Computational and Structural Biotechnology Journal 17 (2019) 1367–1376

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

Mini Review

Current computational methods for predicting protein interactions of natural products

Aurélien F.A. Moumbock, Jianyu Li, Pankaj Mishra, Mingjie Gao, Stefan Günther *

Institute of Pharmaceutical Sciences, Research Group Pharmaceutical Bioinformatics, Albert-Ludwigs-Universität Freiburg, Germany

ARTICLE INFO

Article history:
Received 19 March 2019
Received in revised form 9 August 2019
Accepted 23 August 2019
Available online 28 October 2019

Keywords:
Natural products
Pharmacological space
Drug-target interactions
Virtual screening
Target fishing
Drug discovery

ABSTRACT

Natural products (NPs) are an indispensable source of drugs and they have a better coverage of the pharmacological space than synthetic compounds, owing to their high structural diversity. The prediction of their interaction profiles with druggable protein targets remains a major challenge in modern drug discovery. Experimental (off-)target predictions of NPs are cost- and time-consuming, whereas computational methods, on the other hand, are much faster and cheaper. As a result, computational predictions are preferentially used in the first instance for NP profiling, prior to experimental validations. This review covers recent advances in computational approaches which have been developed to aid the annotation of unknown drug-target interactions (DTIs), by focusing on three broad classes, namely: ligand-based, target-based, and target—ligand-based (hybrid) approaches. Computational DTI prediction methods have the potential to significantly advance the discovery and development of novel selective drugs exhibiting minimal side effects. We highlight some inherent caveats of these methods which must be overcome to enable them to realize their full potential, and a future outlook is given.

Contents

* Corresponding author.
E-mail address: stefan.guenther@pharmazie.uni-freiburg.de (S. Günther).

https://doi.org/10.1016/j.csbj.2019.08.008

1. Introduction

Since the earliest times, for the treatment of diseases, humans have heavily depended on medicinal plants whose "active principles" are secondary metabolites termed natural products (NPs). Precisely, NPs are "genetically encoded small molecules" originating from microorganisms, plants, or animals [1,2]. They have better coverage of the biologically relevant chemical space (pharmacological space) than synthetic molecules. It is estimated that about 60% of all medicines approved in the last three decades are either NPs or their semisynthetic derivatives [3–5]. Notable examples of approved drugs of NP origin (Fig. 1) include: the antibiotic penicillin G, isolated from the fungus *Penicillium chrysogenum*; the antibiotic streptomycin, isolated from the bacterium *Streptomyces griseus*; the anthelmintics avermectins (B_{1a} and B_{1b}), isolated from the bacterium *Streptomyces avermitilis*, and the antimalarial artemisinin, isolated from the plant *Artemisia annua*. Their discoverers received the Nobel Prize (in Physiology or Medicine) in 1945, 1952, and 2015, respectively [6]. There is a huge number of secondary metabolites annotated in focused chemical libraries such as StreptomeDB 2.0 [7] and NANPDB [8], which have not yet been investigated for their medicinal potential. Furthermore, for the vast majority of NPs whose activities have been evaluated in bioassays, their interaction profiles with drug targets (mostly proteins) are still unknown.

The "magic bullet" concept formulated in 1900 by Paul Ehrlich, is the foundation of single-target pharmacology. It states that a compound will exhibit a given biological activity unless it binds to a specific target [9,10]. This principle has been successfully applied during the last century in the design of numerous approved drugs. However, the development of specific binders is a challenging task and many drugs have been withdrawn from the market due to their undesirable side effects, resulting from their target promiscuity. In recent years, there has been a quantum leap from single-target pharmacology to multi-target pharmacology (polypharmacology). With increasing knowledge about drug–target interactions (DTIs), more effective drugs can be developed by specifically modulating multiple targets simultaneously [11,12]. Polypharmacology can therefore be an asset in synergistic therapy.

Generally, NPs have high structural diversity and complexity, and very often exhibit target promiscuity. Bearing in mind that high throughput *in vitro*/*vivo* experiments for studying the polypharmacology of NPs are cost- and time-consuming, highly efficient prospective *in silico* predictions could serve as promising, rapid, and cost-effective strategies to decipher NP–target associations, prior to experimental validation [13,14]. The prediction of ligand–receptor interactions, most commonly known as DTIs, is carried out in several stages of the drug discovery and development process, for on-target as well as off-target interactions. DTI prediction, and thereby prediction of the mechanism of action, can either be performed in a forward manner for virtual screening to predict putative ligands of a given druggable target, or in a reverse manner for target fishing to predict putative target proteins of bioactive ligand(s) [15–17].

In this review, we focus on the three current approaches dealing with computational DTI prediction, namely ligand-based, target-based, and target–ligand-based (hybrid) approaches (Fig. 2).

2. Computational methods for DTI prediction

2.1. Ligand-based approaches

These methods stem from the chemical similarity principle, which states that similar molecules typically have similar physicochemical properties and bind to similar drug targets [18]. Based on this principle, ligand-based similarity approaches predict DTIs via comparison of query ligands to known active ligands of a specific drug target. They are the methods of choice for drug targets whose macromolecular structures have not yet been solved, such as several G-protein-coupled receptors (GPCRs), transporters, or ion channels [18,19]. Ligand-based similarity comparisons can be subdivided into pharmacophore modeling, chemical similarity searching, and quantitative structure–activity relationship (QSAR).

2.1.1. Pharmacophore screening

Historically, the concept of pharmacophore was formulated by Paul Ehrlich in 1909 [20,21]. According to IUPAC, a pharmacophore is defined as "an ensemble of steric and electronic features that is

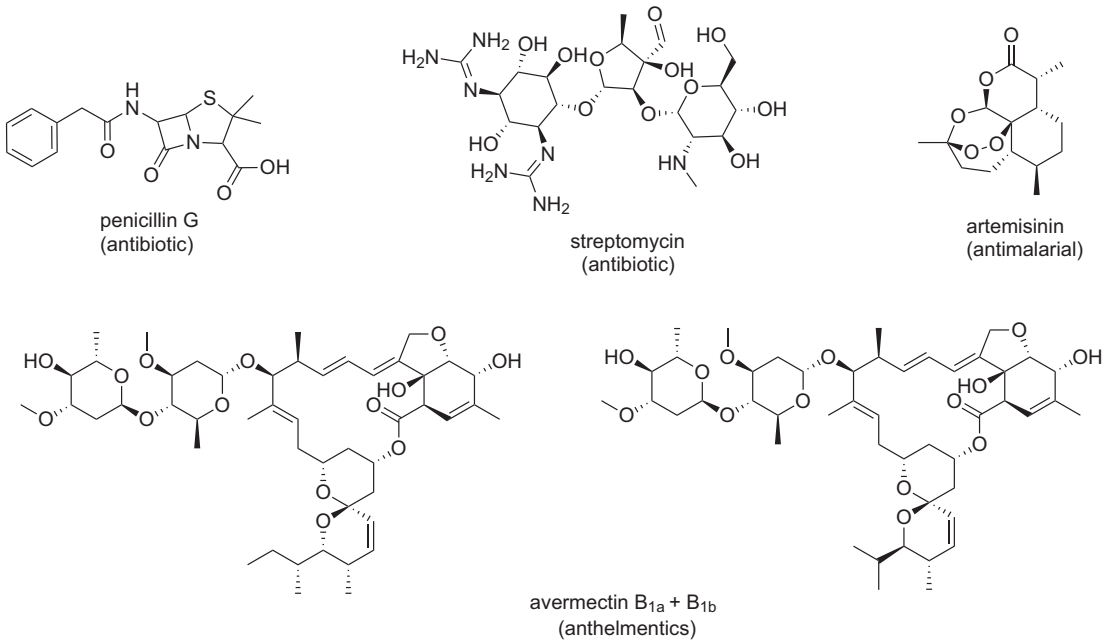

Fig. 1. Structures of some notable approved drugs of NP origin.

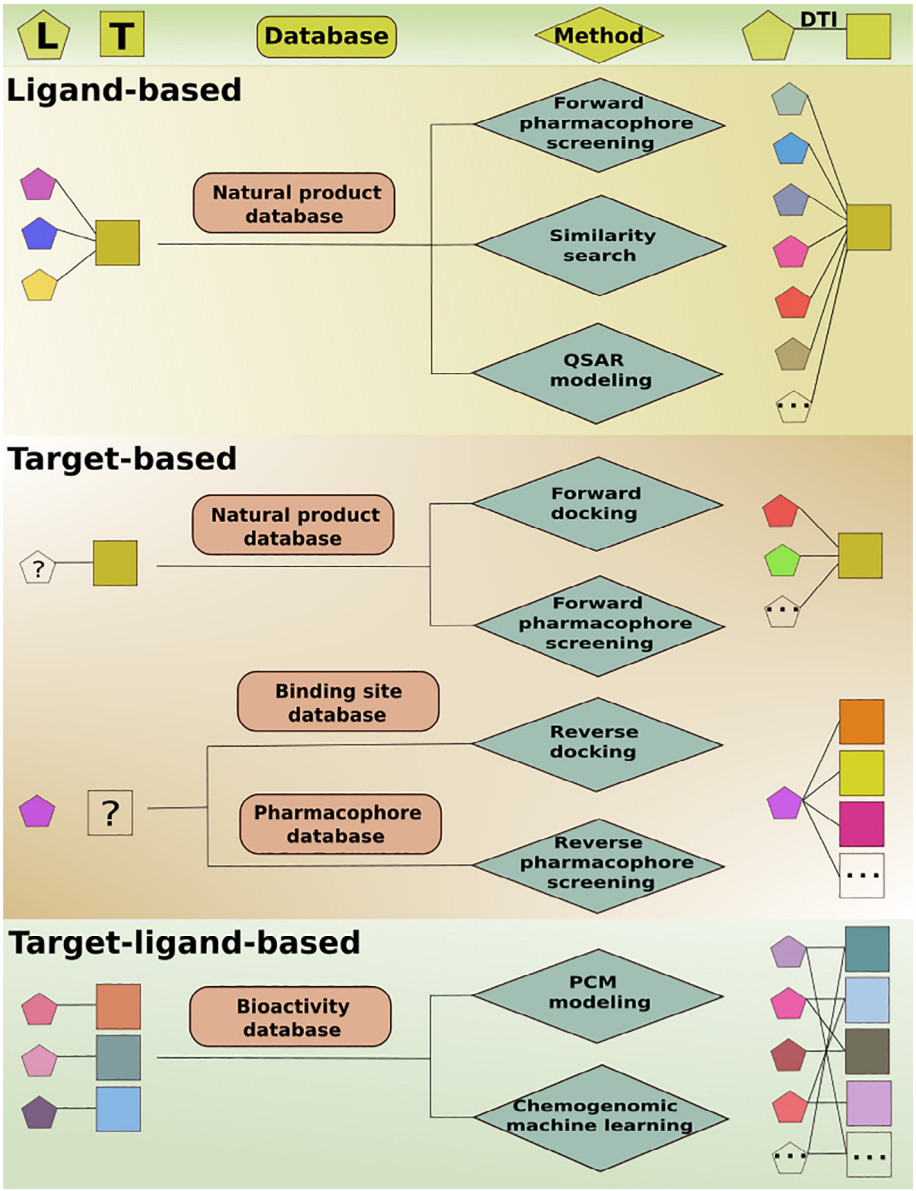

Fig. 2. Overview of computational approaches for DTI prediction; L and T represent ligand (including NPs and synthetic drugs) and target, respectively.

necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response" [22]. These pharmacophoric features include mainly aromatic, hydrophobic, charged ionizable and hydrogen bonding moieties. Pharmacophore perception involves the overlap of energy minimized conformations of a set of known active ligands and the extraction of the recurrent pharmacophoric features in a single model. Once a pharmacophore model has been generated, a query can be done using database molecules in a forward manner in search of novel putative hits, or in a reverse manner when a ligand is compared with multiple pharmacophore models in search of putative targets (parallel screening) [23].

Generally, the pharmacophore query is done by the overlay of generated 3D conformers and tautomers of each database molecule onto the pharmacophore model derived from bioactive ligands to identify the maximal common subsets [24,25]. Alternatively, a bit-wise comparison of generated fingerprints of the pharmacophore model and those of the database molecules is made. Pharmacophoric fingerprints are bit strings encoding distances between sets of three (or four) pharmacophoric points in a ligand structure, counted in bonds and distance-binning at the 2D and 3D levels, respectively [25,26]. The fit between a given query ligand and pharmacophore model can be measured either by rmsd-based or overlay-based scoring functions. The former scoring functions are superior in predicting the highest number of hits for large chemical libraries, whereas the latter have the advantage of producing the highest ratio of correct/incorrect hits [27,28]. Some of the most popular programs used for pharmacophore modeling/search are Pharmer [29], Discovery Studio [30], LigandScout [31], Phase [32], Screen [33], and MOE [34]. Pharmacophore web servers include ZINCPharmer [35], PharmMapper [36], Pharmit [37], and CavityPlus [38]. Kirchweger *et al.* [39], used the pharmacophore program LigandScout [31] to generate two ligand-based pharmacophore models from known activators of the G protein-coupled bile acid receptor 1 (GPBAR1). These models were used to screen an NP library, leading to the identification of two NPs, farnesiferol B and microlobidene, which were confirmed to activate GPBAR1 with potencies similar to that of the endogenous ligand, lithocholic acid (Fig. 3).

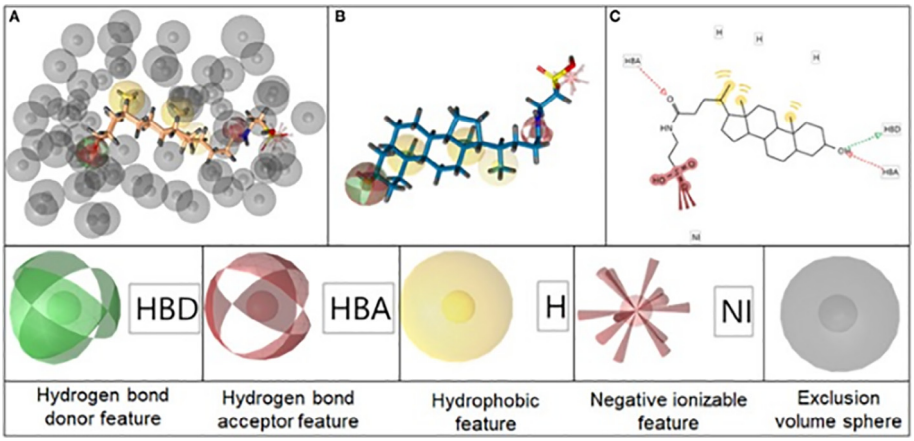

Fig. 3. Representation of one of the generated pharmacophore hypotheses, aligned to lithocholic acid in 3D with exclusion volume spheres (A), without exclusion volumes (B), and in 2D (C) [39]. The original figure was published under a Creative Commons License.

Due to advances in techniques for macromolecular structure determination, the paradigm has moved from ligand-centric to receptor-centric pharmacophore modeling. Briefly, 3D pharmacophoric features here are established on the ligand within the binding pocket of its co-crystallised protein [40–42]. During a receptor-centric pharmacophoric query, excluded volume spheres, corresponding to spatial positions occupied by the protein side chains, are usually added as constraints. This is done in order to ensure shape complementarity of the matches, meanwhile occasioning unfavorable steric clashes for bogus hits. Three databases exist which contain pharmacophore models extracted from PDB protein—ligand complexes, namely PharmaDB [42], PharmTargetDB [36], and Inte:PharmacophoreDB [43]. These databases are often used for target fishing of NPs, by implementation in a pharmacophore software. Rollinger *et al.* [44] used the latter database, along with the software Discovery Studio [30], to identify putative targets for 16 NPs isolated from the medicinal plant *Ruta graveolens*. These NPs exhibited *in vitro* micromolar inhibitory concentrations (IC_{50}) to acetylcholinesterase, the human rhinovirus coat protein and the cannabinoid receptor type-2, identified from target fishing.

2.1.2. Chemical similarity searching

In the late 1980s, chemical similarity screening (also called nearest-neighbor searching or shape screening) was reported as an alternative to pharmacophore modeling [45,46]. It involves the use of a similarity metric to assess the global intermolecular structural similarity between a query structure and each compound in a database, with the most-similar structures (nearest-neighbors) emerging as the top-ranked by the metric. The query (reference) structure can either be a whole molecule or a substructure (e.g. a "privileged scaffold"). In this approach, the molecules are structurally represented by 2D/3D molecular descriptors, principally fingerprints which can be either circular-, topological-, or substructure keys-based [26,47–49]. A molecular fingerprint is an advanced form of the fundamental structural key. Unlike its precursor, the molecular fingerprint does not use predefined sets of structural patterns, and consequently has in general a higher information content and is less computationally expensive. However, similarity indices are highly dependent on the subjected chemical properties (such as the size of the molecule) or the relevance of specific chemical features (such as charged groups). To circumvent this drawback, the combination of different similarity indices was successfully applied (similarity fusion). An alternative strategy is the combination of several reference ligands as initial model for similarity screenings (group fusion) [19,50,51]. This method provides satisfactory predictions and is generally recommended for nearest-neighbor searching when numerous known active ligands are available [52]. For both approaches, it could be shown that they were at least as effective as the best individual similarity searches, and that the combination of fingerprints or multiple reference ligands could reduce substantial variations as compared to conventional approaches of similarity-based screening.

Among the various existing similarity metrics, the Tanimoto coefficient (T_c) has been established as the gold standard [53],

$$T_c = c(a + b - c)^{-1}$$

where a, b, and c are the number of bits: in the fingerprints of molecule A only, in the fingerprints of molecule B only, and common to the fingerprints of both molecules, respectively. T_c values range from 0 (complete dissimilarity) to 1 (identity). The higher the structural similarity between two molecules, the higher the probability that they might have similar activities for a given target [54,55]. By virtue of its simplicity and speed, nearest-neighbor searching is incorporated in almost every drug design software package, as well as in online chemical databases. Different methods for encoding fingerprints, such as ECFP (circular-based), FP2 (topological-based), and MACCS (substructure-based), are in use. Several web servers for ligand-based target fishing exist, such as SwissSimilarity [56], SuperPred [57], TargetHunter [58], HybridSim-VS [59], PASS [60], SEA search server [61], and USR-VS [62]. Xu *et al.* [63] identified muscarinic acetylcholine receptor 2, cannabinoid receptor 1, cannabinoid receptor 2, and dopamine receptor 2 with TargetHunter, as potential targets for salvinorin A, the major component of the Mexican plant *Salvia divinorum* and a potent hallucinogen. These targets were validated by means of both *in vitro* and *in vivo* assays. Zatelli *et al.* [64] employed the similarity ensemble approach (SEA) to rationalize the anti-inflammatory effect of miconidin acetate (major metabolite of the Brazilian plant *Eugenia hiemalis*), whereby it was compared to annotated similar molecule ensembles for a given target from the ChEMBL16 binding database. The inflammation related protein 5-lipoxygenase, was the most promising predicted target and its inhibition by miconidin acetate was validated in cell-based assays (Fig. 4).

2.1.3. Quantitative structure—activity relationship (QSAR)

Since its origin in the 1962 seminal paper of Hansch *et al.* [65], quantitative structure—activity relationship (QSAR) has been one of the main computational methods applied in medicinal chemistry [66]. QSAR attempts to build mathematical models which quantitatively correlate structural properties of substances and their biological activities using statistical analysis such as multiple

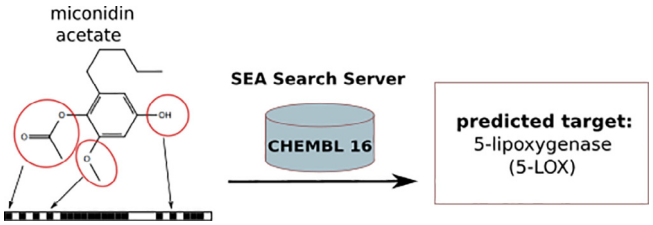

Fig. 4. Target fishing of miconidin acetate with the SEA Search sever.

linear regression (MLR), partial least-squares (PLS), k-nearest neighbors (kNN), etc [67]. QSAR models can be used to optimize existing leads or to predict DTIs for new compounds. As previously mentioned, the fundamental idea underlying QSAR modeling is that compounds sharing structural similarity should also share similar biological activity [18]. Based on the descriptors representing properties of (or differences between) compounds, QSAR methods can be classified into classical QSAR (2D-QSAR), 3D-QSAR, and higher dimensionalities (4D-7D QSAR) [68,69].

Classical QSAR correlates activity with 2D-structural patterns and physicochemical properties of drugs such as p*Ka*, logP, molecular weight, and polarizability [70]. However, the specific DTI depends on a shape complementarity between the ligand and the ligand-binding pocket in the 3D arrangement. It is not surprising that classical QSAR, considering neither the conformation nor the chirality of drugs, suffers from limitations. As a natural extension of classical QSAR, 3D-QSAR emerged for correlating steric and electrostatic potential interaction energies with biological activities, with CoMFA (comparative molecular field analysis) as the first successful demonstration [71]. The contour maps from CoMFA show key features and deeper insight into the mechanism of DTIs, which make it a powerful 3D QSAR method applied successfully in many cases. CoMSIA integrates electrostatic, steric, hydrophobic, hydrogen bond donor and acceptor effects [72]. However, in CoMFA analysis a mutual alignment of all 'bioactive' conformations of compounds is needed, which constitutes one of the most time-consuming aspects of alignment-dependent 3D-QSAR [73]. Thus, alignment-independent 3D QSAR methods have been developed such as COMPASS [74], CoMMA [75], HQSAR [76], and GRIND [77]. An advanced software tool implementing GRIND is Pentacle from Molecular Discovery [78]. The Schrodinger software suite offers AutoQSAR for 3D-QSAR modeling [79]. In order to refine ligand-based 3D QSAR models, receptor-based 3D-QSAR emerged, including COMBINE [80] and AFMoC [81].

QSAR techniques consider the interaction of a group of compounds with only one single target. When trained on these compounds, a QSAR model mostly has limited ability to extrapolate into novel areas of chemical space (to identify new classes of ligands or new binding modes of similar compounds outside the training data). In order to build a statistically meaningful model, QSAR requires enough data on a specific target, which is rarely the case when predicting DTIs for a newly identified target [82]. However, it could be shown that QSAR methods can be successfully applied to identify natural products and related derivatives as inhibitors for various targets, such as monoamine oxidase (MAO). In this study, Helguera *et al.* [83] combined 0D, 1D and 2D molecular descriptors including pure topological descriptors, connectivity indices, walk and path counts, information indices, or 2D-autocorrelations. Linear discriminant analysis (LDA) for modeling, replacement method (RM) for feature selection and Y-randomization test to ensure model robustness, were applied for generating structurally diverse and statistically meaningful QSAR models (Fig. 5). The combinatorial QSAR approach allowed derivation of chemical features which are important for the hMAO-B selectivity.

2.2. Target-based approaches

Molecular docking and the aforementioned receptor-centric pharmacophore modeling are the two existing computational approaches for target-based (structure-based) DTI prediction, and are generally used in conjunction. Central to these methods is the 3D structure of the target protein, determined experimentally by X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, or cryo-electron microscopy (cryo-EM) [84–86]. Alternatively, comparative (homology) modeling can be used to predict an unknown protein structure, based on the solved 3D structure of a template protein sharing high sequence similarity with the protein of interest [87].

2.2.1. Molecular docking

Docking predicts the binding mode (pose) of a ligand towards a target protein's binding site forming a stable (non-)covalent complex, by evaluating and ranking the predicted binding affinities of various poses. During the pose identification phase of a docking simulation, the flexibility of the ligand is accounted as part of the molecular recognition process, whereas that of the protein is normally neglected (rigid receptor docking) [84]. Three types of scoring functions have traditionally been used to measure the binding affinities of the docking poses, namely: force fields, empirical, and knowledge-based scoring functions. Their inability to correctly rank the binding poses, partially due to the unaccounted solvation effect and protein flexibility, impede on their predictive reliability [88–91]. Consensus scoring, involving the combination of two or more scoring functions, has been shown to produce more reliable ranking of docking poses [92,93]. Also, machine learning scoring functions based on protein—ligand interactions data available in chemical databases, have emerged as promising surrogates of the classical scoring functions [94–96]. Furthermore, the binding affinities of top-ranked docking poses can be more accurately predicted via end-point free energy calculations such as molecular mechanics Poisson-Boltzmann or generalized Born surface area (MM/PBSA and MM/GBSA), combined with molecular dynamics (MD) simulations [97–99]. It is worth mentioning that, while induced-fit docking considers both ligand and protein flexibility, its high computational cost greatly penalises the number of evaluated ligands and docking poses [100].

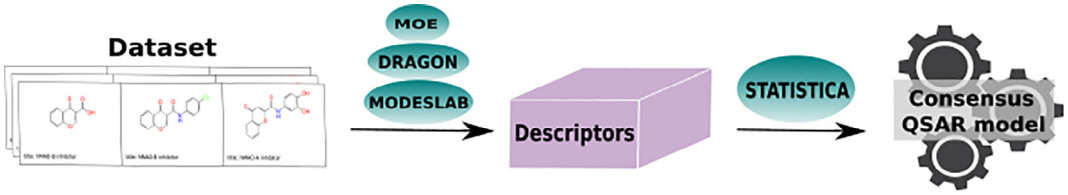

Fig. 5. QSAR modeling workflow. Different sets of descriptors were generated with MOE, DRAGON, and MODESLAB software. LDA and RM are implemented in the STATISTICA software.

The on- and off-target effects of several clinically approved drugs have been successfully predicted with the help of docking programs such as Gold [101], Glide [102], FlexX [103], Autodock [104], and DOCK [105], or web servers such as TarFisDock [106], INVDOCK [107] and idTarget [108] among others. Recently, Yang *et al.* [109] performed docking studies with the program Glide [102] to elucidate the stereoselective complementarity of (20S)-ginsenoside Rh2 over its 20R-epimer (constituents of ginseng), to the platelet P2Y12 receptor, which could be explained by their simulated binding modes, displaying disparate hydrogen bonding interactions with key residues such Asp266, Tyr105 and Glu188. In a view to rationalise the anti-tumor activity of epigallocatechin-3-gallate (EGCG), the major component of green tea, Wang *et al.* [110] constructed a dataset of tumor-related proteins and performed a reverse docking using the program Autodock Vina [111]. The authors established that EGCG anti-tumor mechanism may implicate 33 proteins (4 of which were previously unreported) via 12 signaling transduction pathways (Fig. 6). The inhibition of the 4 unreported proteins by EGCG was confirmed by means of *in vitro* enzymatic activity assay.

2.3. Target—ligand-based approaches

As an extension of QSAR (ligand-based), computational chemogenomic approaches and proteochemometric modeling (PCM) constitute the two computational approaches for target—ligand-based (hybrid) DTI prediction, which integrate both the chemical information of the compounds as well as the genomic space of target proteins in a single machine learning model. In chemogenomics, active compounds are applied as chemical probes to characterize the function of a specific protein. The modulation of the protein by the active compound induces a specific phenotype. If the phenotype can be related to a therapeutic mechanism, the protein comes into question as a drug target (reverse chemogenomics). If a molecule induces a specific phenotype but the target is not yet known, the main challenge lies in the development of methods for target identification (forward chemogenomics) [112].

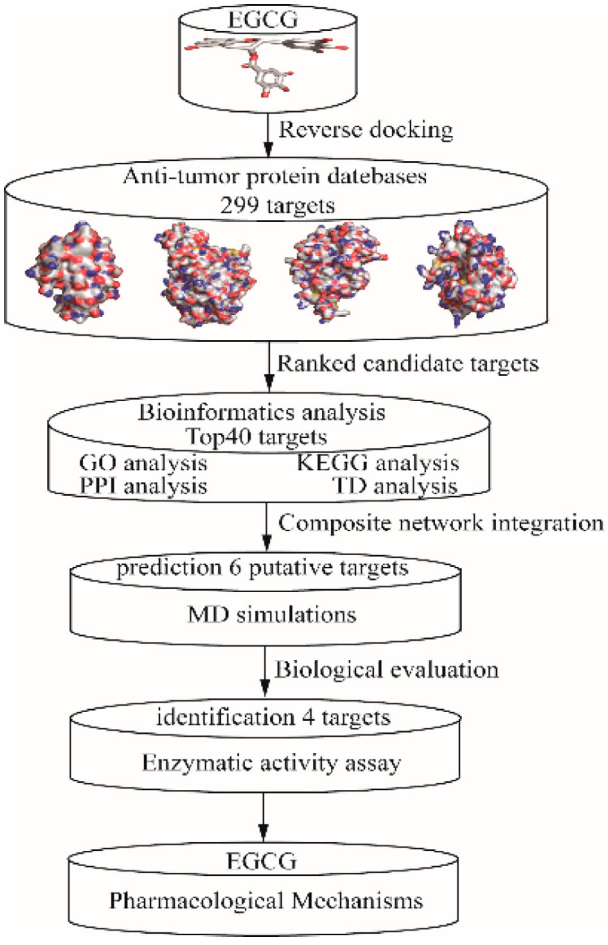

Fig. 6. Workflow of EGCG anti-tumour mechanism prediction, starting from reverse docking [110]. The original figure was published under a Creative Commons License.

2.3.1. Chemogenomic machine—learning approaches

With increasing knowledge about DTIs, machine learning (ML) methods are becoming increasingly popular and can extend and complement classical rule-based approaches such as network- and graph-based methods [113,114]. These ML methods for prediction of drug targets are normally supervised or semi-supervised, which requires a set of input variables or feature vectors (such as chemical fingerprints or physicochemical properties) and protein descriptors (such as amino acid composition, dipeptide composition, sequence order, etc.). The supervised ML algorithms for DTI predictions are trained on datasets that include labeled data containing information about the type of interaction and thus guide the algorithm to learn which features are important for DTIs. Consequently, known DTIs are a valuable resource for the development of ML prediction methods. For example, the latest release of DrugBank includes DTIs of about 12,000 drug entries including 2500 approved small molecule drugs and nearly 6000 experimental drugs [115]. Databases such as ChEMBL [116], PubChem Bioassay [117], and BindingDB [118] provide information about thousands of experimentally validated drug—target data pairs.

The majority of similarity-based ML are based on the guilt-by-association (GBA) principle, which states that similar proteins may be targeted by the same drug or vice-versa [119]. Although it cannot be generalized, genes with related functions often share common properties or physical interactions in gene networks [120]. Traditionally, the nearest profile method (NN) and the weighted profile method were widely utilized to predict new drugs or targets using chemical and interaction information about known compounds and targets [121,122]. In recent years, several new and optimized similarity-based methods have been published. Rodrigues *et al.* developed a random forest regression based DTI prediction workflow named DEcRyPT (Drug–Target Relationship Predictor) and it was successfully used to identify β-lapachone as an allosteric modulator of 5-lipoxygenase [123]. Semi-supervised machine learning algorithms, on the other hand, are trained on a combination of labeled and unlabeled data. Xia *et al.* utilized a manifold regularization semi-supervised learning method for predicting the DTIs from heterogeneous biological data sources [124] Schneider and co-workers developed SPiDER (self-organizing map-based prediction of drug equivalence relationships) utilizing the concept of unsupervised self-organizing map (SOM) algorithm applied in combination with pharmacophore feature representations for macromolecular target prediction. This software tool has been utilized in de-orphaning several natural products [125,126]. In a further development TIGER (Target Inference GEneratoR) was created, which utilizes a combination of multiple SOMs and was validated for the target prediction of numerous natural products [127,128].

2.3.2. Proteochemometric modeling

In contrast to chemogenomic machine—learning methods, proteochemometric modeling (PCM) allows both inter- and extrapolation to (novel) compounds and (novel) targets and can fulfill the need in hit identification of orphan targets [129–131]. PCM modeling requires three essential elements: descriptors

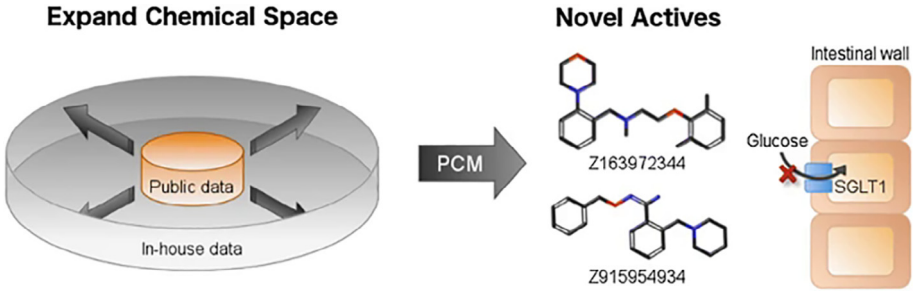

Fig. 7. Application of PCM to identify inhibitors of SGLT1 [147]. The original figure was published under a Creative Commons License.

(including target descriptors, ligand descriptors and additional cross-term descriptors describing information on ligand-target interaction), bioactivity data as well as appropriate modeling techniques linking the descriptors to the activity data. Ligand descriptors used in PCM include binary descriptors, physicochemical descriptors, 2D topological descriptors, 2D circular fingerprints and alignment based 3D descriptors. Physicochemical numerical (real-valued) descriptors are better interpretable than binary descriptors [132]. 3D descriptors require alignments of compounds in their active conformation in 3D space, which is error prone and may introduce noise into the data [133].

As compared to ligands, protein targets are in general larger and need also other descriptor sets. A reduction to a selection of residues (e.g. the binding sites) depends on the availability of related crystal structures. Information derived from sequence can be used to calculate similarity between various entities, such as binding pockets, physicochemical properties, topological properties, or 3D electrostatic potentials [134,135]. Protein descriptors can be also generated based on the availability of specific residues, substructures, or domains. It was shown that a related feature-based semi-binary protein descriptors could outperform sequential descriptors [136]. Cross-term descriptors derived from the multiplication of ligand and protein descriptors (MLPD) were used in early PCM modeling research [137–140]. Although it can describe the two entities simultaneously, its significance is not easy to evaluate [141]. Later, cross-terms not generated by multiplication were developed. A new type of cross-term descriptors introduced in PCM is protein—ligand interaction fingerprint (PLIF), which has been shown that it can outperform the MLPD-based descriptors [142]. Machine learning and data processing techniques implemented in PCM include support vector machines (SVM), random forest (RF), gaussian processes (GP), principal component analysis (PCA) [143,144].

Since PCM considers related targets in addition to multiple ligands, it is able to quantify the similarity between different binding sites, such as the subpockets of a given protein target. PCM can provide advantages in identification for novel allosteric inhibitors, which show advantages in treatment by not disrupting essential physiological process completely [145]. Similarly, considering the induced-fit interaction between drugs and targets, PCM allows distinction between different protein conformations and binding modes. When these related targets refer to similar targets from different species, PCM modeling is able to extrapolate bioactivity data between species and provide intra-species selectivity [146]. Burggraaff *et al.* [147] recently applied PCM in identification of inhibitors for sodium-dependent glucose co-transporter 1 (SGLT1), by implementation of ligand- and protein-based information into random forest models. The authors used an in-house collection of natural products and synthetic compounds. 30 out of 77 identified compounds were validated *in vitro,* showing submicromolar activities (Fig. 7).

3. Summary and outlook

This review presents the current advances and challenges of the state-of-the-art approaches in tackling DTI prediction in small molecule drug discovery from a computational point of view, with a special focus on NPs, which have been and will continue to be an indispensable source of drugs. Although, the rate of approved new molecular entities (NMEs) of NP origin has recently dropped, there is still a largely untapped reservoir of hitherto NPs that could fill the gap.

Computational DTI prediction speeds up as well as reduce the cost of the rather expensive drug discovery and development process. The various *in silico* approaches for DTI prediction have their specific field of applicability. The method of choice in each drug discovery campaign will depend on the type of target protein under consideration, the availability of the protein's macromolecular structure, the number of known active ligands and the availability of annotated DTIs in databases. The main caveat of ligand-based pharmacophore screening and similarity searching is the decrease in their predictive reliability when there is a low number of (or zero) known active ligands for a target of interest. In addition, there exist activity cliffs: molecules with high structural similarity but dissimilar biological activities for the same target. Regarding target-based approaches, the absence of the 3D macromolecular structure of the target protein, the lack of good scoring functions and the high computational costs, are the main drawbacks. As for ligand—ltarget-based approaches which mostly rely on machine learning algorithms, the quality of the curated drug-target annotations stored in chemogenomic databases is a matter of great concern. Also, there is a risk of chance correlation or overfitting because of the large number of descriptors. The hierarchical combination of several DTI prediction approaches has shown to provide superior predictions as opposed to the use of a single approach. These computational methods are still to reveal their full potential, where the completion of the Human Genome Project (HGP), improvements in cryo-EM for protein macromolecular structure determination and dynamics, advances in scoring algorithms and computing power, could be potential game changers.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

AFAM was supported by a doctoral research grant from the German Academic Exchange Service [DAAD, Award No. 91653768]. JL was supported by the German National Research Foundation [DFG,

Research Training Group 1976] and by the Baden-Württemberg Foundation [BWST_WSF-043].

References

[1] Moumbock AFA, Simoben CV, Wessjohann L, Sippl W, Günther S, Ntie-Kang F. Computational studies and biosynthesis of natural products with promising anticancer properties. Nat Prod Cancer Drug Discov, InTech 2017. https://doi.org/10.5772/67650.

[2] Walsh CT, Fischbach MA. Natural products version 2.0: connecting genes to molecules. J Am Chem Soc 2010;132:2469–93. https://doi.org/10.1021/ja909118a.

[3] Newman DJ, Cragg GM. Natural products as sources of new drugs from 1981 to 2014. J Nat Prod 2016;79:629–61. https://doi.org/10.1021/acs.jnatprod.5b01055.

[4] Patridge E, Gareiss P, Kinch MS, Hoyer D. An analysis of FDA-approved drugs: natural products and their derivatives. Drug Discov Today 2016;21:204–7. https://doi.org/10.1016/j.drudis.2015.01.009.

[5] Li F, Wang Y, Li D, Chen Y, Dou QP. Are we seeing a resurgence in the use of natural products for new drug discovery? Expert Opin Drug Discov 2019:1–4. https://doi.org/10.1080/17460441.2019.1582639.

[6] All Nobel Prizes in Physiology or Medicine. https://www.nobelprize.org/prizes/lists/all-nobel-laureates-in-physiology-or-medicine/ (accessed March 8, 2019).

[7] Klementz D, Döring K, Lucas X, Telukunta KK, Erxleben A, Deubel D, et al. StreptomeDB 2.0—an extended resource of natural products produced by streptomycetes. Nucleic Acids Res 2016;44:D509–14. https://doi.org/10.1093/nar/gkv1319.

[8] Ntie-Kang F, Telukunta KK, Döring K, Simoben CVA, Moumbock AF, Malange YI, et al. NANPDB: a resource for natural products from Northern African sources. J Nat Prod 2017;80:2067–76. https://doi.org/10.1021/acs.jnatprod.7b00283.

[9] Cabantchik ZI, Drakesmith H. From one Nobel Prize (P. Ehrlich) to another (Tu Youyou): 100 years of chemotherapy of infectious diseases. Clin Microbiol Infect 2016;22:213–4. https://doi.org/10.1016/j.cmi.2015.11.011.

[10] Strebhardt K, Ullrich A. Paul Ehrlich's magic bullet concept: 100 years of progress. Nat Rev Cancer 2008;8:473–80. https://doi.org/10.1038/nrc2394.

[11] Proschak E, Stark H, Merk D. Polypharmacology by design: a medicinal chemist's perspective on multitargeting compounds. J Med Chem 2019;62:420–44. https://doi.org/10.1021/acs.jmedchem.8b00760.

[12] Anighoro A, Bajorath J, Rastelli G. Polypharmacology: challenges and opportunities in drug discovery. J Med Chem 2014;57:7874–87. https://doi.org/10.1021/jm5006463.

[13] Medina-Franco JL, Giulianotti MA, Welmaker GS, Houghten RA. Shifting from the single to the multitarget paradigm in drug discovery. Drug Discov Today 2013;18:495–501. https://doi.org/10.1016/j.drudis.2013.01.008.

[14] Lavecchia A, Cerchia C. In silico methods to address polypharmacology: current status, applications and future perspectives. Drug Discov Today 2016;21:288–98. https://doi.org/10.1016/j.drudis.2015.12.007.

[15] Patel H, Lucas X, Bendik I, Günther S, Merfort I. Target fishing by cross-docking to explain polypharmacological effects. ChemMedChem 2015;10:1209–17. https://doi.org/10.1002/cmdc.201500123.

[16] Huang H, Zhang G, Zhou Y, Lin C, Chen S, Lin Y, et al. Reverse screening methods to search for the protein targets of chemopreventive compounds. Front Chem 2018;6:138. https://doi.org/10.3389/fchem.2018.00138.

[17] Chaudhari R, Tan Z, Huang B, Zhang S. Computational polypharmacology: a new paradigm for drug discovery. Expert Opin Drug Discov 2017;12:279–91. https://doi.org/10.1080/17460441.2017.1280024.

[18] Bender A, Glen RC. Molecular similarity: a key technique in molecular informatics. Org Biomol Chem 2004;2:3204–18. https://doi.org/10.1039/B409813G.

[19] Maggiora G, Vogt M, Stumpfe D, Bajorath J. Molecular similarity in medicinal chemistry. J Med Chem 2014;57:3186–204. https://doi.org/10.1021/jm401411z.

[20] Ehrlich P. Über den jetzigen Stand der Chemotherapie. Berichte Der Dtsch Chem Gesellschaft 1908:17–47. https://doi.org/10.1002/cber.19090420105.

[21] Langer T, Wolber G. Pharmacophore definition and 3D searches. Drug Discov Today Technol 2004;1:203–7. https://doi.org/10.1016/J.DDTEC.2004.11.015.

[22] Wermuth CG, Ganellin CR, Lindberg P, Mitscher LA. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). Pure Appl Chem 1998;70:1129–43. https://doi.org/10.1351/pac199870051129.

[23] Steindl TM, Schuster D, Laggner C, Langer T. Parallel screening: a novel concept in pharmacophore modeling and virtual screening. J Chem Inf Model 2006;46:2146–57. https://doi.org/10.1021/ci6002043.

[24] Podolyan Y, Karypis G. Common pharmacophore identification using frequent clique detection algorithm. J Chem Inf Model 2009;49:13–21. https://doi.org/10.1021/ci8002478.

[25] Van Drie JH. History of 3D pharmacophore searching: commercial, academic and open-source tools. Drug Discov Today Technol 2010;7:e255–62. https://doi.org/10.1016/j.ddtec.2010.12.002.

[26] Hu G, Kuang G, Xiao W, Li W, Liu G, Tang Y. Performance evaluation of 2D fingerprint and 3D shape similarity methods in virtual screening. J Chem Inf Model 2012;52:1103–13. https://doi.org/10.1021/ci300030u.

[27] Seidel T, Ibis G, Bendix F. Strategies for 3D pharmacophore-based virtual screening. Drug Discov Today Technol 2010;7:e221–8. https://doi.org/10.1016/J.DDTEC.2010.11.004.

[28] Sanders MPA, Barbosa AJM, Zarzycka B, Nicolaes GAF, Klomp JPG, de Vlieg J, et al. Comparative analysis of pharmacophore screening tools. J Chem Inf Model 2012;52:1607–20. https://doi.org/10.1021/ci2005274.

[29] Koes DR, Camacho CJ. Pharmer: efficient and exact pharmacophore search. J Chem Inf Model 2011;51:1307–14. https://doi.org/10.1021/ci200097m.

[30] BIOVA Discovery Studio n.d. http://www.3dsbiovia.com/products/collaborative-science/biovia-discovery-studio/.

[31] Wolber G, Langer T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J Chem Inf Model 2005;45:160–9. https://doi.org/10.1021/ci049885e.

[32] Dixon SL, Smondyrev AM, Rao SN. PHASE: a novel approach to pharmacophore modeling and 3D database searching. Chem Biol Drug Des 2006;67:370–2. https://doi.org/10.1111/j.1747-0285.2006.00384.x.

[33] ChemAxon Screen Suite. https://chemaxon.com/products/screen-suite.

[34] Molecular Operating Environment (MOE). Chemical Computing Group. https://www.chemcomp.com/.

[35] Koes DR, Camacho CJ. ZINCPharmer: pharmacophore search of the ZINC database. Nucleic Acids Res 2012;40:W409–14. https://doi.org/10.1093/nar/gks378.

[36] Wang X, Shen Y, Wang S, Li S, Zhang W, Liu X, et al. PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. Nucleic Acids Res 2017;45:W356–60. https://doi.org/10.1093/nar/gkx374.

[37] Sunseri J, Koes DR. Pharmit: interactive exploration of chemical space. Nucleic Acids Res 2016;44:W442–8. https://doi.org/10.1093/nar/gkw287.

[38] Xu Y, Wang S, Hu Q, Gao S, Ma X, Zhang W, et al. CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. Nucleic Acids Res 2018;46:W374–9. https://doi.org/10.1093/nar/gky380.

[39] Kirchweger B, Kratz JM, Ladurner A, Grienke U, Langer T, Dirsch VM, et al. In silico workflow for the discovery of natural products activating the G protein-coupled bile acid receptor 1. Front Chem 2018;6:1–14. https://doi.org/10.3389/fchem.2018.00242.

[40] Dong X, Ebalunode JO, Yang S-Y, Zheng W. Receptor-based pharmacophore and pharmacophore key descriptors for virtual screening and QSAR modeling. Curr Comput Aided Drug Des 2011;7:181–9.

[41] Loving K, Salam NK, Sherman W. Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. J Comput Aided Mol Des 2009;23:541–54. https://doi.org/10.1007/s10822-009-9268-1.

[42] Meslamani J, Li J, Sutter J, Stevens A, Bertrand H-O, Rognan D. Protein−ligand-based pharmacophores: generation and utility assessment in computational ligand profiling. J Chem Inf Model 2012;52(4):943–55. https://doi.org/10.1021/ci300083r.

[43] Inte:PharmacophoreDB. http://www.inteligand.com/pharmdb/.

[44] Rollinger JM, Schuster D, Danzl B, Schwaiger S, Markt P, Schmidtke M, et al. In silico target fishing for rationalized ligand discovery exemplified on constituents of Ruta graveolens. Planta Med 2009;75:195–204. https://doi.org/10.1055/s-0028-1088397.

[45] Willett P, Winterman V, Bawden D. Implementation of nearest-neighbor searching in an online chemical structure search system. J Chem Inf Model 1986;26:36–41. https://doi.org/10.1021/ci00049a008.

[46] Sheridan RP, Miller MD, Underwood DJ, Kearsley SK. Chemical similarity using geometric atom pair descriptors. J Chem Inf Comput Sci 1996;36:128–36. https://doi.org/10.1021/ci950275b.

[47] Riniker S, Landrum GA. Open-source platform to benchmark fingerprints for ligand-based virtual screening. J Cheminform 2013;5:26. https://doi.org/10.1186/1758-2946-5-26.

[48] O'Boyle NM, Sayle RA. Comparing structural fingerprints using a literature-based similarity benchmark. J Cheminform 2016;8:36. https://doi.org/10.1186/s13321-016-0148-0.

[49] Cereto-Massagué A, Ojeda MJ, Valls C, Mulero M, Garcia-Vallvé S, Pujadas G. Molecular fingerprint similarity search in virtual screening. Methods 2015;71:58–63. https://doi.org/10.1016/J.YMETH.2014.08.005.

[50] Flower DR. On the properties of bit string-based measures of chemical similarity. J Chem Inf Model 1998;38:379–86. https://doi.org/10.1021/CI9704377.

[51] Fligner MA, Verducci JS, Blower PE. A modification of the jaccard-tanimoto similarity index for diverse selection of chemical compounds using binary strings. Technometrics 2002;44:110–9. https://doi.org/10.1198/004017002317375064.

[52] Willett P. Fusing similarity rankings in ligand-based virtual screening. Comput Struct Biotechnol J 2013;5:. https://doi.org/10.5936/csbj.201302002e201302002.

[53] Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? J Cheminform 2015;7:20. https://doi.org/10.1186/s13321-015-0069-3.

[54] Jasial S, Hu Y, Vogt M, Bajorath J. Activity-relevant similarity values for fingerprints and implications for similarity searching. F1000Research 2016;5. https://doi.org/10.12688/f1000research.8357.2.

[55] Martin YC, Kofron JL, Traphagen LM. Do structurally similar molecules have similar biological activity? J Med Chem 2002;45:4350–8. https://doi.org/10.1021/JM020155C.

[56] Zoete V, Daina A, Bovigny C, Michielin O. SwissSimilarity: a web tool for low to ultra high throughput ligand-based virtual screening. J Chem Inf Model 2016;56:1399–404. https://doi.org/10.1021/acs.jcim.6b00174.

[57] Dunkel M, Günther S, Ahmed J, Wittig B, Preissner R. SuperPred: drug classification and target prediction. Nucleic Acids Res 2008;36:W55–9. https://doi.org/10.1093/nar/gkn307.

[58] Wang L, Ma C, Wipf P, Liu H, Su W, Xie X-Q. TargetHunter: an in silico target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database. AAPS J 2013;15:395–406. https://doi.org/10.1208/s12248-012-9449-z.

[59] Shang J, Dai X, Li Y, Pistolozzi M, Wang L. HybridSim-VS: a web server for large-scale ligand-based virtual screening using hybrid similarity recognition techniques. Bioinformatics 2017;33:3480–1. https://doi.org/10.1093/bioinformatics/btx418.

[60] Lagunin A, Filimonov D, Poroikov V. Multi-targeted natural products evaluation based on biological activity prediction with PASS. Curr Pharm Des 2010;16:1703–17. https://doi.org/10.2174/138161210791164063.

[61] Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein pharmacology by ligand chemistry. Nat Biotechnol 2007;25:197–206. https://doi.org/10.1038/nbt1284.

[62] Li H, Leung K-S, Wong M-H, Ballester PJ. USR-VS: a web server for large-scale prospective virtual screening using ultrafast shape recognition techniques. Nucleic Acids Res 2016;44:W436–41. https://doi.org/10.1093/nar/gkw320.

[63] Xu X, Ma S, Feng Z, Hu G, Wang L, Xie X-Q. Chemogenomics knowledgebase and systems pharmacology for hallucinogen target identification—Salvinorin A as a case study. J Mol Graph Model 2016;70:284–95. https://doi.org/10.1016/J.JMGM.2016.08.001.

[64] Zatelli G, Temml V, Kutil Z, Landa P, Vanek T, Schuster D, et al. Miconidin acetate and primin as potent 5-lipoxygenase inhibitors from brazilian eugenia hiemalis (Myrtaceae). Planta Medica Lett 2016;3:e17–9. https://doi.org/10.1055/s-0042-102460.

[65] Hansch C, Muir RM, Fujita T, Maloney PP, Geiger F, Streich M. The correlation of biological activity of plant growth regulators and chloromycetin derivatives with hammett constants and partition coefficients. J Am Chem Soc 1963;85:2817–24. https://doi.org/10.1021/ja00901a033.

[66] Singh DA, Singh DR. QSAR and its role in target-ligand interaction. Open Bioinforma J 2013;7:63–7. https://doi.org/10.2174/1875036201307010063.

[67] Lo Y-C, Rensi SE, Torng W, Altman RB. Machine learning in chemoinformatics and drug discovery. Drug Discov Today 2018;23:1538–46. https://doi.org/10.1016/j.drudis.2018.05.010.

[68] Polanski J. Receptor dependent multidimensional QSAR for modeling drug–receptor interactions. Curr Med Chem 2009;16:3243–57.

[69] Lill MA. Multi-dimensional QSAR in drug discovery. Drug Discov Today 2007;12:1013–7. https://doi.org/10.1016/j.drudis.2007.08.004.

[70] Baskin II, Consonni V, Muratov EN, Todeschini R, Rathman J, Varnek A, et al. QSAR Modeling: where have you been? Where are you going to? J Med Chem 2013;57:4977–5010. https://doi.org/10.1021/jm4004285.

[71] Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 1988;110:5959–67. https://doi.org/10.1021/ja00226a005.

[72] Klebe G, Abraham U, Mietzner T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem 1994;37:4130–46. https://doi.org/10.1021/jm00050a010.

[73] Kim KH. Comparative molecular field analysis (CoMFA). Mol. Similarity Drug Des. Weinheim, Germany: Wiley-VCH Verlag GmbH; 2011. p. 291–331.

[74] Jain AN, Koile K, Chapman D. Compass: predicting biological activities from molecular surface properties. Performance comparisons on a steroid benchmark. J Med Chem 1994;37:2315–27. https://doi.org/10.1021/jm00041a010.

[75] Silverman BD, Platt DE. Comparative molecular moment analysis (coMMA): 3D-QSAR without molecular superposition. J Med Chem 1996;39:2129–40. https://doi.org/10.1021/jm950589q.

[76] Heritage TW, Lowis DR. Molecular hologram QSAR. Ration Drug Des 2009:212–25. https://doi.org/10.1021/bk-1999-0719.ch014.

[77] Pastor M, Cruciani G, McLay I, Pickett S, Clementi S. GRid-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. J Med Chem 2000;43:3233–43. https://doi.org/10.1021/jm000941m.

[78] Durán Á, Zamora I, Pastor M. Suitability of GRIND-based principal properties for the description of molecular similarity and ligand-based virtual screening. J Chem Inf Model 2009;49:2129–38. https://doi.org/10.1021/ci900228x.

[79] Dixon SL, Duan J, Smith E, Von Bargen CD, Sherman W, Repasky MP. AutoQSAR: an automated machine learning tool for best-practice quantitative structure-activity relationship modeling. Future Med Chem 2016;8:1825–39. https://doi.org/10.4155/fmc-2016-0093.

[80] Ortiz AR, Pisabarro MT, Gago F, Wade RC. Prediction of drug binding affinities by comparative binding energy analysis. J Med Chem 1995;38:2681–91. https://doi.org/10.1021/jm00014a020.

[81] Gohlke H, Klebe G. Drugscore meets CoMFA: adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. J Med Chem 2002;45:4153–70. https://doi.org/10.1021/jm020808p.

[82] Koutsoukas A, Simms B, Kirchmair J, Bond PJ, Whitmore AV, Zimmer S, et al. From in silico target prediction to multi-target drug design: current databases, methods and applications. J Proteomics 2011;74:2554–74. https://doi.org/10.1016/j.jprot.2011.05.011.

[83] Helguera AM, Pérez-Garrido A, Gaspar A, Reis J, Cagide F, Vina D, et al. Combining QSAR classification models for predictive modeling of human monoamine oxidase inhibitors. Eur J Med Chem 2013;59:75–90. https://doi.org/10.1016/j.ejmech.2012.10.035.

[84] Gohlke H, Klebe G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew Chemie Int Ed 2002;41:2644–76. https://doi.org/10.1002/1521-3773(20020802)41:15<2644::AID-ANIE2644>3.0.CO;2-O.

[85] Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. Biophys Rev 2017;9:91–102. https://doi.org/10.1007/s12551-016-0247-1.

[86] Cheng Y, Grigorieff N, Penczek PA, Walz T. A primer to single-particle cryo-electron microscopy. Cell 2015;161:438–49. https://doi.org/10.1016/j.cell.2015.03.050.

[87] Muhammed MT, Aki-Yalcin E. Homology modeling in drug discovery: overview, current applications, and future perspectives. Chem Biol Drug Des 2019;93:12–20. https://doi.org/10.1111/cbdd.13388.

[88] Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, Lambert MH, et al. A critical assessment of docking programs and scoring functions. J Med Chem 2006;49:5912–31. https://doi.org/10.1021/jm050362n.

[89] Gilson MK, Zhou H-X. Calculation of protein-ligand binding affinities. Annu Rev Biophys Biomol Struct 2007;36:21–42. https://doi.org/10.1146/annurev.biophys.36.040306.132550.

[90] Liu Z, Su M, Han L, Liu J, Yang Q, Li Y, et al. Forging the basis for developing protein-ligand interaction scoring functions. Acc Chem Res 2017;50:302–9. https://doi.org/10.1021/acs.accounts.6b00491.

[91] Guedes IA, Pereira FSS, Dardenne LE. Empirical scoring functions for structure-based virtual screening: applications, critical aspects, and challenges. Front Pharmacol 2018;9:1089. https://doi.org/10.3389/fphar.2018.01089.

[92] Charifson PS, Corkery JJ, Murcko MA, Walters WP. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. J Med Chem 1999;42:5100–9. https://doi.org/10.1021/JM990352K.

[93] Ericksen SS, Wu H, Zhang H, Michael LA, Newton MA, Hoffmann FM, et al. Machine learning consensus scoring improves performance across targets in structure-based virtual screening. J Chem Inf Model 2017;57:1579–90. https://doi.org/10.1021/acs.jcim.7b00153.

[94] Khamis MA, Gomaa W, Ahmed WF. Machine learning in computational docking. Artif Intell Med 2015;63:135–52. https://doi.org/10.1016/j.artmed.2015.02.002.

[95] Wójcikowski M, Ballester PJ, Siedlecki P. Performance of machine-learning scoring functions in structure-based virtual screening. Sci Rep 2017;7:46710. https://doi.org/10.1038/srep46710.

[96] Ballester PJ, Mitchell JBO. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. Bioinformatics 2010;26:1169–75. https://doi.org/10.1093/bioinformatics/btq112.

[97] Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov 2015;10:449–61. https://doi.org/10.1517/17460441.2015.1032936.

[98] Homeyer N, Gohlke H. Free energy calculations by the molecular mechanics Poisson−Boltzmann surface area method. Mol Inform 2012;31:114–22. https://doi.org/10.1002/minf.201100135.

[99] Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. J Chem Inf Model 2011;51:69–82. https://doi.org/10.1021/ci100275a.

[100] Xu M, Lill MA. Induced fit docking, and the use of QM/MM methods in docking. Drug Discov Today Technol 2013;10:e411–8. https://doi.org/10.1016/j.ddtec.2013.02.003.

[101] Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking 1 1Edited by F. E. Cohen. J Mol Biol 1997;267:727–48. https://doi.org/10.1006/jmbi.1996.0897.

[102] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 2004;47:1739–49. https://doi.org/10.1021/jm0306430.

[103] Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. J Mol Biol 1996;261:470–89. https://doi.org/10.1006/jmbi.1996.0477.

[104] Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 2009;30:2785–91. https://doi.org/10.1002/jcc.21256.

[105] Ewing TJA, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des 2001;15:411–28. https://doi.org/10.1023/A: 1011115820450.

[106] Li H, Gao Z, Kang L, Zhang H, Yang K, Yu K, et al. TarFisDock: a web server for identifying drug targets with docking approach. Nucleic Acids Res 2006;34:W219–24. https://doi.org/10.1093/nar/gkl114.

[107] Chen YZ, Zhi DG. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. Proteins Struct Funct Genet 2001;43:217–26. https://doi.org/10.1002/1097-0134(20010501)43:2<217::AID-PROT1032>3.0.CO;2-G.

[108] Wang J-C, Chu P-Y, Chen C-M, Lin J-H. idTarget: a web server for identifying protein targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach. Nucleic Acids Res 2012;40: W393–9. https://doi.org/10.1093/nar/gks496.
[109] Yang Q, Wang N, Zhang J, Chen G, Xu H, Meng Q, et al. In vitro and in silico evaluation of stereoselective effect of ginsenoside isomers on platelet P2Y12 receptor. Phytomedicine 2019. https://doi.org/10.1016/J.PHYMED.2019.152899.
[110] Wang W, Xiong X, Li X, Zhang Q, Yang W, Du L. In silico investigation of the anti-tumor mechanisms of epigallocatechin-3-gallate. Molecules 2019;24:1–17. https://doi.org/10.3390/molecules24071445.
[111] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2010;31:455–61. https://doi.org/10.1002/jcc.21334.
[112] Wuster A, Madan Babu M. Chemogenomics and biotechnology. Trends Biotechnol 2008;26:252–8. https://doi.org/10.1016/j.tibtech.2008.01.004.
[113] Aittokallio T, Schwikowski B. Graph-based methods for analysing networks in cell biology. Brief Bioinform 2006;7:243–55. https://doi.org/10.1093/bib/bbl022.
[114] Wu Z, Li W, Liu G, Tang Y. Network-based methods for prediction of drug-target interactions. Front Pharmacol 2018;9:1134. https://doi.org/10.3389/fphar.2018.01134.
[115] Lo EJ, Iynkkaran I, Li C, Le D, Sajed T, Maciejewski A, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 2017;46: D1074–82. https://doi.org/10.1093/nar/gkx1037.
[116] Gaulton A, Hersey A, Nowotka ML, Patricia Bento A, Chambers J, Mendez D, et al. The ChEMBL database in 2017. Nucleic Acids Res 2017;45:D945–54. https://doi.org/10.1093/nar/gkw1074.
[117] Wang Y, Gindulyte A, Thiessen PA, Cheng T, He S, Wang J, et al. PubChem BioAssay: 2017 update. Nucleic Acids Res 2016;45:D955–63. https://doi.org/10.1093/nar/gkw1118.
[118] Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic Acids Res 2016;44:D1045–53. https://doi.org/10.1093/nar/gkv1072.
[119] Oliver S. Guilt-by-association goes global. Nature 2000;403:601–3. https://doi.org/10.1038/35001165.
[120] Gillis J, Pavlidis P. "Guilt by association" is the exception rather than the rule in gene networks. PLoS Comput Biol 2012;8:. https://doi.org/10.1371/journal.pcbi.1002444e1002444.
[121] van Laarhoven T, Marchiori E. Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile. PLoS ONE 2013;8:. https://doi.org/10.1371/journal.pone.0066952e66952.
[122] Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics 2008;24:i232–40. https://doi.org/10.1093/bioinformatics/btn162.
[123] Rodrigues T, Werner M, Roth J, da Cruz EHG, Marques MC, Akkapeddi P, et al. Machine intelligence decrypts β-lapachone as an allosteric 5-lipoxygenase inhibitor. Chem Sci 2018;9:6899–903. https://doi.org/10.1039/c8sc02634c.
[124] Xia Z, Wu LY, Zhou X, Wong STC. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. BMC Syst Biol 2010;4:S6. https://doi.org/10.1186/1752-0509-4-6.
[125] Kremer L, Schultz-Fademrecht C, Baumann M, Habenberger P, Choidas A, Klebl B, et al. Discovery of a novel inhibitor of the hedgehog signaling pathway through cell-based compound discovery and target prediction. Angew Chemie Int Ed 2017;56:13021–5. https://doi.org/10.1002/anie.201707394.
[126] Rodrigues T, Lin Y-C, Hartenfeller M, Renner S, Lim YF, Schneider G. Repurposing de novo designed entities reveals phosphodiesterase 3B and cathepsin L modulators. Chem Commun (Camb) 2015;51:7478–81. https://doi.org/10.1039/c5cc01376c.
[127] Schneider P, Schneider G. De-orphaning the marine natural product (±)-marinopyrrole A by computational target prediction and biochemical validation. Chem Commun 2017;53:2272–4. https://doi.org/10.1039/c6cc09693j.
[128] Grisoni F, Merk D, Friedrich L, Schneider G. Design of natural-product-inspired multitarget ligands by machine learning. ChemMedChem 2019;14:1129–34. https://doi.org/10.1002/cmdc.201900097.
[129] Ain QU, Méndez-Lucio O, Ciriano IC, Malliavin T, van Westen GJP, Bender A. Modelling ligand selectivity of serine proteases using integrative proteochemometric approaches improves model performance and allows the multi-target dependent interpretation of features. Integr Biol 2014;6:1023–33. https://doi.org/10.1039/C4IB00175C.
[130] Malliavin TE, van Westen GJP, Méndez-Lucio O, Lenselink EB, Prusis P, Wohlfahrt G, et al. Polypharmacology modelling using proteochemometrics (PCM): recent methodological developments, applications to target families, and future prospects. Medchemcomm 2014;6:24–50. https://doi.org/10.1039/c4md00216d.
[131] IJzerman AP, Paricharak S, Bender A, Cortés-Ciriano I, Malliavin TE. Proteochemometric modelling coupled to in silico target prediction: an integrated approach for the simultaneous prediction of polypharmacology and binding affinity/potency of small molecules. J Cheminform 2015;7:1–11. doi:10.1186/s13321-015-0063-9.
[132] Lapinsh M. Proteochemometric mapping of the interaction of organic compounds with melanocortin receptor subtypes. Mol Pharmacol 2004;67:50–9. https://doi.org/10.1124/mol.104.002857.
[133] Van Westen GJP, Wegner JK, Ijzerman AP, Van Vlijmen HWT, Bender A. Proteochemometric modeling as a tool to design selective compounds and for extrapolating to novel targets. Medchemcomm 2011;2:16–30. https://doi.org/10.1039/c0md00165a.
[134] van Westen GJ, Bender A, Swier RF, van Vlijmen HW, Wegner JK, IJzerman AP. Benchmarking of protein descriptor sets in proteochemometric modeling (part 1): comparative study of 13 amino acid descriptor sets. J Cheminform 2013;5:1. doi:10.1186/1758-2946-5-41.
[135] van Westen GJ, Swier RF, Cortes-Ciriano I, Wegner JK, Overington JP, IJzerman AP, et al. Benchmarking of protein descriptor sets in proteochemometric modeling (part 2): modeling performance of 13 amino acid descriptor sets. J Cheminform 2013;5:42. doi:10.1186/1758-2946-5-42.
[136] Doddareddy MR, van Westen GJP, van der Horst E, Peironcely JE, Corthals F, Ijzerman AP, et al. Chemogenomics: looking at biology through the lens of chemistry. Stat Anal Data Min 2009;2:149–60. https://doi.org/10.1002/sam.10046.
[137] Lapinsh M, Prusis P, Gutcaits A, Lundstedt T, Wikberg JES. Development of proteo-chemometrics: a novel technology for the analysis of drug-receptor interactions. Biochim Biophys Acta - Gen Subj 2001;1525:180–90. https://doi.org/10.1016/S0304-4165(00)00187-2.
[138] Lapinsh M. Proteochemometrics modeling of the interaction of amine G-protein coupled receptors with a diverse set of ligands. Mol Pharmacol 2003;61:1465–75. https://doi.org/10.1124/mol.61.6.1465.
[139] Lapinsh M, Prusis P, Mutule I, Mutulis F, Wikberg JES. QSAR and proteo-chemometric analysis of the interaction of a series of organic compounds with melanocortin receptor subtypes. J Med Chem 2003;46:2572–9. https://doi.org/10.1021/jm020945m.
[140] Freyhult E, Prusis P, Lapinsh M, Wikberg JES, Moulton V, Gustafsson MG. Unbiased descriptor and parameter selection confirms the potential of proteochemometric modelling. BMC Bioinf 2005;6:1–14. https://doi.org/10.1186/1471-2105-6-50.
[141] Huang Q, Jin H, Liu Q, Wu Q, Kang H, Cao Z, et al. Proteochemometric modeling of the bioactivity spectra of HIV-1 protease inhibitors by introducing protein-ligand interaction fingerprint. PLoS ONE 2012;7:1–8. https://doi.org/10.1371/journal.pone.0041698.
[142] Qiu T, Qiu J, Feng J, Wu D, Yang Y, Tang K, et al. The recent progress in proteochemometric modelling: focusing on target descriptors, cross-term descriptors and application scope. Brief Bioinform 2017;18:125–36. https://doi.org/10.1093/bib/bbw004.
[143] Ding H, Takigawa I, Mamitsuka H, Zhu S. Similarity-based machine learning methods for predicting drug-target interactions: a brief review. Brief Bioinform 2013;15:734–47. https://doi.org/10.1093/bib/bbt056.
[144] Atas H, Rifaioglu AS, Cetin-Atalay R, Atalay V, Doğan T, Martin MJ. Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. Brief Bioinform 2018:1–36. https://doi.org/10.1093/bib/bby061.
[145] Tresadern G, Trabanco AA, Pérez-Benito L, Overington JP, Van Vlijmen HWT, Van Westen GJP. Identification of allosteric modulators of metabotropic glutamate 7 receptor using proteochemometric modeling. J Chem Inf Model 2017;57:2976–85. https://doi.org/10.1021/acs.jcim.7b00338.
[146] van Westen GJP, Bender A, Overington JP. Towards predictive resistance models for agrochemicals by combining chemical and protein similarity via proteochemometric modelling. J Chem Biol 2014;7:119–23. https://doi.org/10.1007/s12154-014-0112-2.
[147] Burggraaff L, Oranje P, Gouka R, van der Pijl P, Geldof M, van Vlijmen HWT, et al. Identification of novel small molecule inhibitors for solute carrier SGLT1 using proteochemometric modeling. J Cheminform 2019;11:15. https://doi.org/10.1186/s13321-019-0337-8.