelbaatar M, et al. 2023. ZINC-22—a free multi-billion-scale database of tangible compounds for ligand discovery. *J. Chem. Inf. Model.* 63(4):1166–76
126. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, et al. 2017. The ChEMBL database in 2017. *Nucleic Acids Res.* 45(D1):D945–54
127. Tse EG, Aithani L, Anderson M, Cardoso-Silva J, Cincilla G, et al. 2021. An open drug discovery competition: experimental validation of predictive models in a series of novel antimalarials. *J. Med. Chem.* 64(22):16450–63
128. Huang K, Fu T, Gao W, Zhao Y, Roohani Y, et al. 2022. Artificial intelligence foundation for therapeutic science. *Nat. Chem. Biol.* 18(10):1033–36
129. Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, et al. 2018. MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.* 9(2):513–30
130. Méndez-Lucio O, Nicolaou C, Earnshaw B. 2022. MolE: a molecular foundation model for drug discovery. arXiv:2211.02657 [q-bio.QM]
131. Sushko I, Novotarskyi S, Körner R, Pandey AK, Rupp M, et al. 2011. Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. *J. Comput. Aided Mol. Des.* 25(6):533–54
132. Sosnin S, Vashurina M, Withnall M, Karpov P, Fedorov M, Tetko IV. 2019. A survey of multi-task learning methods in chemoinformatics. *Mol. Inform.* 38(4):e1800108
133. Varnek A, Gaudin C, Marcou G, Baskin I, Pandey AK, Tetko IV. 2009. Inductive transfer of knowledge: application of multi-task learning and feature net approaches to model tissue-air partition coefficients. *J. Chem. Inf. Model.* 49(1):133–44

134. Zakharov AV, Zhao T, Nguyen D-T, Peryea T, Sheils T, et al. 2019. Novel consensus architecture to improve performance of large-scale multitask deep learning QSAR models. *J. Chem. Inf. Model.* 59(11):4613–24
135. Govinda KCB, Bocci G, Verma S, Hassan MM, Holmes J, et al. 2021. A machine learning platform to estimate anti-SARS-CoV-2 activities. *Nat. Mach. Intell.* 3(6):527–35
136. Siramshetty V, Williams J, Nguyễn Đ-T, Neyra J, Southall N, et al. 2021. Validating ADME QSAR models using marketed drugs. *SLAS Discov.* 26(10):1326–36
137. Xiong G, Wu Z, Yi J, Fu L, Yang Z, et al. 2021. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.* 49(W1):W5–14
138. Renz P, Van Rompaey D, Wegner JK, Hochreiter S, Klambauer G. 2019. On failure modes in molecule generation and optimization. *Drug Discov. Today Technol.* 32–33:55–63
139. Walters WP, Murcko M. 2020. Assessing the impact of generative AI on medicinal chemistry. *Nat. Biotechnol.* 38(2):143–45
140. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, et al. 2020. A deep learning approach to antibiotic discovery. *Cell* 180(4):688–702.e13
141. Lemonick S. 2020. AI finds molecules that kill bacteria, but would they make good antibiotics? *Chemical & Engineering News*, Feb. 26. <https://cen.acs.org/physical-chemistry/computational-chemistry/AI-finds-molecules-kill-bacteria/98/web/2020/02>
142. Hu F, Santagostino SF, Danilenko DM, Tseng M, Brumm J, et al. 2022. Assessment of skin toxicity in an in vitro reconstituted human epidermis model using deep learning. *Am. J. Patbol.* 192(4):687–700
143. Hasselgren C, Myatt GJ. 2018. Computational toxicology and drug discovery. *Methods Mol. Biol.* 1800:233–44
144. Vo AH, Van Vleet TR, Gupta RR, Liguori MJ, Rao MS. 2020. An overview of machine learning and big data for drug toxicity evaluation. *Chem. Res. Toxicol.* 33(1):20–37
145. Liu A, Seal S, Yang H, Bender A. 2023. Using chemical and biological data to predict drug toxicity. *SLAS Discov.* 28(3):53–64
146. Gardiner L-J, Carrieri AP, Wilshaw J, Checkley S, Pyzer-Knapp EO, Krishna R. 2020. Using human in vitro transcriptome analysis to build trustworthy machine learning models for prediction of animal drug toxicity. *Sci. Rep.* 10(1):9522
147. Whitehead TM, Irwin BWJ, Hunt P, Segall MD, Conduit GJ. 2019. Imputation of assay bioactivity data using deep learning. *J. Chem. Inf. Model.* 59(3):1197–204
148. Williams DP, Lazic SE, Foster AJ, Semenova E, Morgan P. 2020. Predicting drug-induced liver injury with Bayesian machine learning. *Chem. Res. Toxicol.* 33(1):239–48
149. Johnson M, Patel M, Phipps A, van der Schaar M, Boulton D, Gibbs M. 2023. The potential and pitfalls of artificial intelligence in clinical pharmacology. *CPT Pharmacometrics Syst. Pharmacol.* 12(3):279–84
150. Bean DM, Wu H, Iqbal E, Dzahini O, Ibrahim ZM, et al. 2017. Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records. *Sci. Rep.* 7(1):16416
151. Int. Counc. Harmon. Tech. Requir. Pharm. Hum. Use (ICH). 2023. M7: assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk. Multidiscip. Guidel., ICH, Geneva. <https://ich.org/page/multidisciplinary-guidelines#7-2>
152. EMA (Eur. Med. Agency). 2018. *Reflection paper on the qualification of non-genotoxic impurities*. Rep., EMA, London. [https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-qualification-non-genotoxic-impurities\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-qualification-non-genotoxic-impurities_en.pdf)
153. Hines PA, Herold R, Pinheiro L, Frias Z, Arlett P. 2022. Artificial intelligence in European medicines regulation. *Nat. Rev. Drug Discov.* <https://www.nature.com/articles/d41573-022-00190-3>
154. FDA (US Food Drug Admin.). 2023. *Using artificial intelligence and machine learning in the development of drug and biological products*. Discuss. Pap., FDA, Silver Spring, MD. <https://www.fda.gov/media/167973/download?attachment>
155. FDA (US Food Drug Admin.). 2023. *Artificial intelligence in drug manufacturing*. Discuss. Pap., Cent. Drug Eval. Res., FDA, Silver Spring, MD. <https://www.fda.gov/media/165743/download?attachment>
156. Myatt GJ, Ahlberg E, Akahori Y, Allen D, Amberg A, et al. 2018. In silico toxicology protocols. *Regul. Toxicol. Pharmacol.* 96:1–17

157. Myatt GJ, Bassan A, Bower D, Crofton KM, Cross KP, et al. 2022. Increasing the acceptance of in silico toxicology through development of protocols and position papers. *Comput. Toxicol.* 21:100209
158. Bercu J, Masuda-Herrera MJ, Trejo-Martin A, Hasselgren C, Lord J, et al. 2021. A cross-industry collaboration to assess if acute oral toxicity (Q)SAR models are fit-for-purpose for GHS classification and labelling. *Regul. Toxicol. Pharmacol.* 120:104843
159. Graham JC, Trejo-Martin A, Chilton ML, Kostal J, Bercu J, et al. 2022. An evaluation of the occupational health hazards of peptide couplers. *Chem. Res. Toxicol.* 35(6):1011–22
160. Musuamba FT, Skottheim Rusten I, Lesage R, Russo G, Bursi R, et al. 2021. Scientific and regulatory evaluation of mechanistic in silico drug and disease models in drug development: building model credibility. *CPT Pharmacometr. Syst. Pharmacol.* 10(8):804–25
161. Exscientia. 2020. *DSP-1181*. Press Release, Jan. 30. <https://www.exscientia.ai/dsp-1181>
162. Yoshinaga H, Uemachi H, Ohno T, Besnard J. 2020. *2,6-Disubstituted pyridine derivative*. US Patent 10,800,755B2
163. Wills T. 2022. AI drug discovery: assessing the first AI-designed drug candidates to go into human clinical trials. *CAS*, Sept. 23. <https://www.cas.org/resources/cas-insights/drug-discovery/ai-designed-drug-candidates>
164. Yoshinaga H, Ikuma Y, Ikeda J, Adachi S, Mitsunuma H, et al. 2020. *Condensed lactam derivative*. US Patent 10,745,401B2
165. Kornauth C, Pemovska T, Vladimer GI, Bayer G, Bergmann M, et al. 2022. Functional precision medicine provides clinical benefit in advanced aggressive hematologic cancers and identifies exceptional responders. *Cancer Discov.* 12(2):372–87
166. Fultinavičiūtė U. 2023. Insilico's AI drug enters Phase II IPF trial. *Clinical Trials Arena*, June 27. <https://www.clinicaltrialsarena.com/news/insilico-medicine-ins018055-ai/>
167. Pun FW, Liu BHM, Long X, Leung HW, Leung GHD, et al. 2022. Identification of therapeutic targets for amyotrophic lateral sclerosis using PandaOmics—an AI-enabled biological target discovery platform. *Front. Aging Neurosci.* 14:914017
168. Ivanenkov YA, Polykovskiy D, Bezrukov D, Zagribelnyy B, Aladinskiy V, et al. 2023. Chemistry42: an AI-driven platform for molecular design and optimization. *J. Chem. Inf. Model.* 63(3):695–701
169. Aliper A, Aladinskiy V, Zavoronkovs A. 2020. *Methods of inhibiting kinases*. WO Patent 2020/170203
170. InSilico Med. 2023. *Hong Kong Stock Exchange application proof of InSilico Medicine Cayman TopCo*. Appl. Proof, InSilico Med., Hong Kong. <https://www1.hkexnews.hk/app/sehk/2023/105489/documents/sehk23062701667.pdf>
171. Williams B, Taylor A, Orozco O, Owen C, Kelley E, et al. 2022. Discovery and characterization of the potent, allosteric SHP2 inhibitor GDC-1971 for the treatment of RTK/RAS driven tumors. *Cancer Res.* 82(12\_Suppl.):3327 (Abstr.)
172. Relay Ther. 2020. *Relay Therapeutics® announces a worldwide license and collaboration agreement with Genentech for RLY-1971*. Press Release, Dec. 14. <https://ir.relaytx.com/news-releases/news-release-details/relay-therapeutics-announces-worldwide-license-and-collaboration>
173. Park JO, Wai Meng DT, Hollebecque A, Borad M, Goyal L, et al. 2022. 76MO efficacy of RLY-4008, a highly selective FGFR2 inhibitor in patients (pts) with an FGFR2-fusion or rearrangement (f/r), FGFR inhibitor (FGFRi)-naïve cholangiocarcinoma (CCA): ReFocus trial. *Ann. Oncol.* 33:S1461–62
174. Subbiah V, Sahai V, Maglic D, Bruderek K, Toure BB, et al. 2023. RLY-4008, the first highly selective FGFR2 inhibitor with activity across *FGFR2* alterations and resistance mutations. *Cancer Discov.* 13(9):2012–31
175. King RD, Rowland J, Oliver SG, Young M, Aubrey W, et al. 2009. The automation of science. *Science* 324(5923):85–89
176. Bilsland E, van Vliet L, Williams K, Feltham J, Carrasco MP, et al. 2018. *Plasmodium* dihydrofolate reductase is a second enzyme target for the antimalarial action of triclosan. *Sci. Rep.* 8(1):1038
177. Schneider G. 2017. Automating drug discovery. *Nat. Rev. Drug Discov.* 17(2):97–113
178. Bohacek RS, McMartin C, Guida WC. 1996. The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev.* 16(1):3–50
179. Tropsha A. 2010. Best practices for QSAR model development, validation, and exploitation. *Mol. Inform.* 29(6–7):476–88

180. Gramatica P, Sangion A. 2016. A historical excursus on the statistical validation parameters for QSAR models: a clarification concerning metrics and terminology. *J. Chem. Inf. Model.* 56(6):1127–31
181. Collins FS, Tabak LA. 2014. Policy: NIH plans to enhance reproducibility. *Nature* 505(7485):612–13
182. Prinz F, Schlange T, Asadullah K. 2011. Believe it or not: How much can we rely on published data on potential drug targets? *Nat. Rev. Drug Discov.* 10(9):712
183. Begley CG, Ellis LM. 2012. Drug development: raise standards for preclinical cancer research. *Nature* 483(7391):531–33
184. Ioannidis JPA. 2005. Why most published research findings are false. *PLOS Med.* 2(8):e124
185. Rodgers P, Collings A. 2021. What have we learned? *eLife* 10:e75830
186. Errington TM, Mathur M, Soderberg CK, Denis A, Perfito N, et al. 2021. Investigating the replicability of preclinical cancer biology. *eLife* 10:e71601
187. Errington TM, Denis A, Perfito N, Iorns E, Nosek BA. 2021. Challenges for assessing replicability in preclinical cancer biology. *eLife* 10:e67995
188. Brainard J. 2018. Rethinking retractions. *Science* 362(6413):390–93
189. Else H, Van Noorden R. 2021. The fight against fake-paper factories that churn out sham science. *Nature* 591(7851):516–19
190. Green DVS. 2019. Using machine learning to inform decisions in drug discovery: an industry perspective. In *Machine Learning in Chemistry: Data-Driven Algorithms, Learning Systems, and Predictions*, ed. EO Pyzer-Knapp, T Laino, pp. 81–101. Washington, DC: Am. Chem. Soc.
191. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23(1–3):3–25
192. Benet LZ, Hosey CM, Ursu O, Oprea TI. 2016. BDDCS, the rule of 5 and drugability. *Adv. Drug Deliv. Rev.* 101:89–98
193. O'Donovan DH, De Fusco C, Kuhnke L, Reichel A. 2023. Trends in molecular properties, bioavailability, and permeability across the Bayer compound collection. *J. Med. Chem.* 66(4):2347–60
194. van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL. 2023. ChatGPT: five priorities for research. *Nature* 614(7947):224–26
195. OpenAI. 2023. *GPT-4 technical report*. Rep., OpenAI, San Francisco, CA. <https://cdn.openai.com/papers/gpt-4.pdf>
196. Bran AM, Cox S, White AD, Schwaller P. 2023. ChemCrow: augmenting large-language models with chemistry tools. arXiv:2304.05376 [physics.chem-ph]



# Contents

Gene-Environment Interactions: My Unique Journey <i>Daniel W. Nebert</i> .....	1
Introduction to the Theme “Pharmacological Individuality: New Insights and Strategies for Personalized and Precise Drug Treatment” <i>Urs A. Meyer, Susan G. Amara, Terrence F. Blaschke, and Paul A. Insel</i> .....	27
Pharmacogenomics Beyond Single Common Genetic Variants: The Way Forward <i>Volker M. Lauschke, Yitian Zhou, and Magnus Ingelman-Sundberg</i> .....	33
Addressing Ancestry and Sex Bias in Pharmacogenomics <i>Manuel Corpas, Moneeza K. Siddiqui, Opeyemi Soremekun, Robini Mathur, Dipender Gill, and Segun Fatumo</i> .....	53
Circulating Biomarkers Instead of Genotyping to Establish Metabolizer Phenotypes <i>Roman Tremmel, Ute Hofmann, Mathias Haag, Elke Schaeffeler, and Matthias Schwab</i> .....	65
Circadian Regulation of Drug Responses: Toward Sex-Specific and Personalized Chronotherapy <i>Francis A. Lévi, Alper Okyar, Eva Hadadi, Pasquale F. Innominato, and Annabelle Ballesta</i> .....	89
Anthracycline Toxicity: Light at the End of the Tunnel? <i>Romina B. Cejas, Kateryna Petrykey, Yadav Sapkota, and Paul W. Burridge</i> .....	115
LP(a): Structure, Genetics, Associated Cardiovascular Risk, and Emerging Therapeutics <i>Erfan Tasdighi, Rishav Adbikari, Omar Almaadawy, Thorsten M. Leucker, and Michael J. Blaha</i> .....	135
Health Digital Twins in Life Science and Health Care Innovation <i>Kaushik P. Venkatesh, Gabriel Brito, and Maged N. Kamel Boulos</i> .....	159

The Nocebo Effect <i>Luana Colloca</i> .....	171
High-Throughput Screening to Advance In Vitro Toxicology: Accomplishments, Challenges, and Future Directions <i>Caitlin Lynch, Srilatha Sakamuru, Masato Ooka, Ruili Huang, Carleen Klumpp-Thomas, Paul Shinn, David Gerhold, Anna Rossosbek, Sam Michael, Warren Casey, Michael F. Santillo, Suzanne Fitzpatrick, Russell S. Thomas, Anton Simeonov, and Menghang Xia</i> .....	191
Hear and Now: Ongoing Clinical Trials to Prevent Drug-Induced Hearing Loss <i>John Lee, Katharine Fernandez, and Lisa L. Cunningham</i> .....	211
Direct K-Ras Inhibitors to Treat Cancers: Progress, New Insights, and Approaches to Treat Resistance <i>Ruth Nussinov and Hyunbum Jang</i> .....	231
Alcohol Use Disorder Treatment: Problems and Solutions <i>George F. Koob</i> .....	255
Targeting the Actions of Muscarinic Receptors on Dopamine Systems: New Strategies for Treating Neuropsychiatric Disorders <i>Eric J. Nunes, Nii A. Addy, P. Jeffrey Conn, and Daniel J. Foster</i> .....	277
From Thalidomide to Rational Molecular Glue Design for Targeted Protein Degradation <i>Vladas Oleinikovas, Pablo Gainza, Thomas Ryckmans, Bernhard Fasching, and Nicolas H. Thomä</i> .....	291
Apathy and Motivation: Biological Basis and Drug Treatment <i>Harry Costello, Masud Husain, and Jonathan P. Roiser</i> .....	313
Targeting Efferocytosis in Inflammaging <i>Ivan K.H. Poon and Kodi S. Ravichandran</i> .....	339
Orexin Receptor Antagonism: Normalizing Sleep Architecture in Old Age and Disease <i>Jarrah O.-Z.J. Kron, Ryan J. Keenan, Daniel Hoyer, and Laura H. Jacobson</i> .....	359
G Protein–Coupled Receptor Signaling: New Insights Define Cellular Nanodomains <i>Martin J. Lobse, Andreas Bock, and Manuela Zaccolo</i> .....	387
Molecular Activation of NLRP3 Inflammasome by Particles and Crystals: A Continuing Challenge of Immunology and Toxicology <i>Qiang Ma and Chol Seung Lim</i> .....	417
Translational In Vivo Assays in Behavioral Biology <i>Sarah L. Withey, Diego A. Pizzagalli, and Jack Bergman</i> .....	435



Mass Spectrometry–Based Proteogenomics: New Therapeutic Opportunities for Precision Medicine <i>Sunil K. Joshi, Paul Piehowski, Tao Liu, Sara J.C. Gosline, Jason E. McDermott, Brian J. Druker, Elie Traer, Jeffrey W. Tyner, Anupriya Agarwal, Cristina E. Tognon, and Karin D. Rodland</i> .....	455
Novel Immunopharmacological Drugs for the Treatment of Allergic Diseases <i>Ekaterini Tiligada, Daria Gafarov, Maria Zaimi, Joana Vitte, and Francesca Levi-Schaffer</i> .....	481
Deciphering Drug Targets and Actions with Single-Cell and Spatial Resolution <i>Zhengyuan Pang, Benjamin F. Cravatt, and Li Ye</i> .....	507
Artificial Intelligence for Drug Discovery: Are We There Yet? <i>Catrin Hasselgren and Tudor I. Oprea</i> .....	527
Oral Anticoagulants Beyond Warfarin <i>Renske H. Olie, Kristien Winckers, Bianca Rocca, and Hugo ten Cate</i> .....	551
Antiepileptic Drugs as Potential Dementia Prophylactics Following Traumatic Brain Injury <i>Laszlo F. Locskai, Hadeel Alyenbaawi, and W. Ted Allison</i> .....	577

## Indexes

Cumulative Index of Contributing Authors, Volumes 60–64 .....	599
Cumulative Index of Article Titles, Volumes 60–64 .....	604

## Errata

An online log of corrections to *Annual Review of Pharmacology and Toxicology* articles may be found at <http://www.annualreviews.org/errata/pharmtox>

## Related Articles

From the *Annual Review of Biochemistry*, Volume 92 (2023)

The Proteins of mRNA Modification: Writers, Readers, and Erasers

*Mathieu N. Flamand, Matthew Tegowski, and Kate D. Meyer*

The Activation Mechanism of the Insulin Receptor: A Structural Perspective

*Eunhee Choi and Xiao-Chen Bai*

The Inseparable Relationship Between Cholesterol and Hedgehog Signaling

*Christian Siebold and Rajat Robatgi*

Polyamines in Parkinson's Disease: Balancing Between Neurotoxicity  
and Neuroprotection

*Stephanie Vrijssen, Marine Houdou, Ana Cascalho, Jan Eggermont,  
and Peter Vangheluwe*

From the *Annual Review of Cancer Biology*, Volume 7 (2023)

Reinventing Radiobiology in the Light of FLASH Radiotherapy

*Charles L. Limoli and Marie-Catherine Vozenin*

Stalled CARs: Mechanisms of Resistance to CAR T Cell Therapies

*Diego Salas-Benito, Trisha R. Berger, and Marcela V. Maus*

New Tools for Lineage Tracing in Cancer In Vivo

*Matthew G. Jones, Dian Yang, and Jonathan S. Weissman*

Strategies for Heating Up Cold Tumors to Boost Immunotherapies

*Daniel J. Zabransky, Mark Yarchoan, and Elizabeth M. Jaffee*

The Potent and Paradoxical Biology of Cellular Senescence in Cancer

*Paul B. Romesser and Scott W. Lowe*

CAR NK Cells: The Future Is Now

*Martin J. Raftery, Alexander Sebastian Franzén, and Gabriele Pecher*

Rationales for Combining Therapies to Treat Cancer: Independent Action,  
Response Correlation, and Collateral Sensitivity Versus Synergy

*Emmett V. Schmidt, Linda Z. Sun, Adam C. Palmer, and Cong Chen*

The Blood-Brain Barrier: Implications for Experimental Cancer Therapeutics

*Joelle P. Straehla, David A. Reardon, Patrick Y. Wen, and Nathalie Y.R. Agar*

On the Biology and Therapeutic Modulation of Macrophages and Dendritic  
Cells in Cancer

*Matthew D. Park, Meriem Belabed, Steven T. Chen, Pauline Hamon, Samarth Hegde,  
Raphaël Mattiuz, Thomas U. Marron, and Miriam Merad*

Next-Generation Estrogen Receptor–Targeted Therapeutics  
*Tharu M. Fernando, Heather M. Moore, Matthew J. Wongchenko, and Ciara Metcalfe*

Targeting Driver Oncogenes and Other Public Neoantigens Using T Cell  
Receptor–Based Cellular Therapy  
*Tijana Martinov and Philip D. Greenberg*

From the *Annual Review of Genetics*, Volume 57 (2023)

Interplay Between Antimicrobial Resistance and Global Environmental Change  
*María Mercedes Zambrano*

From the *Annual Review of Immunology*, Volume 41 (2023)

Designing Cancer Immunotherapies That Engage T Cells and NK Cells  
*Oleksandr Kyrysyuk and Kai W. Wucherpfennig*

Structural Mechanisms of NLRP3 Inflammasome Assembly and Activation  
*Jianing Fu and Hao Wu*

T Cell Responses to SARS-CoV-2  
*Alessandro Sette, John Sidney, and Shane Crotty*

Systemic Lupus Erythematosus Pathogenesis: Interferon and Beyond  
*Simone Caielli, Zurong Wan, and Virginia Pascual*

Interfering with Interferons: A Critical Mechanism for Critical  
COVID-19 Pneumonia  
*Helen C. Su, Huie Jing, Yu Zhang, and Jean-Laurent Casanova*

From the *Annual Review of Medicine*, Volume 74 (2023)

COVID-19 Thrombotic Complications and Therapeutic Strategies  
*Alexander C. Fanaroff and Renato D. Lopes*

SARS-CoV-2 Vaccination-Induced Thrombotic Thrombocytopenia:  
A Rare but Serious Immunologic Complication  
*Charles S. Abrams and Geoffrey D. Barnes*

Cytomegalovirus Therapy: Role of Letermovir in Prophylaxis and Treatment  
in Transplant Recipients  
*Jennifer L. Saullo and Rachel A. Miller*

New Frontiers in Obesity Treatment: GLP-1 and Nascent Nutrient-Stimulated  
Hormone-Based Therapeutics  
*Ania M. Jastreboff and Robert F. Kushner*

Valvular Heart Disease: New Concepts in Pathophysiology  
and Therapeutic Approaches  
*Mackram F. Eleid, Vuyisile T. Nkomo, Sorin V. Pislaru, and Bernard J. Gersh*

Lessons Learned from the ISCHEMIA Trial for the Management of Patients  
with Stable Ischemic Heart Disease

*William E. Boden and Peter H. Stone*

Maternal Mortality in the United States: Trends and Opportunities for Prevention

*Siwen Wang, Kathryn M. Rexrode, Andrea A. Florio, Janet W. Rich-Edwards,  
and Jorge E. Chavarro*

Primary Aldosteronism and the Role of Mineralocorticoid Receptor Antagonists  
for the Heart and Kidneys

*Jordana B. Cohen, Irina Bancos, Jennifer M. Brown, Harini Sarathy, Adina F. Turcu,  
and Debbie L. Cohen*

Adeno-Associated Virus Gene Therapy for Hemophilia

*Benjamin J. Samelson-Jones and Lindsey A. George*

Multispecific CAR T Cells Deprive Lymphomas of Escape via Antigen Loss

*Fateeba Furqan and Nirav N. Shah*

Regulation of Erythropoiesis by the Hypoxia-Inducible Factor Pathway:  
Effects of Genetic and Pharmacologic Perturbations

*Gregg L. Semenza*

Systemic Lupus Erythematosus: New Diagnostic and Therapeutic Approaches

*Stephanie Lazar and J. Michelle Kablenberg*

SGLT2 Inhibitors: The Sweet Success for Kidneys

*Atit Dharra, Abid Khan, Vikas S. Sridhar, and David Z.I. Cherney*

Cystic Fibrosis Modulator Therapies

*Shijing Jia and Jennifer L. Taylor-Cousar*

Club Cell Secretory Protein in Lung Disease: Emerging Concepts  
and Potential Therapeutics

*Tereza Martinu, Jamie L. Todd, Andrew E. Gelman, Stefano Guerra,  
and Scott M. Palmer*

Diverse Approaches to Gene Therapy of Sickle Cell Disease

*Shanna L. White, Kevyn Hart, and Donald B. Kohn*

From the *Annual Review of Neuroscience*, Volume 46 (2023)

Therapeutic Potential of PTBP1 Inhibition, If Any, Is Not Attributed to  
Glia-to-Neuron Conversion

*Lei-Lei Wang and Chun-Li Zhang*

Therapeutic Potential of PTB Inhibition Through Converting Glial Cells to  
Neurons in the Brain

*Xiang-Dong Fu and William C. Mobley*

Deep Brain Stimulation for Obsessive-Compulsive Disorder and Depression

*Sameer A. Sheth and Helen S. Mayberg*

From the *Annual Review of Pathology: Mechanisms of Disease*, Volume 18 (2023)

Tumor Microenvironment in Pancreatic Cancer Pathogenesis  
and Therapeutic Resistance

*Mara H. Sherman and Gregory L. Beatty*

Tumor-Derived Extracellular Vesicles: Multifunctional Entities  
in the Tumor Microenvironment

*James W. Clancy and Crislyn D'Souza-Schorey*

New Insights into the Pathogenesis of Mastocytosis: Emerging Concepts in  
Diagnosis and Therapy

*Peter Valent, Cem Akin, Wolfgang R. Sperr, Hans-Peter Horny, Michel Arock,  
Dean D. Metcalfe, and Stephen J. Galli*

The Pathology of Asthma: What Is Obstructing Our View?

*Helena Aegerter and Bart N. Lambrecht*

From the *Annual Review of Physiology*, Volume 85 (2023)

Molecular Physiology of TRPV Channels: Controversies and Future Challenges

*Tamara Rosenbaum and León D. Islas*

Insulin Clearance in Health and Disease

*Sonia M. Najjar, Sonia Caprio, and Amalia Gastaldelli*

From the *Annual Review of Statistics and Its Application*, Volume 10 (2023)

Surrogate Endpoints in Clinical Trials

*Michael R. Elliott*

Statistical Data Privacy: A Song of Privacy and Utility

*Aleksandra Slavković and Jeremy Seeman*