Computers in Biology and Medicine 106 (2019) 54–64

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

In silico approaches and tools for the prediction of drug metabolism and fate: A review

Sayada Reemsha Kazmi[1], Ren Jun[1], Myeong-Sang Yu, Chanjin Jung, Dokyun Na*

School of Integrative Engineering, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul, 06974, Republic of Korea

ARTICLE INFO

Keywords:
In silico tools
Toxicity prediction
Metabolism prediction
drug discovery
drug metabolism

ABSTRACT

The fate of administered drugs is largely influenced by their metabolism. For example, endogenous enzyme–catalyzed conversion of drugs may result in therapeutic inactivation or activation or may transform the drugs into toxic chemical compounds. This highlights the importance of drug metabolism in drug discovery and development, and accounts for the wide variety of experimental technologies that provide insights into the fate of drugs. In view of the high cost of traditional drug development, a number of computational approaches have been developed for predicting the metabolic fate of drug candidates, allowing for screening of large numbers of chemical compounds and then identifying a small number of promising candidates. In this review, we introduce *in silico* approaches and tools that have been developed to predict drug metabolism and fate, and assess their potential to facilitate the virtual discovery of promising drug candidates. We also provide a brief description of various recent models for predicting different aspects of enzyme-drug reactions and provide a list of recent *in silico* tools used for drug metabolism prediction.

1. Introduction

As our understanding of drug fate–determining metabolic reactions has increased, drug metabolism has attracted greater attention as a critical factor in drug discovery [1–3]. The fate of substances such as drugs and xenobiotics administered into our body is largely governed by the three phases of drug metabolism: phase I, introduction of a reactive group by oxidation, reduction or hydrolysis, among others; phase II, conjugation with various moieties; and phase III, removal of xenobiotics and metabolites from cells in the liver and intestine. These transformation processes may convert compounds to inactive, active, or toxic metabolites. Not surprisingly, because it is responsible for the clearance of ~70% of clinical drugs, metabolism has been intensively investigated as part of drug development efforts [4].

Natural compounds have recently attracted considerable research attention owing to their inherent advantages and high potential as drug candidates [5]. Moreover, the structural similarity of certain natural compounds with metabolites found in the human body makes metabolism a critical factor in determining the effectiveness of natural drugs [6–8]. For instance, historical opioid drug candidates are metabolized into more potent metabolites, such as (dihydro) codeine, which, in turn, is metabolized into (dihydro) morphine [9,10]. Considering the large number of endogenous enzymatic reactions that influence drug modification through (de)activation and (de)toxification, determining how a drug is metabolized is an important step in drug discovery.

Numerous experimental technologies have been used in recent decades to study the metabolism and fate of drugs [11–13]. The traditional drug discovery method—target-to-hit, hit-to-lead, and lead optimization—is expensive, costing more than $200 million for the average drug, and time-consuming, with a typical discovery period of 4–5 years [14,15]. In addition, because it is impossible to exactly replicate biological *in vivo* environments, such methods are relatively inaccurate and are still considered low-throughput, given the scale of combinatorial structural variations of chemical compounds.

Numerous advances in predicting drug metabolism using *in silico* approaches have been made as part of drug discovery efforts, and different aspects of these advances have been reviewed [16–22]. These include tools to predict drug metabolism based on the interactions of drugs with cytochrome P450 (CYP450) enzymes and their metabolic endpoints [18,22], tools to predict ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties of drugs and related solubility, permeability and bioavailability issues [16], as well as approaches for predicting the inducibility of drug-metabolizing enzymes and transporters that affect the plasma concentration of drugs, which can cause undesirable or prolonged action or adverse effects [21].

Given these observations, *in silico* approaches have been

* Corresponding author.
E-mail address: blisszen@cau.ac.kr (D. Na).
[1] These authors contributed equally.

https://doi.org/10.1016/j.compbiomed.2019.01.008
Received 19 September 2018; Received in revised form 14 January 2019; Accepted 14 January 2019

increasingly used to predict the metabolic conversion of drugs [23] and as such are considered the best "fail early and fail cheap strategy," allowing for reduced costs, time savings, and thus decreased attrition rates in late drug discovery phases. Herein, we present an introduction to fundamental approaches to *in silico* model development for predicting the metabolic fate of drugs as well as their toxicity. In addition, we summarize currently available *in silico* tools and some recent achievements in predicting drug metabolism. Finally, we discuss the future and challenges of *in silico* drug discovery.

2. *In silico* approaches

2.1. Predictions based on quantitative structure-activity relationships and machine-learning approaches

The quantitative structure-activity relationship (QSAR) concept, developed in the early 1960s by Hansch/Fujita [24] and Free/Wilson [25] and widely used in drug discovery, assumes that molecules with similar structures potentially exhibit similar chemical and biological activities [26–28]. The original concept of structure-activity relationship dates back to 1868, when Crum-Brown and Fraser introduced the idea of correlating the chemical composition of a certain compound with its physiological properties in biological systems [29]. QSAR-based models are commonly used in the lead-optimization step of drug discovery to assess various drug properties (particularly toxicity) and consequently reduce the number of promising lead compounds identified by screening, thereby ultimately minimizing time, cost, and labor. The European Commission Regulation (REACH: Registration, Evaluation and Authorization of Chemicals) [30] accepts the use of various approaches such as QSAR once the results are proven to be highly reliable [31].

The QSAR approach employs experimental datasets comprising the biological activity of chemical compounds; their chemical and physical features, represented as molecular descriptors [32]; and statistical methods for correlating these molecular descriptors with biological activity [33] (Fig. 1). Molecular descriptors are arithmetical values representing the physicochemical properties of compounds, and can be categorized into 1D, 2D or 3D descriptors, depending on the provided amount/type of information. The most common types of descriptors used in QSARs are constitutional, electronic, topological and geometrical descriptors, and include molecular weight, total number of atoms, total number of carbon atoms, atomic net, total number of bonds and Van der Waals area, among others. A broad range of software and web-based tools are available for computing molecular descriptors, as shown in Table 1; there are also various QSAR systems with their own integrated descriptor-generators, including CASE Ultra (http://www.multicase.com/case-ultra) and Leadscope (http://www.leadscope.com/).

Generally, QSARs that predict the metabolic conversion of endogenous or exogenous compounds are constructed for hepatic CYP450 family enzymes (which metabolize the majority of drugs into toxic chemical compounds [34–36]) and are known for their robustness in predicting toxicity; as such, they provide valuable information for large-scale, virtual screening of drug efficacy.

Table 2 lists common QSAR-based models built for the prediction of drug metabolism reactions. Several other models listed in the table (e.g., IDsite, SMARTcyp) can also predict the site at which a metabolic transformation occurs in a chemical compound. In addition, the C-QSAR database, constructed in 2003 and available to users [37], comprises over 18000 QSAR equations and associated biophysical data. The QSAR Databank, another repository that archives *in silico* QSAR-type descriptive and predictive models, enables the research community to share and present their QSAR data [38]. QSARs have been used since the early era of drug discovery, but their application was limited to only small linear datasets. However, advanced methods based on direct scoring and/or machine-learning algorithms that are capable of modeling complex nonlinear datasets have more recently been employed [39].

The rapid progress in developing new machine learning methods in computer science has inspired the development of thousands of QSAR models for accurate drug metabolism prediction that are based on methods other than linear and multiple linear regression [30,40–43].

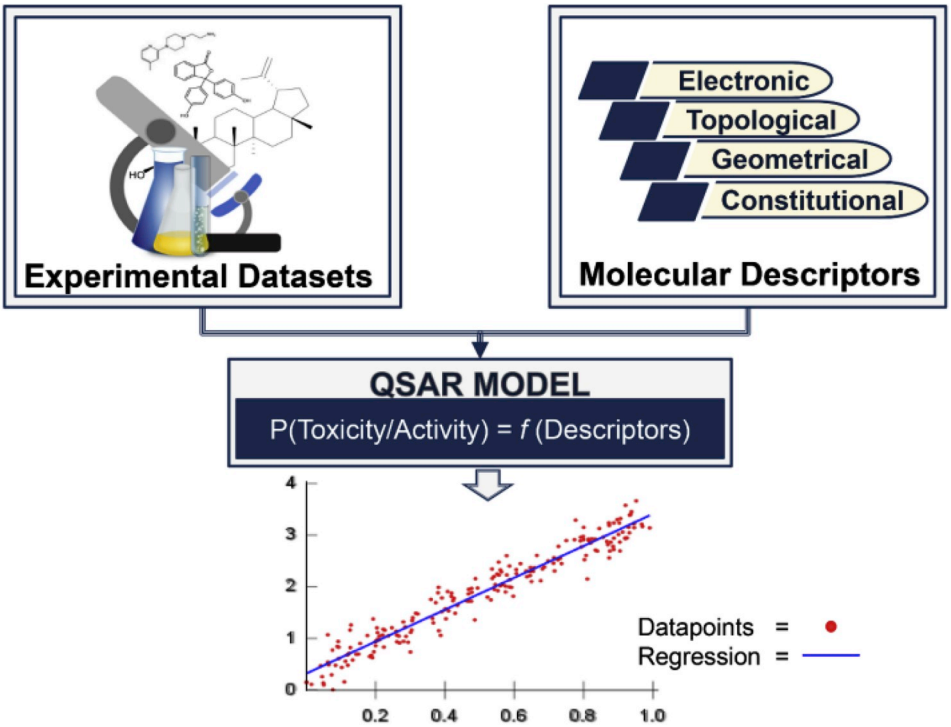

Fig. 1. Molecular descriptors, calculated from curated experimental datasets, and measured descriptors are used in QSAR models. For example, the term "P(toxicity/activity)" represents a predictor of the biological activity of known compounds as a function of their structural descriptors.

Table 1
A non-exhaustive list of tools available for computing molecular descriptors.

Software	Operating System	No. of available descriptors	License	Description
ADAPT[a]	Unix/Linux	265	Free	– Software for development of QSAR/QSPR. – Supports feature selection, i.e., it selects the best subset of descriptors.
ADMET Predictor[b] (PCB module)	Windows	297	Commercial	– A module of QSAR/QSPR development software. – Artificial Neural Network Ensemble (ANNE)-based calculation.
ChemAxon[c] (Calculator plugins)	Windows/Unix/Linux	> 500	Commercial	– Descriptor-calculation module of ChemAxon's cheminformatics development platform. – An extension for KNIME (Konstanz Information Miner) is also available.
Codessa[d]	Windows	> 1400	Commercial	– Multilinear/Nonlinear QSAR/QSPR model construction tool which supports descriptor calculation.
Corina Symphony[e]	Windows	786	Commercial	– Cheminformatics application which supports *in silico* experiment optimization. – Chemotype descriptors (structural fragments with embedded physicochemical properties) are included.
DRAGON 7.0[f]	Windows/Unix/Linux	5270	Commercial	– Parameters can be customized for the generation of molecular fragments and fingerprints. – Supports molecular descriptor calculation of disconnected structures such as salts or ionic liquids. – An extension for KNIME is also available.
E-Dragon[g]	Web service	1666	Free	– Web-service java applet based on a previous version of DRAGON.
MOE[h]	Windows/Unix/Linux	> 400	Commercial	– Suite of medicinal chemistry tools providing an integrative drug discovery environment. – Custom descriptors using built-in language can be added.
Molconn-Z[i]	Windows/Unix/Linux	> 1000	Commercial	– Descriptor calculation tool for generation of chemical information representing molecular connectivity.
MOLGEN QSPR[j]	Windows	708	Commercial	– Tool for optimization of combinatorial chemistry experiments that support descriptor calculation.
PaDEL-descriptor (CDK)[k]	Java JRE	1875	Free	– Software to calculate descriptors and fingerprints based on the Chemical Development Kit. – Max 16092 bits of fingerprints can also be generated. – An extension of KNIME is also available.
PowerMV[l]	Windows	122	Partially Free[m]	– Software environment supports molecular viewing and statistical analysis. – Over 10,000 “atom pair” or “fragment pair” fingerprints can be calculated.
PreADMET[n]	Windows	955	Commercial	– Information about chemical compounds involved in ADMET can be predicted. – Drug-likeness/ADME/toxicity prediction can be conducted for free using the web application.
Open Babel[o]	Windows/Linux	> 20	Free	– Supports conversion of chemical data files, molecular searching, chiral detection and superimposing molecules – Also supports descriptor calculation.
QikProp[p]	Windows/Linux	> 20	Commercial	– Facilitates prediction of relevant pharmaceutical properties of drug candidates – Computes over 20 physical descriptors to improve prediction accuracy and generate QSAR models.
ACD Labs/Percepta[q]	Web service & modules	> 40	Free	– Predicts the ADMET and physicochemical properties of the molecules.
MOPAC[r]	Windows/Linux	24	Free	– A semi-empirical molecular orbital package used to study molecular structures and reactions.
EPI Suite[s]	Windows	13	Free	– Consists of various programs used to estimate the physical, chemical, aquatic toxicity and environmental fate of compounds.

[a] http://research.chem.psu.edu/pcjgroup/adapt.html.
[b] http://www.simulations-plus.com/software/admet-property-prediction-qsar/.
[c] https://www.chemaxon.com/products/calculator-plugins/.
[d] http://www.codessa-pro.com/.
[e] https://www.mn-am.com/products/corinasymphony.
[f] https://chm.kode-solutions.net/products_dragon.php.
[g] http://www.vcclab.org/lab/edragon/.
[h] http://www.chemcomp.com/MOE-Cheminformatics_and_QSAR.htm.
[i] http://www.edusoft-lc.com/molconn/.
[j] http://molgen.de/download.html.
[k] http://www.yapcwsoft.com/dd/padeldescriptor/.
[l] https://www.niss.org/research/software/powermv.
[m] Commercial affiliates available.
[n] https://preadmet.bmdrc.kr/.
[o] http://openbabel.org.
[p] https://www.schrodinger.com/qikprop.
[q] https://www.acdlabs.com/products/percepta/.
[r] http://openmopac.net/.
[s] https://www.epa.gov/tsca-screening-tools/download-epi-suitetm-estimation-program-interface-v411.

Table 2
A non-exhaustive list of tools available for QSAR-based enzyme reaction prediction.

Software	Operating system	Target	License	Description
ADME WORKS Predictor[a]	Windows/Linux	CYP450 isoforms[b]	Commercial	– Provides high-speed virtual screening system for ADME/T property prediction with quantitative/qualitative SAR. – Models can be customized with optional tools. – Kinetics of some CYP450 isoforms and inhibitor of CYP450 3A4 can be predicted.
ADMET Predictor[c] (Metabolism module)	Windows	CYP450 isoforms[d]	Commercial	– Offers advanced data mining, clustering, and molecular pair analysis. – Inhibition, kinetics, regioselectivity of CYP450 isoforms can be predicted based on ANNE.
admetSAR[e]	Web service	CYP450 isoforms[f] and P-glycoprotein	Free	– Database/tool for prediction of ADME/T properties of drug candidates. – QSAR-based classification models can predict inhibitor/substrate of CYP450 isoforms or P-glycoprotein.
PreADMET[g]	Web service Windows (PC version)	CYP450 isoforms[h] and P-glycoprotein	Free (Web service) Commercial (PC version)	– Provides QSAR-based ADME/T prediction on the web service. Commercial PC software is also available. – Inhibitors or substrates of some CYP450 isoforms and inhibitors of P-glycoprotein can be predicted.
SMARTCyp[i]	Web service	CYP450 isoforms[j]	Free	– Predicts molecule metabolic sites that are most prone to CYP450 mediated metabolism. – Predicts the metabolism site directly from the molecule's 2D structure without generating 3D structures or electronic descriptor calculations.
SOMP (Way2Drug)[k]	Web service	CYP450 isoforms[l] and UDP-glucuronosyltransferase	Free	– Predicts site of metabolism based on structural formula of chemicals. – Prediction is based on PASS and LMNA descriptors.
MetaSite[m]	Windows/Linux	CYP450, FMO3, and AOX1	Commercial	– Predicts metabolic conversion reactions in phase I metabolism by CYP450 and FMO3. – Improved prediction of novel pharmaceuticals as a result of its unique algorithm that is independent of the training dataset.
RS-WebPredictor[n]	Web service	CYP450 isoforms[o]	Free	– Predicts CYP-mediated metabolism on user-submitted molecules. – Predicts region-selectivity of CYP450 isoforms.
Meteor Nexus[p]	Windows	User query structure	Commercial	– Consists of an expert knowledge base of biotransformation to predict the metabolic fate of the drugs. – Provides information for decision-making when there is little or no experimental metabolism data available.
ACD Labs/Percepta[q]	Windows/Linux	User 2D structure/SMILE	Commercial	– Predicts the ADMET and physicochemical properties of molecules. – Predicts CYP inhibitors and substrates.
MetabolExpert[r]	Windows/Linux	User 2D structure/SMILE	Commercial	– Predicts the metabolic fate of compounds – User can add their own rules (open-knowledge base)
Meta-PC[s]	Unix/Linux	Query chemical structure	Commercial	– Expert rule based tool which consists of four transformation dictionaries. – Predicts the biotransformation (metabolic and degradation products) of chemical compounds.
syGMa[t]	Windows/Linux	Query chemical structure	Free (for academic institutions)	– A Python library used to predict the potential metabolites of a query structure. – Consists of wide ranging reaction rules including phase 1 and 2 metabolism.
TIMES[u]	Unix/Linux	User query structure	Commercial	– Predicts the metabolic biotransformation in a specific metabolic environment by generating multiple metabolic pathway maps and using different data sources to train the metabolic simulator.
MetaPath (OASIS)[v]	Unix/Linux	User query structure	Commercial	– Collects and compiles metabolism data in a systematic database. – Comprised of tools to access and extract information such as biotransformation and metabolites responsible for a given transformation.
IDSite[w]	–	CYP isoform 2D6	–	– Predicts the sites of metabolism and 3D structure of protein-ligand complexes. – Uses the docking program Glide (Table 3) and PLOP as a protein refinement module to model induced-fit complexes and predict site of metabolism. – There is no web service available, but algorithm's details could be found in paper DOI: 10.1021/ct200462q.
Metabolizer (ChemAxon)[x]	Windows/Linux	User query structure	Commercial	– A metabolic pathway prediction tool that predicts xenobiotics and other major metabolites and estimates their metabolic stability. – The product has been discontinued as of July 2016 but can be downloaded in its prior JChem metabolizer version.

(*continued on next page*)

Table 2 (*continued*)

Software	Operating system	Target	License	Description
MetaDrug[y]	Unix	Small molecule compounds	Commercial	– Predicts biological effects of small molecule compounds. – Comprised of curated information on over 6000 human compounds that consists of metabolic fate, ADME properties, and therapeutic and side effects.

[a] http://www.fqs.pl/en/chemistry/products/admeworks-predictor.
[b] 2D and 3A4.
[c] http://www.simulations-plus.com/software/admet-property-prediction-qsar/metabolism/.
[d] 1A2, 2A6, 2B6, 2C8, 2C19, 2C9, 2D6, 2E1, and 3A4.
[e] http://lmmd.ecust.edu.cn:8000/.
[f] 1A1, 1A2, 2A5, 2C9, etc.
[g] https://preadmet.bmdrc.kr/.
[h] 2C9, 2C19, 2D6, and 3A4.
[i] https://smartcyp.sund.ku.dk/.
[j] 3A4 isoform.
[k] http://www.way2drug.com/SOMP/.
[l] 1A2, 2C9, 2C19, 2D6 and 3A4.
[m] http://www.moldiscovery.com/software/metasite/.
[n] http://reccr.chem.rpi.edu/Software/RS-WebPredictor/.
[o] 2C9, 2D6, 3A4, 1A2, 2A6, 2B6, 2C8, 2C19 and 2E1.
[p] https://www.lhasalimited.org/products/meteor-nexus.htm.
[q] https://www.acdlabs.com/products/percepta/.
[r] http://www.compudrug.com/metabolexpert.
[s] http://www.multicase.com/meta-pc.
[t] https://sygma.readthedocs.io.
[u] http://oasis-lmc.org/products/software/times.aspx.
[v] http://oasis-lmc.org/products/software/metapath.aspx.
[w] https://pubs.acs.org/doi/abs/10.1021/ct200462q.
[x] https://docs.chemaxon.com/display/docs/Metabolizer.
[y] https://portal.genego.com/.

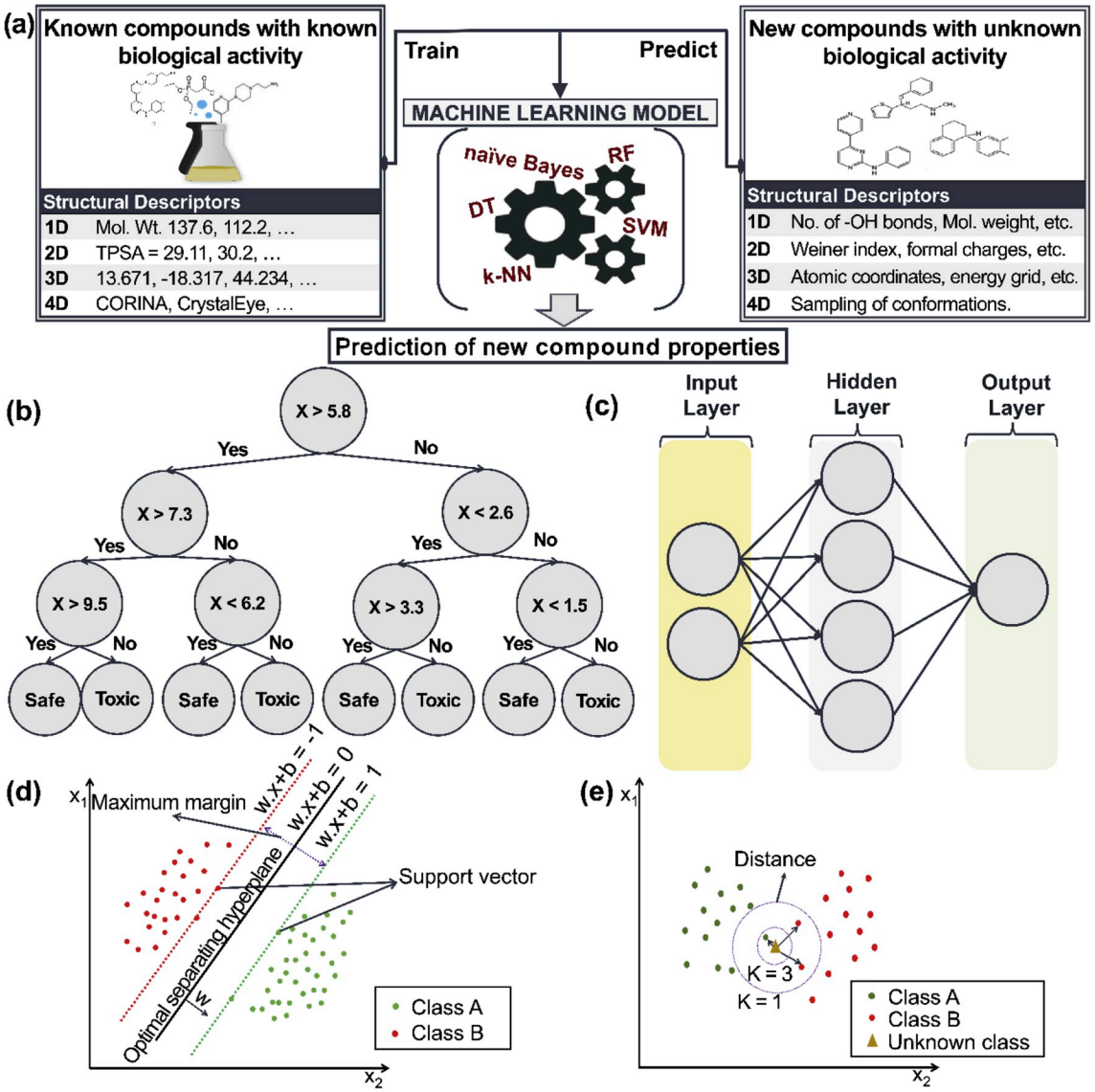

Fig. 2. Schematic illustration of machine learning algorithms commonly used to predict the biological activities of unknown compounds. (a) Conventional development of a machine-learning model to predict the fate of unknown compounds; (b) decision tree; (c) simple neural network; (d) support vector machine; and (e) k-nearest neighbor.

Machine learning, defined as a computational method that learns from a set of test data to build a model for classifying unknown data [46], has been primarily used to develop QSAR models [44,45]. The application of machine learning approaches in modern drug discovery has accelerated the process of scanning and filtering out ineffective compounds, achieving a significant time and cost reduction compared with experimental screening methods [47]. Machine learning is better suited for extracting non-parametric and nonlinear relationships from datasets, which enables the development of *in silico* models with better predictive performance [45,48]. Several machine learning methods (e.g., neural network, decision tree, support vector machine, k-nearest neighbor) have been successfully utilized to build more accurate QSAR models [49–51] that receive a set of descriptors from a large dataset as input and create a classification model that predicts the biological activity of a query compound as output (see Fig. 2).

Currently, machine learning is widely used in the field of computer-aided drug discovery, which allows for predicting the interaction between a ligand and a target protein, and hence facilitates the development of new drugs [52–54]. It also seeks to predict the ADMET properties of drugs, thus ultimately facilitating the development of safe and promising agents [55–57]. Drug metabolism is broken down into several phases, each with numerous enzymes that play a role in metabolizing the drug; therefore, a large number of machine learning models have been built to classify a drug's fate based on whether or not the drug will be metabolized by certain enzymes [58].

Recently, increased attention has been directed toward drug toxicity prediction using other machine learning methods, such as neural networks and deep learning [59,60], that involve the use of strong multilayered interconnected neuronal networks comprising processing units, denoted as nodes [61,62]. Examples of architectures used for predicting biological activity include convolutional, autoencoder, and recurrent neural networks [63,64]. The rapid increase in the volume of pharmaceutical data and computational power has inspired the application of neural networks and deep learning in a variety of fields, including bioinformatics [65], cheminformatics [66], structure prediction [67–69], and drug discovery [70–73].

The advent of computational prediction represents a major turning point in the history of drug discovery, with a number of machine learning–based models now being available for drug toxicity prediction [74,75]. However, their application is still limited by drawbacks such as

a tendency to over-fit data and difficulties in selecting the appropriate algorithm and descriptors for a problem from among those available [76]. An over-fitted model arises when the model is too complex or the number of features/descriptors is too large relative to the size of the dataset. These issues lead to a biased model that works well on the training dataset used to build the model, but is unable to accurately predict using external datasets [77].

2.2. Structure-based computational approaches

The identification of a protein's structural properties provides insight into its biological activity and enables the design of effective ligands for its binding site. Often, metabolic reactions occur at the site where the ligand binds to the target protein. This site tells us a lot about the metabolic fate of the drug, and thus whether the drug will be therapeutically active, inactive, or toxic. It also often provides information that aids in optimizing lead compounds in the drug-development process. To date, structure-based approaches have been one of the most successful and best-established methodologies used in various pharmaceutical research areas for drug design [78–81]. These state-of-the-art methods involve techniques such as computational docking and molecular dynamics, which have been intensively used to study drug metabolism by identifying the site of metabolism [82] and molecular interactions, information that contributes significantly to the drug discovery process [83,84].

The docking method considers the interaction between a small molecule and an active site on a target protein, and predicts the affinity of their binding interactions from their docking orientation and the interacting forces between them. Protein-ligand interactions are simulated using powerful computational tools, such as docking algorithms [85] implemented in AutoDock Vina [86], GOLD [87], and DOCK [88] software packages, that predict the most favorable reaction. The construction of these models is predicated on the assumption that structural information is closely related to the metabolic fate of the drug [89]. The docking approach has been widely used to quickly identify promising lead compounds from large compound libraries. Ligand-protein interactions often require structural changes to achieve better interactions, and can be simulated with the aid of molecular dynamics simulation. Thus, molecular dynamics simulations are often used in combination with docking algorithms to further refine docking complexes to include other parameters, such as solvent effects, thereby yielding more accurate drug candidates [85]; they are also used in predicting the site of metabolism [90]. Table 3 lists some of the common tools used for protein-ligand docking simulation.

Despite their numerous advantages, structure-based approaches require high computational power to model structural flexibility [92]. The processes of calculating binding energy and evaluating docking conformation require various methods that can take from seconds up to several days, making these calculations computationally expensive. In addition, a target protein and its ligand may undergo structural changes to adapt their structures to a suitable conformational state [93]; thus, generating an exact modeling replica is still a daunting challenge. However, additional methods such as the use of rotamer libraries [94] or soft docking modeling [95] have been used to improve the accuracy of simulations. Rotamer libraries are used to predict the most suitable side chain conformations and remove unfavorable conformations, resulting in the selection of low-energy side chain conformations and thereby increasing modeling accuracy and reducing modeling time. Soft docking can be performed using soft scoring functions to bring about minor changes in the conformation of protein receptors, an approach that is known to be computationally efficient [96].

Docking approaches have been meticulously applied for *in silico* prediction of drug toxicity, allowing identification of lead compound binding to unfavorable proteins and prediction of undesirable side effects and consequences [97–99]. For instance, using docking simulation, Ji et al. searched for potential protein binding partners of 11 marketed anti-HIV drugs among 147 known adverse drug reaction-related proteins deposited in the DART (Drug Adverse Reaction Target) database to predict adverse effects of the drugs. They confirmed that the predicted adverse drug reaction-related proteins that caused side/toxic effects were consistent with reported adverse reactions resulting from drug-target interactions [100]. In 2011, the same approach was adopted to predict the toxicities of melamine and its major derivative, cyanuric acid. This analysis identified potential toxicity-related target proteins and provided detailed insights into the toxicity mechanism. Specifically, in addition to nephrotoxicity, melamine was also predicted to exert lung toxicity [101]. Hence, the computational docking strategy can substantially facilitate drug toxicity prediction.

Despite the significant progress in drug discovery achieved using structure-based approaches, the widespread use of this strategy is hindered by numerous limitations, not least of which is the challenge of pro-drugs and their metabolic conversion to other active compound(s) [102]. For instance, high-resolution experimental structural data for all target proteins (e.g., enzymes) are required to reliably predict the metabolic fate of a given drug [103]. Additionally, analyses of protein-ligand complexes (and hence, accurate reaction prediction) are hindered by the structural flexibility of proteins [96].

3. Applications of developed *in silico* tools in predicting drug metabolic fates

3.1. Prediction of drug conversion to toxic metabolites

The metabolism of xenobiotics such as drugs and other foreign substances involves certain important enzymatic reactions, such as those mediated by CYP450 family enzymes expressed in liver and small intestines. According to literature reports, ~90% of drugs can be efficiently metabolized by six CYP450 enzymes [104], the activity of which can be altered by factors such as genetic polymorphisms, cytokine regulation, disease state, sex, age, and hormones [105–107]. Another example is provided by the membrane-bound P-glycoprotein (encoded by the multi-drug resistant-1 gene), which is expressed in various tissues including the intestinal epithelium, liver cells, and cells constituting the blood-brain barrier. These tissues are known to act as biological barriers that restrict the uptake of various substances into cells; hence, they affect the distribution of drugs for further metabolism [108–110].

As listed in Tables 2 and 3, a large number of computational models for predicting enzymatic reactions are available, reflecting the strong influence of such reactions on ADMET properties that result in a decrease or increase in the pharmaceutical effect of the drug [111].

3.2. Prediction of enzymatic reactions of drugs and enzymes

Endogenous enzymes in the human body may mediate the metabolic conversion of administered drugs into inactive, active or toxic chemical compounds, highlighting the practical significance of predicting potential chemical modifications of drugs. Drug metabolism involves enzyme-catalyzed reactions; thus, a number of attempts have been recently made to predict enzyme-mediated reactions. An example of this is the reported prediction of hydrolysis and redox reactions. In this study, a machine-learning–based model was built to predict the classes/subclasses of hydrolysis reactions (EC 3.b.c.d, b is 1, 2, and 5) and redox reactions (EC 1.b.c.d, b is 1, 2, 3, 4, 5, 8, 13, and 14) [112], allowing metabolic transformations of a molecule to be predicted.

To predict enzymatic reactions involved in metabolic pathways, one study used the novel approach of building a substrate-enzyme-product interaction network based on the k-nearest neighbor method to provide toxicity-related information in metabolic pathways. The substrate, enzyme and products were encoded by molecular descriptors and physicochemical properties, and the k-nearest neighbor algorithm was adopted to construct the predictive model. Substrate-enzyme-product

Table 3
A non-exhaustive list of protein-ligand docking simulation tools.

Software	Operating System	License	Description
AutoDock Vina[a]	Windows/Linux/Unix	Free Open source	– Improved version of popular ligand-protein docking software, AutoDock 4, using brand new calculation algorithms. – Almost no input variable limitations, such as the size of search space or maximum number of atoms in a ligand. – Qualified projects can run the software on the World Community Grid[b] for free.
BetaDock[c]	Linux	Free	– Shape-priority docking simulation software based on the beta-complex theory. – Requires relatively little human intervention.
BSP-SLIM[d]	Web service	Free	– Blind molecular docking method for low-resolution protein structures. – Protein structure is predicted from the sequence and docking conformation.
DOCK 6.8[e]	Windows/Linux/Unix	Free Open source	– Flexible protein-ligand docking simulation based on a geometric matching algorithm. – Can predict binding modes of small molecule-protein complexes and examine protein-protein or protein-DNA complexes.
Docking Server[f]	Web service	Partially Free[g]	– Introduces quantum mechanical semi-empirical calculations to predict partial charge of proteins. – Supports high-throughput docking of ligand library. – Docking procedure is an integration of various computational chemistry packages.
FlexAID[h]	Windows/Linux/Unix	Free Open source	– Accounts for ligand side-chain flexibility and utilizes a soft scoring function. – Using an iterative Monte Carlo method, pairwise energy data is extracted from the PDBbind database[i].
Glide[j]	Windows/Linux/Unix	Commercial	– A component of Schrödinger suites, which provide an integrative environment for various scientific fields. – Exhaustive search-based docking simulation that supports calculation speed control affecting accuracy. – An extension for KNIME is also available.
GOLD Suite[k]	Windows/Linux	Commercial	– Simulates docking between flexible proteins and ligands. – Highly configurable and customizable docking protocol is available. – Efficient and accurate virtual screening can be conducted on the compound library.
idTarget[l]	Web service	Free	– Predicts binding between biomolecular targets and small chemicals using a divide-and-conquer docking approach. – Introduces a scoring function based on robust regression and quantum chemical charge models.
MOE[m]	Windows/Unix/Linux	Commercial	– Suite of medicinal chemistry tools which provides an integrative drug discovery environment. – User can choose one of various scoring functions available. – Supports automated structure preparation and correction.
MOLS 2.0[n]	Java	Free Open source	– Deals with “induced-fit” side chain receptor flexibility for docking peptide ligands. – Flexible ligand–flexible receptor docking prediction.
ParDock[o]	Web service	Free	– Simulation of rigid protein–ligand docking. – Uses Monte Carlo docking protocol based on all-atom energy calculation.
rDock[p]	Linux	Free Open source	– Developed for high-throughput virtual screening applications, predicts docking between small molecules and proteins. – User can incorporate additional constraints or information for biased guided docking.
SwissDock[q]	Web service	Free	– Blind docking simulation based on EADock dihedral space sampling. – Force field is calculated and evaluated using CHARMM [91].
Virtual ToxLab[r]	Windows/Unix/Linux	Free (for academic institutions)	– Toxic potential prediction of small molecules. – Using flexible docking and multi-dimensional QSAR, this tool can predict ligands of 16 proteins and CYP450 isoforms.

[a] http://vina.scripps.edu/.
[b] https://www.worldcommunitygrid.org/.
[c] http://voronoi.hanyang.ac.kr/software.htm.
[d] http://zhanglab.ccmb.med.umich.edu/BSP-SLIM/.
[e] http://dock.compbio.ucsf.edu/DOCK_6/index.htm.
[f] http://www.dockingserver.com/web.
[g] Commercial premium licenses.
[h] http://biophys.umontreal.ca/nrg/NRG/FlexAID.html.
[i] http://sw16.im.med.umich.edu/databases/pdbbind/index.jsp.
[j] https://www.schrodinger.com/glide.
[k] https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/.
[l] http://idtarget.rcas.sinica.edu.tw/.
[m] https://www.chemcomp.com/MOE-Structure_Based_Design.htm.
[n] https://sourceforge.net/projects/mols2-0/.
[o] http://www.scfbio-iitd.res.in/dock/pardock.jsp.
[p] http://rdock.sourceforge.net/.
[q] http://www.swissdock.ch/.
[r] http://www.biograf.ch/index.php?id=projects&subid=virtualtoxlab.

interaction networks were represented as main factors, and optimal features were selected using the maximum relevance minimum redundancy and incremental feature selection (mRMR-IFS) method. Out of 290 features, 160 were retained and grouped into 10 different categories, including amino acid composition, predicted secondary structure, hydrophobicity and polarity, among others [113].

Another study recently used a machine-learning–based approach to computationally predict the potential reactions of 1449 enzymes (including CYP450 enzymes) deposited in the databases, BRENDA (Braunschweig Enzyme Database) [114] and HMDB (Human Metabolome Database) [115]. In particular, it was assumed that if a known molecule interacts with a certain enzyme, the query molecule should also interact with this enzyme if the physicochemical descriptors of the query molecule are similar to those of the known molecule. Interestingly, this model was shown to predict enzymatic conversion by CYP450 enzymes and concomitant formation of toxic metabolites, and was therefore concluded to be useful for predicting drug metabolism in terms of drug biological activity and toxicity [116].

Thus, the above methods for predicting potential enzymatic reactions have revolutionized *in silico* approaches and have substantially contributed to drug screening and the identification of potential new drug leads.

3.3. Prediction of drug-target interactions based on the concept of pharmacological space

Based on the theory that proteins mediating similar reactions are likely to exhibit substrate similarity [117], Yamanishi et al. [118] proposed a QSAR-based model for the prediction of unknown drug-target interactions, introducing the concept of pharmacological space, which integrates chemical structure and protein genomic profile information. With the assumption that compounds that are highly similar structurally are very likely to interact with similar target proteins, chemical and genomic similarity was calculated and unified as pharmacological space. The prediction model was constructed using three datasets: a drug-target interaction dataset extracted from various databases, a chemical dataset composed of chemical compound structures expressed as a similarity matrix between two compounds (chemical space), and a genomic dataset consisting of amino acid sequences of target proteins expressed as a similarity matrix (genomic space). Chemical and genomic protein sequence datasets were collectively combined in pharmacological space and calculated based on the bipartite graph learning model method. Model performance was subsequently evaluated against the drug-target interaction dataset, and the developed model was shown to predict both enzyme-compound interaction activity and the interaction of proteins with other factors, such as ion channels, G-protein–coupled receptors, and nuclear receptors. Therefore, the constructed model allowed reliable prediction of the interaction of a set of protein-compound pairs.

4. Problems associated with predictive model construction

The inconsistency of available experimental data used to build *in silico* models is a major concern [119]. Predictive models strongly rely on experimental data for model construction; thus, high variability in experimental assays produced by biological variation and technical errors can lead to erroneous data and may therefore introduce inaccuracy into predictive models. The inaccuracy of *in silico* models may also result from different experimental conditions for the multiple resources collected, imbalanced datasets, and molecular descriptor values that differ from tool to tool [120]. The reliability of the experimental data is validated if the results are consistent and accurate under a standardized experimental protocol over time [121]. Therefore, in addition to considering the validity of *in silico* models, the quality of the experimental data should also be considered. There have been several attempts to take into account the reliability of experimental data and its degree of uncertainty, efforts that often increase prediction accuracy [122]. This underscores the fact that poor prediction accuracy may not result merely from the *in silico* nature of the predictive tool, but may also reflect the nature of biological experiments. Comprehensive databases such as Drugbank [123], HMDB [115] and others, such as MetaDrug and MetaCore [104], are gradually becoming more reliable through human curation, advanced data mining algorithms, and/or addition of new experimentally proven data, increasing the reliability of datasets used by *in silico* models and thereby improving the accuracy of results. *In silico* methods have been a major innovation in the effort to predict drug fate, yet the construction of reliable predictive models remains a challenge. Therefore, predictive models are tested and their accuracy, precision and robustness confirmed using external validation datasets to judge whether the model is acceptable for a particular purpose. For instance, three web servers, SOMP, SMARTcyp and RS-WebPredictor, used to predict the site of metabolism have been compared for predictive accuracy. Of these, the SOMP server was shown to have higher invariant accuracy of prediction (similar to AUC) than others, with a score of 0.9, and is thus considered an adequate drug metabolism prediction tool [124].

5. Conclusions

Because of its central importance, metabolism in biological systems has been intensively researched, especially in the field of drug discovery. The high impact of drug metabolism on drug efficacy and drug fate in biological systems has given rise to numerous *in silico* approaches and tools for metabolic reaction prediction in recent decades. However, the limitations of these approaches cannot be ignored. Specifically, the fact that these methods are highly dependent on experimental data is a major concern, since inconsistent and erroneous data may lead to inaccurate prediction models. Although, metabolic reaction prediction is an extremely challenging area, it has greatly facilitated the advancement of drug discovery while continuing to show rapid improvement with evolving computational techniques and increased computational power.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements and grants

This research was supported by a National Research Foundation of Korea grant funded by the Korean government (NRF-2016R1D1A1B03935264) and was also supported by the Bio-Synergy Research Project (NRF-2018M3A9C4076474) of the Ministry of Science, ICT and Future Planning through the National Research Foundation.

References

[1] M.A. Lazar, M.J. Birnbaum, Physiology. De-meaning of metabolism, Science 336 (2012) 1651–1652.
[2] R.A. Cairns, I.S. Harris, T.W. Mak, Regulation of cancer cell metabolism, Nat. Rev. Cancer 11 (2011) 85–95.
[3] Y.P. Kang, N.P. Ward, G.M. DeNicola, Recent advances in cancer metabolism: a technological perspective, Exp. Mol. Med. 50 (2018) 31.
[4] P.I. Mackenzie, A.A. Somogyi, J.O. Miners, Advances in drug metabolism and pharmacogenetics research in Australia, Pharmacol. Res. 116 (2017) 7–19.
[5] S. Bernardini, A. Tiezzi, V. Laghezza Masci, E. Ovidi, Natural products for human health: an historical overview of the drug discovery approaches, Nat. Prod. Res. 32 (2018) 1926–1950.
[6] L.A. Stanley, Chapter 27 - drug metabolism A2 - badal, simone, in: R. Delgoda (Ed.), Pharmacognosy, Academic Press, Boston, 2017, pp. 527–545.
[7] U.M. Zanger, M. Schwab, Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation, Pharmacol. Ther. 138 (2013) 103–141.
[8] J.F. Rogers, A.N. Nafziger, J.S. Bertino Jr., Pharmacogenetics affects dosing, efficacy, and toxicity of cytochrome P450-metabolized drugs, Am. J. Med. 113 (2002) 746–750.

[9] G.B. Stefano, R. Ptacek, H. Kuzelova, R.M. Kream, Endogenous morphine: up-to-date review 2011, Folia Biol 58 (2012) 49–56.
[10] L.C. Kirkwood, R.L. Nation, A.A. Somogyi, Characterization of the human cytochrome P450 enzymes involved in the metabolism of dihydrocodeine, Br. J. Clin. Pharmacol. 44 (1997) 549–555.
[11] M. Shadid, Y. Liu, E. Veizaj, J. Huang, J. Johnson, A. Kamel, Evaluation of rat brain subcellular fractions for the prediction of *in vivo* drug metabolism in brain, Drug Metab. Pharmacol. 33 (2018) S79.
[12] X. Diao, J. Carlier, M. Zhu, S. Pang, R. Kronstrand, K.B. Scheidweiler, M.A. Huestis, *In vitro* and *in vivo* human metabolism of a new synthetic cannabinoid NM-2201 (CBL-2201), Forensic Toxicol. 35 (2017) 20–32.
[13] S.A. Wrighton, M. Vandenbranden, J.C. Stevens, L.A. Shipley, B.J. Ring, A.E. Rettie, J.R. Cashman, *In vitro* methods for assessing human hepatic drug metabolism: their use in drug development, Drug Metab. Rev. 25 (1993) 453–484.
[14] J.A. DiMasi, H.G. Grabowski, R.W. Hansen, Innovation in the pharmaceutical industry: new estimates of R&D costs, J. Health Econ. 47 (2016) 20–33.
[15] S.M. Paul, D.S. Mytelka, C.T. Dunwiddie, C.C. Persinger, B.H. Munos, S.R. Lindborg, A.L. Schacht, How to improve R&D productivity: the pharmaceutical industry's grand challenge, Nat. Rev. Drug Discovery 9 (2010) 203–214.
[16] S. Alqahtani, *In silico* ADME-Tox modeling: progress and prospects, Expert Opin. Drug Metab. Toxicol. 13 (2017) 1147–1158.
[17] C.H. Andrade, D.C. Silva, R.C. Braga, *In silico* prediction of drug metabolism by P450, Curr. Drug Metab. 15 (2014) 514–525.
[18] B.H. Tan, Y. Pan, A.N. Dong, C.E. Ong, In *vitro* and *in silico* approaches to study cytochrome P450-mediated interactions, J. Pharm. Sci. 20 (2017) 319–328.
[19] T. Zhang, Q. Chen, L. Li, L.A. Liu, D.Q. Wei, *In silico* prediction of cytochrome P450-mediated drug metabolism, Comb. Chem. High Throughput Screening 14 (2011) 388–395.
[20] M.P. Gleeson, A. Hersey, S. Hannongbua, *In-silico* ADME models: a general assessment of their utility in drug discovery applications, Curr. Top. Med. Chem. 11 (2011) 358–381.
[21] D. Schuster, T.M. Steindl, T. Langer, Predicting drug metabolism induction *in silico*, Curr. Top. Med. Chem. 6 (2006) 1627–1640.
[22] T. Fox, J.M. Kriegl, Machine learning techniques for *in silico* modeling of drug metabolism, Curr. Top. Med. Chem. 6 (2006) 1579–1591.
[23] J.K. Lu Tan, Software for metabolism prediction, in: J. Kirchmair (Ed.), Drug Metabolism Prediction, 2014.
[24] C. Hansch, T. Fujita, p-σ-π analysis. A method for the correlation of biological activity and chemical structure, J. Am. Chem. Soc. 86 (1964) 1616–1626.
[25] S.M. Free Jr., J.W. Wilson, A mathematical contribution to structure-activity studies, J. Med. Chem. 7 (1964) 395–399.
[26] J.C. Gertrudes, V.G. Maltarollo, R.A. Silva, P.R. Oliveira, K.M. Honorio, A.B. da Silva, Machine learning techniques and drug design, Curr. Med. Chem. 19 (2012) 4289–4297.
[27] Q.S. Du, R.B. Huang, K.C. Chou, Recent advances in QSAR and their applications in predicting the activities of chemical molecules, peptides and proteins for drug design, Curr. Protein Pept. Sci. 9 (2008) 248–259.
[28] S.H. Oraphan Phuangsawai, Mathew Paul Gleeson, Quantitative structure–activity relationship (QSAR) methods for the prediction of substrates, inhibitors, and inducers of metabolic enzymes, in: J. Kirchmair (Ed.), Drug Metabolism Prediction, 2014.
[29] A.C. Brown, T.R. Fraser, On the connection between chemical constitution and physiological action; with special reference to the physiological action of the salts of the ammonium bases derived from strychnia, brucia, thebaia, codeia, morphia, and nicotia, J. Anat. Physiol. 2 (1868) 224–242.
[30] S. Kausar, A.O. Falcao, An automated framework for QSAR model building, J. Cheminform. 10 (2018) 1.
[31] G. Domenico, M. Giuseppe Felice, C. Marco, C. Angelo, N. Orazio, Applicability domain for QSAR models: where theory meets reality, International journal of quantitative structure-property relationships (IJQSPR) 1 (2016) 45–63.
[32] K. Roy, S. Kar, R.N. Das, Chapter 2 - Chemical Information and Descriptors, Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment, Academic Press, Boston, 2015, pp. 47–80.
[33] P. Gramatica, A Short History of QSAR Evolution, (2011).
[34] E. Stjernschantz, N.P.E. Vermeulen, C. Oostenbrink, Computational prediction of drug binding and rationalisation of selectivity towards cytochromes P450, Expert Opin. Drug Metab. Toxicol. 4 (2008) 513–527.
[35] M.P. Gleeson, A.M. Davis, K.K. Chohan, S.W. Paine, S. Boyer, C.L. Gavaghan, C.H. Arnby, C. Kankkonen, N. Albertson, Generation of *in-silico* cytochrome P450 1A2, 2C9, 2C19, 2D6, and 3A4 inhibition QSAR models, J. Comput. Aided Mol. Des. 21 (2007) 559–573.
[36] D.F. Lewis, S. Modi, M. Dickins, Structure-activity relationship for human cytochrome P450 substrates and inhibitors, Drug Metab. Rev. 34 (2002) 69–82.
[37] A. Kurup, C-QSAR: a database of 18,000 QSARs and associated biological and physical data, J. Comput. Aided Mol. Des. 17 (2003) 187–196.
[38] V. Ruusmann, S. Sild, U. Maran, QSAR databank repository: open and linked qualitative and quantitative structure-activity relationship models, J. Cheminform. 7 (2015) 32.
[39] Y.C. Lo, S.E. Rensi, W. Torng, R.B. Altman, Machine learning in chemoinformatics and drug discovery, Drug Discov. Today 23 (2018) 1538–1546.
[40] R.R. Ramsay, M.R. Popovic-Nikolic, K. Nikolic, E. Uliassi, M.L. Bolognesi, A perspective on multi-target drug discovery and design for complex diseases, Clin. Transl. Med. 7 (2018).
[41] L. Zaccagnini, S. Brogi, M. Brindisi, S. Gemma, G. Chemi, G. Legname, G. Campiani, S. Butini, Identification of novel fluorescent probes preventing PrP (Sc) replication in prion diseases, Eur. J. Med. Chem. 127 (2017) 859–873.
[42] R.D. Cramer, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[43] J.M. Luco, F.H. Ferretti, QSAR based on multiple linear regression and PLS methods for the anti-HIV activity of a large group of HEPT derivatives, J. Chem. Inf. Comput. Sci. 37 (1997) 392–401.
[44] O. Ivanciuc, Machine learning quantitative structure-activity relationships (QSAR) for peptides binding to the human amphiphysin-1 SH3 domain, Curr. Proteomics 6 (2009) 289–302.
[45] R.D. King, J.D. Hirst, M.J.E. Sternberg, New approaches to QSAR: neural networks and machine learning, Perspect. Drug Discovery Des. 1 (1993) 279–290.
[46] M.I. Jordan, T.M. Mitchell, Machine learning: trends, perspectives, and prospects, Science 349 (2015) 255–260.
[47] N. Fleming, How artificial intelligence is changing drug discovery, Nature 557 (2018) S55–S57.
[48] J.B.O. Mitchell, Machine learning methods in chemoinformatics, Wiley Interdiscip. Rev.: Comput. Mol. Sci. 4 (2014) 468–481.
[49] N. Singh, P. Shah, H. Dwivedi, S. Mishra, R. Tripathi, A.A. Sahasrabuddhe, M.I. Siddiqi, Integrated machine learning, molecular docking and 3D-QSAR based approach for identification of potential inhibitors of trypanosomal N-myristoyltransferase, Mol. Biosyst. 12 (2016) 3711–3723.
[50] Y. Cong, B.-k. Li, X.-g. Yang, Y. Xue, Y.-z. Chen, Y. Zeng, Quantitative structure–activity relationship study of influenza virus neuraminidase A/PR/8/34 (H1N1) inhibitors by genetic algorithm feature selection and support vector regression, Chemom. Intell. Lab. Syst. 127 (2013) 35–42.
[51] L. Zhang, J. Tan, D. Han, H. Zhu, From machine learning to deep learning: progress in machine intelligence for rational drug discovery, Drug Discov. Today 22 (2017) 1680–1685.
[52] Y.-C. Lo, S.E. Rensi, W. Torng, R.B. Altman, Machine learning in chemoinformatics and drug discovery, Drug Discov. Today 23 (2018) 1538–1546.
[53] Y. Sakiyama, H. Yuki, T. Moriya, K. Hattori, M. Suzuki, K. Shimada, T. Honma, Predicting human liver microsomal stability with machine learning techniques, J. Mol. Graph. Model. 26 (2008) 907–915.
[54] J.S. Smith, A.E. Roitberg, O. Isayev, Transforming computational drug discovery with machine learning and AI, ACS Med. Chem. Lett. 9 (2018) 1065–1069.
[55] H. Li, C.W. Yap, C.Y. Ung, Y. Xue, Z.R. Li, L.Y. Han, H.H. Lin, Y.Z. Chen, Machine learning approaches for predicting compounds that interact with therapeutic and ADMET mated proteins, J. Pharm. Sci. 96 (2007) 2838–2860.
[56] V.G. Maltarollo, J.C. Gertrudes, P.R. Oliveira, K.M. Honorio, Applying machine learning techniques for ADME-Tox prediction: a review, Expert Opin. Drug Metab. Toxicol. 11 (2015) 259–271.
[57] S. Wang, H. Sun, H. Liu, D. Li, Y. Li, T. Hou, ADMET evaluation in drug discovery. 16. Predicting hERG blockers by combining multiple pharmacophores and machine learning approaches, Mol. Pharm. 13 (2016) 2855–2866.
[58] V.G. Maltarollo, J.C. Gertrudes, P.R. Oliveira, K.M. Honorio, Applying machine learning techniques for ADME-Tox prediction: a review, Expert Opin. Drug Metab. Toxicol. 11 (2015) 259–271.
[59] S. Ekins, The next era: deep learning in pharmaceutical research, Pharm. Res. (N. Y.) 33 (2016) 2594–2603.
[60] E. Gawehn, J.A. Hiss, G. Schneider, 6 deep learning in drug discovery, Mol. Inf. 35 (2016) 3–14.
[61] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.
[62] S. Min, B. Lee, S. Yoon, Deep learning in bioinformatics, Briefings Bioinf. 18 (2017) 851–869.
[63] A. Gupta, A.T. Muller, B.J.H. Huisman, J.A. Fuchs, P. Schneider, G. Schneider, Generative recurrent networks for *de novo* drug design, Mol. Inf. 37 (2018).
[64] H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, T. Blaschke, The rise of deep learning in drug discovery, Drug Discov, Today Off. 23 (2018) 1241–1250.
[65] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, E.I.C. Chang, Deep learning of feature representation with multiple instance learning for medical image analysis, IEEE Trans. Acoust. Speech Signal Process. (2014) 1626–1630.
[66] C.W. Yap, Z.R. Li, Y.Z. Chen, Quantitative structure–pharmacokinetic relationships for drug clearance by using statistical learning methods, J. Mol. Graph. Model. 24 (2006) 383–395.
[67] M. Spencer, J. Eickholt, C. Jianlin, A deep learning network approach to *ab initio* protein secondary structure prediction, IEEE/ACM Trans. Comput. Biol. Bioinform. 12 (2015) 103–112.
[68] S. Wang, J. Peng, J.Z. Ma, J.B. Xu, Protein secondary structure prediction using deep convolutional neural fields, Sci. Rep. 6 (2016).
[69] T. Jo, J. Hou, J. Eickholt, J. Cheng, Improving protein fold recognition by deep learning networks, Sci. Rep. 5 (2015) 17573.
[70] A.M.C. Sean Ekins, Alexander L. Perryman, Joel S. Freundlich, Alexandru Korotcov, Valery Tkachenko, Accessible machine learning approaches for toxicology, in: Ekins (Ed.), Comput. Toxicol, 2018.
[71] Baskin II, D. Winkler, I.V. Tetko, A renaissance of neural networks in drug discovery, Expert Opin, Drug Discovery 11 (2016) 785–795.
[72] G. Klambauer, T. Unterthiner, A. Mayr, S. Hochreiter, DeepTox: toxicity prediction using deep learning, Toxicol. Lett. 280 (2017) S69-S69.
[73] K.D. Wu, G.W. Wei, Quantitative toxicity prediction using topology based multitask deep neural networks, J. Chem. Inf. Model. 58 (2018) 520–531.
[74] L. Zhang, H. Zhang, H. Ai, H. Hu, S. Li, J. Zhao, H. Liu, Applications of machine learning methods in drug toxicity prediction, Curr. Top. Med. Chem. 18 (2018) 987–997.
[75] A. Roncaglioni, A.A. Toropov, A.P. Toropova, E. Benfenati, *In silico* methods to predict drug toxicity, Curr. Opin. Pharmacol. 13 (2013) 802–806.
[76] A. Lavecchia, Machine-learning approaches in drug discovery: methods and

applications, Drug Discov, Today Off. 20 (2015) 318–331.
[77] R. Venkatapathy, N.C. Wang, Developmental toxicity prediction, Methods Mol. Biol. 930 (2013) 305–340.
[78] W.L. Jorgensen, The many roles of computation in drug discovery, Science 303 (2004) 1813–1818.
[79] D.B. Kitchen, H. Decornez, J.R. Furr, J. Bajorath, Docking and scoring in virtual screening for drug discovery: methods and applications, Nat. Rev. Drug Discovery 3 (2004) 935–949.
[80] M. von Itzstein, W.-Y. Wu, G.B. Kok, M.S. Pegg, J.C. Dyason, B. Jin, T. Van Phan, M.L. Smythe, H.F. White, S.W. Oliver, P.M. Colman, J.N. Varghese, D.M. Ryan, J.M. Woods, R.C. Bethell, V.J. Hotham, J.M. Cameron, C.R. Penn, Rational design of potent sialidase-based inhibitors of influenza virus replication, Nature 363 (1993) 418.
[81] H. Kubinyi, Chance favors the prepared mind–from serendipity to rational drug design, J. Recept. Signal Transduction 19 (1999) 15–39.
[82] M. Zheng, X. Luo, Q. Shen, Y. Wang, Y. Du, W. Zhu, H. Jiang, Site of metabolism prediction for six biotransformations mediated by cytochromes P450, Bioinformatics 25 (2009) 1251–1258.
[83] A. Ganesan, M.L. Coote, K. Barakat, Molecular dynamics-driven drug discovery: leaping forward with confidence, Drug Discov. Today 22 (2017) 249–269.
[84] H.M. Li, H. Xiao, L. Lin, D. Jou, V. Kumari, J. Lin, C.L. Li, Drug design targeting protein-protein interactions (PPIs) using multiple ligand simultaneous docking (MLSD) and drug repositioning: discovery of raloxifene and bazedoxifene as novel inhibitors of IL-6/GP130 interface, J. Med. Chem. 57 (2014) 632–641.
[85] H. Alonso, A.A. Bliznyuk, J.E. Gready, Combining docking and molecular dynamic simulations in drug design, Med. Res. Rev. 26 (2006) 531–568.
[86] O. Trott, A.J. Olson, AutoDock Vina, Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2010) 455–461.
[87] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Bio. 267 (1997) 727–748.
[88] W.J. Allen, T.E. Balius, S. Mukherjee, S.R. Brozell, D.T. Moustakas, P.T. Lang, D.A. Case, I.D. Kuntz, R.C. Rizzo, DOCK 6: impact of new features and current docking performance, J. Comput. Chem. 36 (2015) 1132–1156.
[89] C.C. Melo-Filho, R.C. Braga, C.H. Andrade, Advances in methods for predicting phase I metabolism of polyphenols, Curr. Drug Metab. 15 (2014) 120–126.
[90] Y. Shan, E.T. Kim, M.P. Eastwood, R.O. Dror, M.A. Seeliger, D.E. Shaw, How does a drug molecule find its target binding site? J. Am. Chem. Soc. 133 (2011) 9181–9183.
[91] A. Grosdidier, V. Zoete, O. Michielin, Fast docking using the CHARMM force field with EADock DSS, J. Comput. Chem. 32 (2011) 2149–2159.
[92] M.L. Teodoro, G.N. Phillips, L.E. Kavraki, Molecular docking: a problem with thousands of degrees of freedom, IEEE Int. Conf. Robot. Autom. (2001) 960–965.
[93] D.E. Koshland, The key-lock theory and the induced fit theory, Angew. Chem. Int. Ed. 33 (1994) 2375–2378.
[94] R.P. Shetty, P.I.W. de Bakker, M.A. DePristo, T.L. Blundell, Advantages of fine-grained side chain conformer libraries, Protein Eng. 16 (2003) 963–969.
[95] N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, C.R. Corbeil, Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go, Br. J. Clin. Pharmacol. 153 (2008) S7–S26.
[96] C. B-Rao, J. Subramanian, S.D. Sharma, Managing protein flexibility in docking and its applications, Drug Discov, Today Off. 14 (2009) 394–400.
[97] H.A. Carlson, Protein flexibility and drug design: how to hit a moving target, Curr. Opin. Chem. Biol. 6 (2002) 447–452.
[98] Y.Z. Chen, C.Y. Ung, Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand-protein inverse docking approach, J. Mol. Graph. Model. 20 (2001) 199–218.
[99] Z. Yao, Z. Lin, T. Wang, D. Tian, X. Zou, Y. Gao, D. Yin, Using molecular docking-based binding energy to predict toxicity of binary mixture with different binding sites, Chemosphere 92 (2013) 1169–1176.
[100] Z.L. Ji, Y. Wang, L. Yu, L.Y. Han, C.J. Zheng, Y.Z. Chen, *In silico* search of putative adverse drug reaction related proteins as a potential tool for facilitating drug adverse effect prediction, Toxicol. Lett. 164 (2006) 104–112.
[101] C. Ma, H. Kang, Q. Liu, R.X. Zhu, Z.W. Cao, Insight into potential toxicity mechanisms of melamine: an *in silico* study, Toxicology 283 (2011) 96–100.
[102] J. Rautio, N.A. Meanwell, L. Di, M.J. Hageman, The expanding role of prodrugs in contemporary drug design and development, Nat. Rev. Drug Discovery 17 (2018) 559–587.
[103] C. George Priya Doss, C. Chakraborty, V. Narayan, D. Thirumal Kumar, Chapter ten - computational approaches and resources in single amino acid substitutions analysis toward clinical research, in: R. Donev (Ed.), Adv. Protein Chem. Struct. Biol. Academic Press, 2014, pp. 365–423.
[104] G.R. Wilkinson, Drug therapy - drug metabolism and variability among patients in drug response, N. Engl. J. Med. 352 (2005) 2211–2221.
[105] T. Lynch, A. Price, The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects, Am. Fam. Physician 76 (2007) 391–396.
[106] J.L. Pilgrim, D. Gerostamoulos, O.H. Drummer, Review: pharmacogenetic aspects of the effect of cytochrome P450 polymorphisms on serotonergic drug metabolism, response, interactions, and adverse effects, Forensic Sci. Med. Pathol. 7 (2011) 162–184.
[107] S.F. Zhou, Drugs behave as substrates, inhibitors and inducers of human cytochrome P450 3A4, Curr. Drug Metab 9 (2008) 310–322.
[108] J.H. Lin, M. Yamazaki, Role of P-glycoprotein in pharmacokinetics: clinical implications, Clin. Pharmacokinet. 42 (2003) 59–98.
[109] J.W. Ko, Z. Desta, N.V. Soukhova, T. Tracy, D.A. Flockhart, *In vitro* inhibition of the cytochrome P450 (CYP450) system by the antiplatelet drug ticlopidine: potent effect on CYP2C19 and CYP2D6, Br. J. Clin. Pharmacol. 49 (2000) 343–351.
[110] M.A. Barrand, P-Glycoprotein, in: S. Brenner, J.H. Miller (Eds.), Encyclopedia of Genetics, Academic Press, New York, 2001, pp. 1430–1433.
[111] T. Akabane, N. Gerst, Y. Naritomi, J.N. Masters, K. Tamura, A practical and direct comparison of intrinsic metabolic clearance of several non-CYP enzyme substrates in freshly isolated and cryopreserved hepatocytes, Drug Metab. Pharmacokinet. 27 (2012) 181–191.
[112] Y. Cai, H. Yang, W. Li, G. Liu, P.W. Lee, Y. Tang, Multiclassification prediction of enzymatic reactions for oxidoreductases and hydrolases using reaction fingerprints and machine learning methods, J. Chem. Inf. Model. 58 (2018) 1169–1181.
[113] B. Niu, G. Huang, L. Zheng, X. Wang, F. Chen, Y. Zhang, T. Huang, Prediction of substrate-enzyme-product interaction based on molecular descriptors and physicochemical properties, BioMed Res. Int. (2013) 2013.
[114] I. Schomburg, A. Chang, C. Ebeling, M. Gremse, C. Heldt, G. Huhn, D. Schomburg, BRENDA, the enzyme database: updates and major new developments, Nucleic Acids Res. 32 (2004) D431–D433.
[115] D.S. Wishart, C. Knox, A.C. Guo, R. Eisner, N. Young, B. Gautam, D.D. Hau, N. Psychogios, E. Dong, S. Bouatra, R. Mandal, I. Sinelnikov, J. Xia, L. Jia, J.A. Cruz, E. Lim, C.A. Sobsey, S. Shrivastava, P. Huang, P. Liu, L. Fang, J. Peng, R. Fradette, D. Cheng, D. Tzur, M. Clements, A. Lewis, A. De Souza, A. Zuniga, M. Dawe, Y. Xiong, D. Clive, R. Greiner, A. Nazyrova, R. Shaykhutdinov, L. Li, H.J. Vogel, I. Forsythe, HMDB: a knowledgebase for the human metabolome, Nucleic Acids Res. 37 (2009) D603–D610.
[116] M.S. Yu, H.M. Lee, A. Park, C. Park, H. Ceong, K.H. Rhee, D. Na, *In silico* prediction of potential chemical reactions mediated by human enzymes, BMC Bioinf. 19 (2018) 207.
[117] M.J. Keiser, B.L. Roth, B.N. Armbruster, P. Ernsberger, J.J. Irwin, B.K. Shoichet, Relating protein pharmacology by ligand chemistry, Nat. Biotechnol. 25 (2007) 197–206.
[118] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, M. Kanehisa, Prediction of drug–target interaction networks from the integration of chemical and genomic spaces, Bioinformatics 24 (2008) i232–i240.
[119] L.L. Zhao, W.Y. Wang, A. Sedykh, H. Zhu, Experimental errors in QSAR modeling sets: what we can do and what we cannot do, ACS Omega 2 (2017) 2805–2812.
[120] A.B. Raies, V.B. Bajic, *In silico* toxicology: computational methods for the prediction of chemical toxicity, Wiley Interdiscip. Rev. Comput. Mol. Sci. 6 (2016) 147–172.
[121] H.J. Klimisch, M. Andreae, U. Tillmann, A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data, Regul. Toxicol. Pharmacol. 25 (1997) 1–5.
[122] M. Nendza, T. Aldenberg, E. Benfenati, R. Benigni, M.T.D. Cronin, S. Escher, A. Fernández, S. Gabbert, F. Giralt, M. Hewitt, M. Hrovat, S. Jeram, D. Kroese, J. Madden, I. Mangelsdorf, R. Rallo, A. Roncaglioni, E. Rorije, H. Segner, T. Vemeire, Data Quality Assessment for *in Silico* Methods: a Survey of Approaches and Needs, (2010), pp. 59–117.
[123] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: a major update to the drugbank database for 2018, Nucleic Acids Res. 46 (2018) D1074–D1082.
[124] A. Rudik, A. Dmitriev, A. Lagunin, D. Filimonov, V. Poroikov, SOMP: web server for *in silico* prediction of sites of metabolism for drug-like compounds, Bioinformatics 31 (2015) 2046–2048.