12

Virtual Screening and Bioactivity Modeling for G Protein-Coupled Receptors

Wallace Chan[1,2,3], Jiansheng Wu[1,4], Eric Bell[1], and Yang Zhang[1,2]

[1] *Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA*
[2] *Department of Biological Chemistry, University of Michigan, Ann Arbor, MI, USA*
[3] *Department of Pharmacology, University of Michigan, Ann Arbor, MI, USA*
[4] *School of Geographic and Biological Information, Nanjing University of Posts and Telecommunications, Nanjing, China*

12.1 Introduction

G protein-coupled receptors (GPCRs) are a superfamily of integral membrane proteins and consist of over 800 established members, establishing them as the third-largest family of proteins in humans [1, 2]. Marked by their distinctive seven-pass transmembrane domain, they account for almost 5% of the human proteome. Consequently, they have been implicated in a multitude of diseases, including cancer and diabetes [3, 4]. Moreover, almost a third of all drugs in use today target these receptors, accentuating their importance in drug discovery [5]. Given the interest the pharmaceutical industry has in GPCRs, extensive scientific efforts have been made to develop novel drugs for a variety of medical conditions.

Drug discovery has traditionally used high throughput screens (HTSs) as a means to discover hit compounds from enormous chemical libraries. Unfortunately, these large-scale assays are typically very expensive, time-consuming, and laborious. In the years following its explosive beginnings in the pharmaceutical industry in the early 1980s [6], computer-aided drug design (CADD) methods were developed to computationally predict how well a potential drug would bind to a receptor or to model how it binds; predictions effectively compensate for the brute force approach of HTS and help inform further biochemical experiments. In particular, virtual (or *in silico*) screening complements HTS by reducing the chemical space to be explored. Using various CADD approaches, computational chemists could then assign scores to chemical compounds and rank them accordingly, helping prioritize which compounds to experimentally assay.

GPCRs as Therapeutic Targets, Volume 1, First Edition. Edited by Annette Gilchrist.

In the current chapter, we aim to provide the reader with an introduction to virtual screening and its application to GPCRs. In addition to providing an overview of virtual screening and its required components, we will delve into what will be referred to as classical virtual screening; this includes many well-established approaches with which many medicinal chemists will be familiar, such as chemical similarity comparisons and molecular docking. Subsequently, we will survey the use of chemogenomics and machine learning in virtual screening, including bioactivity prediction. Lastly, various topics on inverse virtual screening will be presented to give the audience a sense of how off-target effects of drugs can be computationally examined or addressed.

12.2 Overview of Virtual Screening

12.2.1 Principle of Virtual Screening

In virtual screening, the overall aim is to computationally screen through a database of chemical compounds that would be tested in the wet lab. A typical workflow can be represented as follows:

1) Prepare inputs (receptor, pharmacophore, etc.) for CADD method
2) Format chemical compound database
3) Screen through database with CADD method
4) Rank compounds in database by prediction scores
5) Select top *n* compounds for experimental validation

Compound databases can vary greatly in size depending on the target of interest, ranging from tens of thousands to millions of compounds. There is a general misconception from the scientific community that virtual screening is a complicated process; to an extent, it is beautifully simple. One way to envision screening is the large-scale repetition of a CADD methodology against each compound in the database, analogous to a loop in a computer program. However, the intricacies and challenges of virtual screening are found primarily in the parameterization of the CADD methodology, as well as the way one processes the resulting predictions. After a virtual screen, the CADD methodology will have assigned a metric or score to each compound. Subsequently, the compounds will be ranked from most likely to least likely to bind or interact with the receptor of interest. The top-ranked compounds are then typically chosen for experimental validation, either by selection after clustering or visual inspection of docked poses.

Before going further, it should be noted to the reader that a proper understanding of various computer representations of chemical compounds is useful for troubleshooting errors related to file formats while using a CADD methodology. Moreover, the proper design of a chemical database is paramount to the success of a

virtual screening campaign. Finally, a stringent implementation of a retrospective or prospective virtual screening for a drug target of interest is a necessity for validation. These will all be detailed in the subsequent Sections 12.2.2–12.2.4.

12.2.2 Computer Representation of Chemical Compounds

12.2.2.1 Line Notation

Line notation allows for the representation of a chemical compound using a string of ASCII characters. Despite looking rather odd to the untrained eye, they are completely readable, and those familiar with the format would be able to convert between it and the corresponding 2D chemical structure. Nowadays, this representation is primarily used for chemical database searching. The *S*implified *M*olecular-*I*nput *L*ine-*E*ntry *S*ystem (SMILES) and *In*ternational *Ch*emical *I*dentifier (InChI) formats are currently the most widely used.

SMILES strings were initially conceived in the 1980s as a means to make chemical compounds machine readable. Each letter represents an atom (B, C, N, O, P, S, F, Cl, Br, or I), single bonds are usually implicit, and double and triple bonds are represented as "=" and "#", respectively. Aromaticity is denoted with alternating equal signs (e.g. pyridine moiety: C4=CC=CC=N4). Additionally, rings are classified by including an opening and closing number (e.g. thiophene moiety: C1=C(SC=C1)). The use of parentheses indicates branching, and stereochemistry is specified at chiral centers with "@". *SM*iles *AR*bitrary *T*arget *S*pecification (SMARTS) strings were developed by the Daylight Chemical Information Systems as a robust extension of the SMILES string that provided expanded functionality, such as the ability to filter a compound database by substructure. However, one of the biggest drawbacks of this format is that there is no standard way to generate the SMILES string [7]. Thus, the heterogeneity of SMILES strings possible for a single compound can complicate chemical database searching, especially when a compound of interest cannot be found due to this problem.

InChI strings were developed in 2005 by the *I*nternational *U*nion of *P*ure and *A*pplied *C*hemistry (IUPAC) in response to the inconsistencies produced by SMILES strings [8]. Additionally, they were able to express more information than SMILES strings. All InChI strings start with "InChI=", followed by the version number and an "S", which corresponds to its standardization. Subsequently, there are six layers of information; the first layer is the most important and gives the chemical formula, atomic connections, and hydrogen atoms, while the others focus on other chemical aspects such as charge, stereochemistry, and isotopes. Also, it should be noted that the InChI format is conspicuously more difficult to read than SMILES. InChI keys, 27-character hashed versions of InChI strings, allow for extremely fast chemical database searches due to their reduced length. A previous study has demonstrated that a single duplicate for the

first 14 characters could theoretically occur 0.014% of the time in a database of 100 million compounds [9]. Given that most chemical databases have well below this number of chemical compounds, it can be assumed that a duplication will likely not occur. A drawback of using the InChI key is that it cannot be converted back to its respective InChI string, thus these two descriptors always need to be paired.

12.2.2.2 Molecular Fingerprints

Molecular fingerprints provide an abstraction of the chemical features of compounds into binary vectors. All have a fixed length for purposes of comparison and can be used to calculate chemical similarity mind-bogglingly fast. Though efficient, they likely have the least specific information packed into their form. Over the years, various developments have aimed to squeeze as much information as possible into small vector lengths.

Substructure key-based fingerprints consist of a predefined set of substructures, and the number of possible bits is defined by the number of substructures. One of the most commonly used fingerprints of this type is *M*olecular *ACC*ess *S*ystem (MACCS), first developed by MDL Information Systems (formerly Molecular Design Limited) in 1979. Interestingly, they were initially intended for use in database searching as opposed to virtual screening [10], which is the common method it is used for today. They assume two different variants: one with 960 substructures, and the other with 166 of the most interesting substructures for drug discovery, paired with corresponding SMARTS strings [10]. Not surprisingly, the latter is far more popular. The principle of how this type of fingerprint works is that each position in the fingerprint corresponds to a substructure. If the compound has the substructure in its chemical structure, then the bit will be set to "1". Otherwise, it would be set to "0". A drawback to using these types of fingerprints is that they are usually relatively sparse in content, such that they will have mostly zeros, as typical molecules will have very few of the substructures.

Path-based fingerprints are constructed by analyzing every possible fragment in a molecule of a given linear path length, then cryptographically mapping them onto a fixed-length fingerprint. An example is given in Figure 12.1a for oliceridine, using a path length of 3. Occasionally, bit collisions occur when the same bit is assigned to two different fragments. However, this is not a common occurrence and can be reduced by increasing the fingerprint length. The Daylight fingerprint, developed by Daylight Chemical Information Systems (hence the namesake), is the most used out of all of the fingerprints of this type and typically consists of 1028 bits.

Circular fingerprints are very similar to path-based fingerprints in that they are mapped from a collection of molecular fragments onto a fixed-length fingerprint. However, their method of fragment analysis is not based on fragments generated

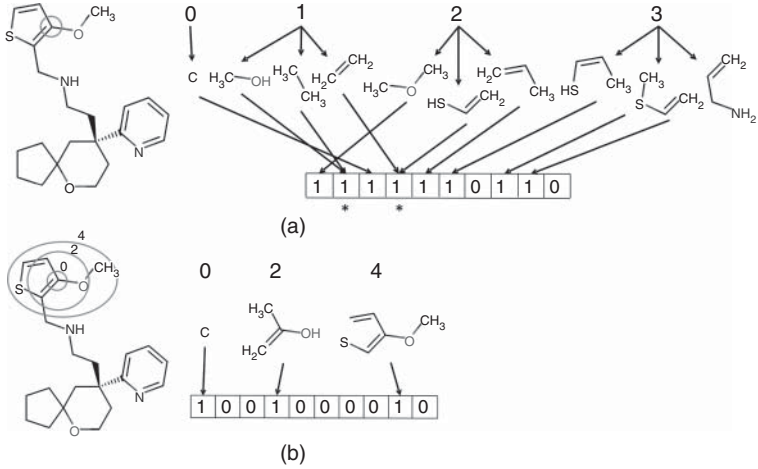

Figure 12.1 Hypothetical 10-bit fingerprints for Oliceridine. (a) A path-based fingerprint with a path length of 3 is used for this example. Only fragments found from a single starting atom (green circle) are shown. The path lengths of the fragments (0, 1, 2, 3) are numbered in bold. The asterisks (*) denote where there are bit collisions. (b) A circular fingerprint with a radius of 2 is used for this example. Only fragments found from a single starting atom (innermost green circle) and onwards are shown. The diameters of the fragments (0, 2, 4) are numbered in bold. For both fingerprints, the fragment-generating process occurs for all atoms on the molecule. MarvinSketch was used for drawing and displaying the chemical structures, MarvinSketch 18.10.0, 2018, ChemAxon (http://www.chemaxon.com).

in a linear path, but rather, the chemical environment centered around each atom within a certain radius. An example for oliceridine is given in Figure 12.1b, where a radius of 2 was used. Here, fragments are generated by moving a certain radius away from a starting atom up until a diameter of 4, resulting in 3 fragments for the specified starting atom. The ECFP4 fingerprint is the industry standard of this type, and not surprisingly, it has been shown to be among the best performing fingerprints in a recent benchmark that ranked diverse structures by similarity [11].

12.2.2.3 Chemical Table Files

Another strategy for storing chemical information in a text file is chemical table file family of file formats. Originally developed by MDL Information Systems starting in the late 1970s [12], they have become one of the most widely used file formats, having been adopted by a vast majority of computational chemistry software. Of those in this family, focus will be upon the *S*tructure-*D*ata *F*ile (SDF) format due to its widespread use. In brief, the file starts with a three-line header block, which is mandatory but can be left empty if desired. This is followed by a "counts" line,

which consists of specifications such as number of atoms, number of bonds, and so forth. The "atoms" block provides information about the coordinates and identity of the atoms, while the "bond" block describes the connectivity between atoms. The properties block denotes any existing charges or isotopes, as well as the end of the molecular description. SDF is unique in that the subsequent associated data allow the inclusion of miscellaneous information not allowed in the main form, such as the IUPAC name and database identifiers. The tag for each data type is included inside angle brackets ("<", ">"), and the relevant data is placed on the line immediately following it.

Originating from the now-defunct Tripos, the Mol2 format has achieved a similar level of popularity and usage as the SDF format. Many aspects are almost identical to the SDF format, where various blocks are designated for counts, atom, and bond information, though with different column formatting. Moreover, each block is recognized starting with a record type indicator (e.g. @<TRIPOS>ATOM), followed by the corresponding data. Apart from the main record type indicators, there exists many others not available in SDF format, such as substructures and rotatable bonds.

12.2.2.4 PDB File Format

Most researchers in biochemistry will be fairly acquainted with the Protein Data Bank (PDB) file format, since it has been primarily used to describe the three-dimensional structure of proteins, DNA, and RNA. This file format was first conceived in 1976 as a means to help researchers exchange protein coordinates through a database [13]. Not surprisingly, its format has been revised and updated numerous times over the years. Essentially, a PDB file is a text file that contains various information about the structure provided in specified ranges of columns. The file contains a variety of data, ranging from resolution and method used to solve the structure to atomic coordinate specifications.

One of the most important pieces of information within the PDB file is the "ATOM" record name. An example is shown in Figure 12.2 that depicts the coordinates for two representative amino acids from a PDB structure. Each line depicts a single atom in the structure. For example, the first line corresponds to the backbone nitrogen of Gly-85. Furthermore, the atomic coordinates of this atom (−2.211, 29.344, −42.463) are given so that whichever algorithm or molecular visualization software is used can correctly process this representation.

12.2.3 Chemical and Biological Databases

As the amount of data available to the scientific community has increased over time, there has become a distinct need to catalogue and organize it so that it can be easily accessible. Truly, gone are the days of hours-long expeditions to the library

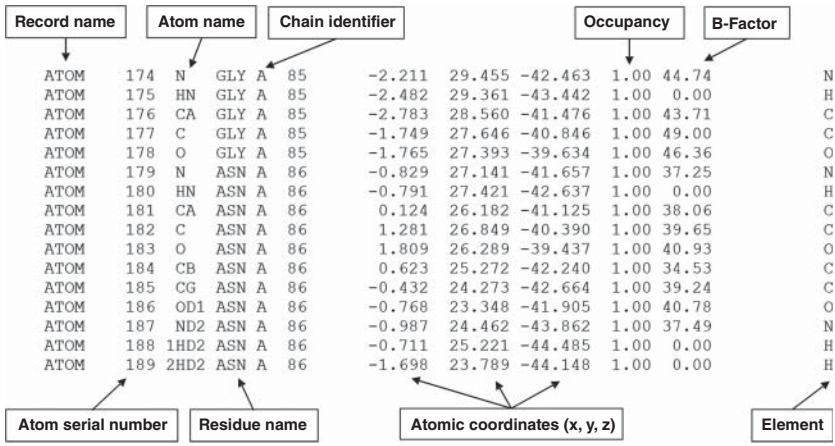

Figure 12.2 Representative portion of PDB file. The portions with the "Record Name" of ATOM helps software understand the identity and location of atoms and therefore help correctly process relevant information from the file. The amino acids, Glycine (GLY) in position 85 and asparagine (ASN) in position 86, from this structure are shown.

in search of publications that may or may not have been helpful to the question at hand. Astoundingly, there now exist public databases that index data anywhere from the primary structures of proteins to various experimental values of ligands for a given receptor.

12.2.3.1 Biological Databases

UniProt is the *de facto* standard source of information for proteins [14]. This database originated from the merging of data from European Bioinformatics Institute (EBI), Swiss Institute of Bioinformatics (SIB), and Protein Information Resource (PIR) into an entity known as the UniProt consortium. The most commonly used portion of the database is referred to as UniProt Knowledgebase (UniProtKB), which is subdivided into Swiss-Prot and TrEMBL. The former collection of data is manually annotated and reviewed by experts of each respective protein, while the latter refers to those that are computationally annotated from genomic data. Not surprisingly, TrEMBL contains a far larger quantity of data than Swiss-Prot. Within Swiss-Prot, a multitude of information about a protein of interest is available, such as primary structure, post-translational modifications, function, subcellular localization, and known protein–protein interactions.

The *P*rotein *D*ata *B*ank (PDB) is the single largest repository for protein, DNA, and RNA structures solved by structural biologists [15]. It began as a united effort in the 1970s to provide the scientific community with protein structures coded into punch cards [13]. As the Internet came into fruition, it became possible to move the data onto an online platform for its higher throughput distribution.

Thus, the first web-server for browsing the PDB was developed at Brookhaven National Laboratory in 1996 [16]. With the explosion of solved structures starting in the 1980s, this resource became increasingly invaluable to life science researchers around the world. In its current state, it serves as the primary resource that provides protein structures for structure-based drug discovery efforts. The G Protein-Coupled Receptors database (GPCRdb) [17] was developed in 1993 as repository for GPCR-related data, and after the GPCR structure boom, it has continually participated in the manual curation of GPCR structures. A more recent effort to catalogue experimental GPCR structures from our group in a user-friendly fashion is GPCR-EXP, which is semi-manually curated and updated weekly (https://zhanglab.ccmb.med.umich.edu/GPCR-EXP/).

12.2.3.2 Chemical Databases

First released in 2009, ChEMBL is arguably the most massive database for molecules with drug-like properties and biological activity [18]. As of its latest release (ChEMBL 25), the database contains 1 879 206 unique compounds corresponding to 12 482 targets and 15 504 603 activities from 72 271 publications, all derived from manual annotation. A similar database founded over a decade earlier at University of California at San Diego is BindingDB [19], which also contains a large amount of manually curated affinity data. However, it has less of a focus on membrane receptors than ChEMBL and more strongly emphasizes enzyme targets [20]. DrugBank is a chemical database whose topic of interest is information on drug molecules and their corresponding targets [21]. Another interesting database of note is Psychoactive Drug Screening Program's (PDSP) K_i database [22], which houses a sizeable number of experimental affinities. A large portion of their data is dedicated to GPCRs. Also, the *I*nternational *U*nion of Basic and Clinical *PHAR*macology's (IUPHAR) Guide to Pharmacology is a chemical database that deals primarily with popular pharmacological targets, such as GPCRs and ion channels [23]. It is manually curated by experts, and only ligands that have been well characterized are included. In contrast, ChEMBL, BindingDB, and PDSP K_i are looser in their criteria for inclusion, where the binding mode or mechanism are largely unknown for most ligands. PubChem is a pure chemical database maintained by the National Center for Biotechnology Information (NCBI) [24], containing approximately 93.9 million chemical compounds. Additionally, they have a gargantuan collection of bioactivity data from about 1.25 million high-throughput screening campaigns, each with several million values.

All of the aforementioned chemical databases contain GPCR-related experimental data. One of the earliest efforts in organizing such data was with G protein-coupled receptor-LIgand DAtabase (GLIDA) [25, 26]. Moreover, our

group developed a database called G protein-coupled receptor-Ligand ASSociation (GLASS) database [27], which processes and unifies GPCR experimental data across ChEMBL, BindingDB, IUPHAR, PDSP K_i, and DrugBank, and remains the most comprehensive database of its type. It has been used in other algorithms as input, such as SwissSimilarity [28] and weighted deep learning and random forest (WDL-RF) [29].

12.2.4 Retrospective and Prospective Virtual Screening

Retrospective virtual screening is performed to computationally validate a method's predictive performance based on a set of known active compounds and their associated decoys. Moreover, it is employed as the primary method of theoretical studies in benchmarking virtual screening methods, as well as serves as a calibration of predictive conditions so that prospective virtual screens are optimally successful. In brief, decoys are compounds that likely do not interact with the receptor of interest but are similar in some way with the active compound. A common way to produce decoys is to generate compounds that are similar in one aspect but different in another. For example, directory of useful decoys (DUD) [30] and directory of useful decoys enhanced (DUD-E) [31] are two such datasets that provide 33 and 50 decoys, respectively, per active compound that are chemically similar yet topologically different. Additionally, GPCR-Bench [32] and GPCR ligand library/GPCR decoy database (GLL/GDD) [33] are GPCR-specific datasets created in a similar fashion. It is important to note that the core assumption of using an abundance of decoys over actives is that most compounds will not bind to a given target by sheer chance, in principle making them hypothetical inactive compounds. Furthermore, it is also recommended to add experimentally determined inactive compounds to the decoy set whenever available, as is done by DUD-E.

After all active compounds and decoys are all scored, they will be ranked accordingly. The goal is to try and get as many active compounds into the top-ranking portion as possible. A typical metric for evaluation is the enrichment factor of the top 1% ($EF_{1\%}$), given as follows:

$$EF_{1\%} = \frac{N_{act}^{1\%}/N_{select}^{1\%}}{N_{act}/N_{tot}} \tag{12.1}$$

where N_{act} and N_{tot} are the total numbers of the active and all compounds, respectively. $N_{act}^{1\%}$ and $N_{select}^{1\%}$ are, respectively, the number of active ligands and the number of all candidates in the top 1% of the ranked database. The numerator essentially accounts for the proportion of active compounds found in the top 1%, while the denominator represents the probability of selecting an active compound

randomly from the entire database. An $EF_{1\%}$ greater than 1 would mean the screening method performed better than randomness, while if it were less than 1, it would mean it performed worse than randomness.

An important metric typically paired with the enrichment factor is the area under the curve (AUC) of the receiver operator characteristic (ROC) curve, where the true and false positive rates are calculated for both the active and decoys, respectively. A value greater than 0.5 would suggest better performance than randomness, while the inverse would suggest worse than randomness. A drawback to this metric is that the entire list of ranked compounds is taken into account, thereby putting great emphasis on the portions of the database that are unlikely to ever have a compound chosen for experimentation. To account for this, some groups have developed ways to give a higher weight to early enrichment. The Shoichet group utilized the logAUC metric in DUD-E [31], while Schrödinger seems to favor the Boltzmann-enhanced discrimination of the receiver operator characteristic (BEDROC) [34, 35]. Either of these metrics will allow a better examination of how well the CADD methodology is able to distinguish active compounds from decoys.

As opposed to its counterpart, prospective virtual screening is far more straightforward: in a prospective screen, a computational chemist chooses several high-scoring compounds for experimental validation. One way of choosing compounds would be to cluster a subset by molecular frameworks (i.e. Bemis–Murcko scaffolds [36]) or chemical similarity, wherein the top-ranking compound in each cluster would be chosen. Another way would be to visually examine the docking poses of the compounds for important interactions with the receptor. Unfortunately, there is no replacement for a prospective virtual screen, as retrospective virtual screens remain theoretical in nature and serve only as a measure of how believable virtual screening can be.

12.3 Conventional Virtual Screening

In the traditional sense, virtual screens are categorized as ligand based or structure based, depending on which CADD methodology is used; the former utilizes pure chemical information in its search process, whereas the latter uses protein structural information to determine how a compound binds. Most of the major modeling suites from various companies (i.e. molecular operating environment [MOE], Schrodinger, Cambridge Crystallographic Data Centre (CCDC), BIOVIA, etc.) have the capability for all or most of the following methods.

12.3.1 Ligand-Based Approaches

12.3.1.1 Chemical Similarity

Molecular fingerprints are most often used in the calculation of chemical similarity. As a brief reminder, they are composed of a fixed-length string of bits; the presence of "1" in a position denotes the presence of a chemical feature, while "0" denotes its absence. The simplicity of this form allows for the possibility of blazing fast calculations. A multitude of similarity metrics are available, but the Tanimoto coefficient has proven to perform the best and therefore has been most popular [37]. Its calculation is shown as follows:

$$\text{Tanimoto Coefficient} = \frac{c}{a + b - c} \tag{12.2}$$

where a is the number of bits in the first molecule, b is the number of bits in the second molecule, and c is the number of shared bits between the two molecules. Only the same type of molecular fingerprint can be compared between molecules; mixing different types will lead to erroneous results. Free software for chemical similarity screening includes OpenBabel [38] and RDKit [39], both of which are user-friendly standard toolkits in cheminformatics.

12.3.1.2 Ligand-Based Pharmacophores

A pharmacophore is a collection of chemical features (H-bond donor, H-bond acceptor, aromatic, etc.) represented spatially, developed from a set of known bioactive compounds. They are typically used when the protein structure of the target is not known, which was historically the case. In brief, the pharmacophore is constructed by structurally superposing low-energy conformers of the bioactive compounds, whereupon chemical features from superposed moieties among the compounds are assigned to the model. It should be noted that it is assumed that the shared chemical features contribute to the bioactivity. When used in a virtual screen, the target compounds will be spatially matched onto the pharmacophore model and scored.

12.3.1.3 Shape-Based Comparison

Molecular shape has long been established as being an important contribution to bioactivity [40]. Thus, another common method in ligand-based virtual screening has involved the use of shape-based matching. Query conformers are geometrically matched to other target conformers, trying to achieve the best 3D electron density overlap. One of the earliest algorithms to implement this approach was ROCS [41] from OpenEye Scientific Software, while other freely available methods include Ultrafast Shape Recognition (USR) [42], LigSift [43], and LS-align [44].

12.3.2 Structure-Based Approaches

12.3.2.1 Structure-Based Pharmacophores

When the drug target of interest has a protein structure available, then a structure-based pharmacophore can be employed in virtual screening. This type operates similarly to the ligand-based pharmacophore, except the spatial distribution of chemical features are informed by the ligand-binding pocket as opposed to a ligand structural superposition. Moreover, these can be used when there is little to no knowledge of how a ligand binds to a receptor, especially in the case of orphan receptors. However, the selection of chemical features to be used in the model is nontrivial, given the large amount of uncertainty of residue importance in the binding site, and warrants careful decision-making.

12.3.2.2 Molecular Docking

Molecular docking is a method used to predict how a ligand binds with a receptor through conformational search and scoring functions. Prior knowledge of the ligand binding site is typically required in order to optimize the area to be examined. Protocols for most docking programs start with adding non-polar hydrogens and partial charges to both the receptor and the compounds. This is then followed by the docking algorithm performing a conformational search for the most favorable ligand pose, which is evaluated with a scoring function at each step. Subsequently, the top poses are generated for the user, who can then visualize them with a molecular viewer. Additionally, the final scores for each predicted pose are also given.

There have been dozens of docking software programs developed over the years, and each has approached docking in a different way. Some of the major differences between these are: (i) the search algorithm, (ii) scoring function, and (iii) conformational flexibility of the ligand and the receptor. Among the top methods employed for conformational searching are the Lamarckian genetic algorithm (AutoDock [45]), genetic algorithm (GOLD [46]), local search global optimizer (AutoDock Vina [47]), ant colony optimization (PLANTS [48]), anchor-and-grow (DOCK 6 [49]), and exhaustive search (Glide [50, 51]). Though these strategies differ greatly in their search algorithms, their basic premise remains the same: they aim to achieve the most favorable ligand pose. To do so, a scoring function must be calculated at each step of conformational sampling to evaluate the pose. Many of the functions used currently are physics-based force fields that approximate the binding energy of the ligand pose in the binding site. For example, the scoring function from DOCK 6 simply uses van der Waals and electrostatic terms for computational efficiency [52]. Various others take other physical terms into account, such as hydrogen bonding, ligand desolvation, and hydrophobic contributions [46]. Additionally, there exist empirical scoring functions, which estimate the binding energy using a set of weighted energy

terms, and knowledge-based scoring functions, which utilize statistical energy potentials derived from experimentally solved structures [53, 54]. Finally, there is the option of how to treat the receptor and ligand during docking with respect to conformational flexibility. Most software packages make the receptor rigid because of the computational rigor involved in sampling all receptor conformations. However, some programs provide an option to make certain side chains of the receptor flexible, such as AutoDock Vina [47] and GOLD [46]. Schrödinger has an induced-fit docking protocol that allows for both ligand flexibility and conformational changes in the binding site, though its application to virtual screening of a large number of compounds is limited due to its computational intensity. Most of the earliest docking methods, such as the original DOCK [55], treated the ligand as rigid in order to find molecules with shape complementarity to the binding site. Nevertheless, this methodology's success is dependent on the conformation of the molecule being docked, which may be a vastly different conformation from what is observed in reality.

12.3.3 Application to GPCRs

To date, there exists a plethora of prospective virtual screening campaigns applied to GPCRs. To list them all would be outside the scope of this chapter, but the reader can find further information from reviews [56–58]. Some particularly interesting studies have included the discovery of: (i) a biased agonist for the μ opioid receptor [59], (ii) selective agonists for the serotonin 1B receptor over the serotonin 2B receptor [60], and (iii) antagonists for the C–X–C chemokine receptor 4 [61].

Several compounds found initially from virtual screen campaigns have actually entered clinical trials. In 2006, a group from Predix Pharmaceuticals (known as Epix Pharmaceuticals before collapsing) performed a docking-based virtual screen on a homology model of the serotonin 1A receptor that resulted in a potent, selective agonist [62]. The reported molecule became the drug candidate, Naluzotan, and proceeded into a phase III clinical trial; ultimately, it failed to perform better than the placebo and was discontinued [63]. From the same company, a separate virtual screening campaign with the serotonin 4 receptor using a similar methodology produced a selective, partial agonist that also made it to clinical trials, though it too failed [64]. In another study from Heptares Therapeutics, a novel adenosine A_{2A} receptor antagonist was discovered through a docking-based virtual screen on homology models, called AZD4635, and is currently in phase II clinical trials for lung cancer [65]. Additionally, they also have a muscarinic M4 receptor agonist, HTL0016878, in phase I clinical trials for Alzheimer's disease, found through similar methods [66]. To the best of the authors' knowledge, there has yet to be an approved drug found from virtual screening targeting GPCRs, though many examples exist for various other targets, such as growth factor-β1

receptor kinase [67]. Despite this, numerous GPCR-targeting drugs resulting from virtual screening await their verdict in clinical trials, and it is only a matter of time before one hits the market, which would undoubtedly validate the current computational methods and increase confidence in virtual screening for GPCRs.

12.3.4 Challenges

Despite the relative successes each type of virtual screening approach has had, there are distinct advantages and disadvantages to each. Ligand-based methods, such as chemical similarity, are computationally inexpensive and can screen millions of compounds within a short time but have the drawback of being biased towards the known ligands used to build the model. Conversely, structure-based methods, such as molecular docking, inherently have no such bias, but they are extremely computationally expensive relative to ligand-based methods. A trend in recent years has culminated in the combination of these methods to address their respective shortcomings [68]. Given the speed of ligand-based methods, several studies experimented with using it to first produce an "enriched" database of top-ranking compounds, followed with molecular docking on this reduced subset [69–71]. This can greatly reduce the computational cost and enable virtual screening for groups without access to high performance computing clusters. Other groups have exploited the complementarity between ligand- and structure-based methods by running them in parallel and employing a consensus scoring system for ranking [72–74]. Regardless of the strategy used, the manual selection of bioactive compounds remains a great challenge.

12.4 Chemogenomics-Based Virtual Screening

Oftentimes, a drug target will have neither structural information nor known active compounds. In cases like this, related proteins can be used to infer what compounds the drug target can bind, based on the assumption that similar receptors bind similar ligands [75]. The sequence of the protein can be used in a sequence-based alignment search to find homologous proteins with sets of bioactive compounds. Alternatively, a structural homology model of the protein can be generated based on the sequence and structurally compared with all known protein structures to find related proteins.

FINDSITE was an early implementation of a chemogenomics-based virtual screening algorithm, which utilized ligand information from structurally homologous receptors found through fold-recognition in a ligand-based virtual screening (VS) [76]. This algorithm later evolved into FINDSITEX, which used modelled structures as structural templates instead [77]. FINDSITE^{comb} and its successor, FINDSITE$^{comb2.0}$, incorporated FINDSITEX along with an improved

version of FINDSITE that filters out false-positive ligands (FINDSITE^{filt}), vastly improving its performance in benchmarks [78, 79]. Additionally, another recent algorithm is PoLi, developed from the same lab; it looks for similar receptors by performing binding pocket structure comparison between the target and templates, followed by a ligand-based screening search [80]. SPOT-Ligand [81] is an algorithm that employs global structure alignment for the acquisition of protein structures that are structurally similar and have sets of bioactive ligands, which are then used in a ligand-based virtual screen. An updated version of the method, SPOT-Ligand 2 [82], included a more comprehensive protein–ligand database, and consequently, it achieved a better performance than its predecessor. Recently, our group has developed a pipeline, *M*ichig*a*n *G* prot*e*in-coup*l*ed receptor *lig*a*n*d-based virtual screen (MAGELLAN), that utilizes structure- and sequence-based similarity to find homologous GPCRs, whereupon their ligands are clustered and then used to construct ligand profiles; using a consensus scoring function, a ligand profile-based virtual screen can then be performed against a database of choice [83].

12.5 Bioactivity Modeling with Machine Learning

Machine learning is the study of algorithms and statistical models where computer systems are used to effectively implement a specific task without explicit instructions, instead drawing information from inference and patterns. Machine learning algorithms construct a mathematical model using sample data, called "training data," to improve the performance P at a task T based on experience E [84]. Machine learning has been applied in a wide variety of applications, such as computer vision and e-mail filtering.

12.5.1 Pipeline

The aim of machine learning in virtual screening is to incorporate data from multiple sources into sensible models for describing and screening compounds with the goal of identifying active drug targets. The typical machine learning pipeline begins from data acquisition, proceeds to feature engineering, then to algorithm selection and model construction, and finally to model evaluation and application. Figure 12.3 shows the overview of a typical machine learning-based virtual screening workflow.

12.5.2 Data Preparation

Data preparation is the step of transforming and cleaning raw data prior to processing and further analysis. *Steve Lohr of The New York Times said:* "*Data scientists,*

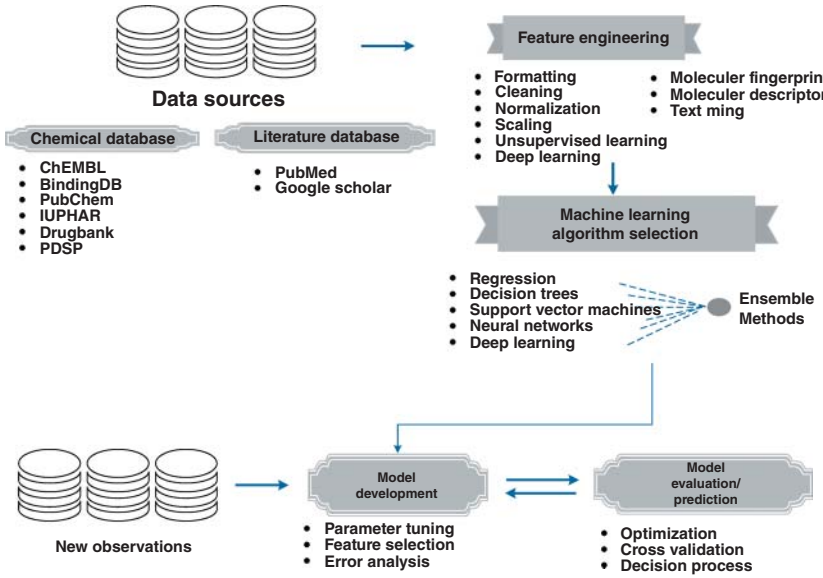

Figure 12.3 Typical machine learning workflow.

according to interviews and expert estimates, spend 50 percent to 80 percent of their time mired in the mundane labor of collecting and preparing unruly digital data, before it can be explored for useful nuggets." Thus, it is important and often involves reformatting data, imputation of missing values, and the combination of datasets to increase sample size. Data preparation usually is a lengthy undertaking for scientists, but it is essential as a prerequisite to put data in its proper context in order to gain meaningful insights and eliminate bias resulting from poor data quality.

12.5.3 Feature Selection

Feature selection aims to reduce the dimensionality of data by using only a subset of the most important features to build a model. Selection criteria usually consist of the minimization of predictive errors for models given diverse feature subsets. Algorithms search for a subset of features that can optimally model measured responses, subject to specific constraints, such as ℓ_1-norm and ℓ_2-norm regularization. As shown in Figure 12.4, there are three general classes of feature selection algorithms: filter methods, wrapper methods and embedded methods.

Filter methods usually adopt a statistical measure to determine a score for each feature. The features can be ranked by the scores and are either selected to be saved or removed from the model. The methods usually are univariate and consider each feature independently or with regards to the dependent variable. Some examples

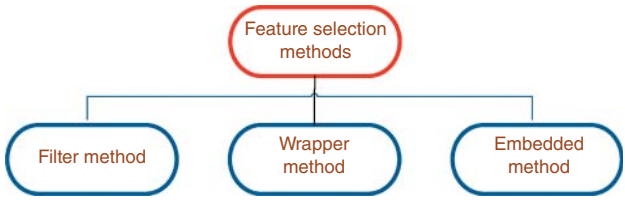

Figure 12.4 Classes of feature selection algorithms.

of filter methods involve the Chi squared test, information gain, and correlation coefficient scores.

Wrapper methods regard the selection of a subset of features as a search problem, in which different combinations are generated, evaluated and compared to each other. A predictive model is built to evaluate the combinations of features and to assign a ranking according to the model accuracy. The search process can be methodical such as a best-first search, or it may be stochastic such as a random hill-climbing algorithm, or it may use heuristics, like forward and backward ways to add and remove features, such as the recursive feature elimination algorithm.

Embedded methods aim to learn which features have the best contribution to the model performance while the model is being built. The most common class of embedded feature selection is regularization-based methods. Regularization methods are also called as penalization methods that introduce additional constraint terms into the objective function of a predictive algorithm that push the model toward lower complexity. Examples of regularization algorithms involve the LASSO, Elastic Net, and Ridge Regression.

12.5.3.1 A Case

We proposed a new method SED to predict ligand bioactivities and to recognize key substructures associated with GPCRs through the coupling of screening for LASSO of long extended-connectivity fingerprints (ECFPs) with deep neural network training [85]. Shown in Figure 12.5, the SED pipeline contains three successive steps: (i) representation of long ECFPs for ligand molecules, (ii) feature selection by screening for LASSO of ECFPs, and (iii) bioactivity prediction through a deep neural network regression model.

12.5.4 Algorithms

12.5.4.1 Traditional Algorithms

Traditional applications of machine learning in virtual screening focus on the use of supervised techniques to train statistical learning algorithms to classify databases of molecules by their activity against a particular drug target.

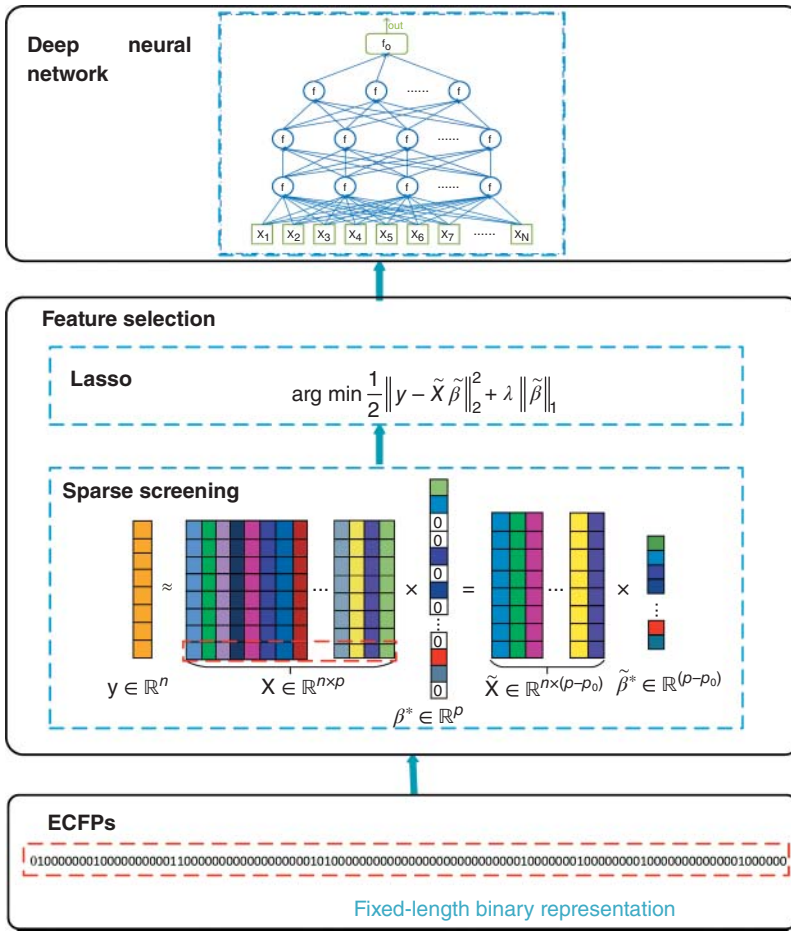

Figure 12.5 Schematic of SED. The approach is composed of three stages: long extended-connectivity fingerprint (ECFP) representation for ligand molecules, feature selection by screening for LASSO, and construction of deep neural network regression prediction models. Source: Wu et al. [85]/with permission of Oxford University Press.

Both ligand-based virtual screening and structure-based docking benefit from machine learning algorithms, including naïve Bayesian classifiers, neural networks, support vector machines, and decision trees, as well as more regression techniques.

The most simple machine learning method in virtual screening is multiple linear regression (MLR), which has been widely used in quantitative structure–activity relationship (QSAR) software [86–89], such as CoMSIA [90]. For instance, Evers et al. proposed a method of integrating linear regression and classification methods

with discriminative analysis to find potential ligands targeting GPCRs from the Molecular Design Limited drug data report (MDDR) [89]. For four GPCR drug targets, Feher et al. employed consensus scoring to combine multiple different ligand-based methods using 2D descriptors, and 3D pharmacophore models [91]. Another usage of linear regression in virtual screening is in structure-based docking approaches [92, 93]. For instance, Jacobsson and Karlén employed partial least squares (PLSs) to correct for the size bias of many docking scoring functions [93].

Conceptually, one of the popular methods to model the activity of a molecule is to search a database for the molecules that are the most similar to it. The predicted score values are the activity of these "nearest neighbors," a method referred to as "k nearest neighbors" (kNN). A few attempts have been made to improve the model performance of virtual screening based on the kNN methods, such as CoLiBRI [94], ENTess [95], MFMNN [96] methods and several lazy methods [97, 98].

Naïve Bayesian is among the simplest classifiers. The probability of activity is determined by the ratio of actives to inactives that share the descriptor value. This approach supposes that each descriptor is statistically independent. The earliest usage of a naïve Bayesian method in virtual screening was for Binary QSAR, introduced by Labute [99]. Following this work, several excellent applications of the naïve Bayesian model have been put forward [100–102]. For instance, Glick et al. adopted a naïve Bayesian model with ECFP descriptors as a post-processing method in order to prioritize hits from experimental HTS data [102]. Then, they extended the usage of the naïve Bayesian model in docking to take the place of HTS as a source of potential active compounds [101].

Support vector machines (SVMs) were proposed by Vapnik [103], where a separation hyperplane boundary was defined, and examples were classified depending on which side of the boundary they are located. The initial work on the usage of SVM in virtual screening was carried out by the Willett group, where 35 991 molecules in the National Cancer Institute (NCI) AIDS data set were tested, using UNITY fingerprints as attributes [104]. Following this work, multiple excellent virtual screening methods based on SVM have been proposed [105–108]. For example, a prospective usage of SVM in virtual screening was developed by Schneider et al., who adopted a model which was built from 331 dopamine receptors inhibitors for screening the SPECS and Interbioscreen databases (over 255 000 molecules combined) which resulted in the identification of 11 compounds with a high selectivity for the D3 receptor [108].

Artificial neural networks (ANNs) have played a long-established role in cheminformatics. ANNs are composed of a set of connected artificial "neurons." A single neuron takes multiple numerical inputs, and outputs a transformed and weighted sum of the inputs; through layers of many parallel neurons, complex classification functions can be defined through training. Multiple uses of ANNs to produce consensus scoring functions for docking have been proposed [109, 110].

For instance, Sem et al. adopted ANNs to generate a consensus score for predicting CYP2D6 binding affinity through combination of the AutoDock and XScore tools [110]. ANNs have also been applied for VS beyond the pharmaceutical area, for example in heterogeneous catalysis [111].

A decision tree denotes the conjunction of a series of "rules," each of which is a predicate concerning a subset of descriptors. The process of training a decision tree model determines which rules are involved in the tree, and what classification is made at each leaf. The process is usually carried out by choosing a valuable rule that can divide the training data into two or more groups. The process is then repeated for each of the subsets, until a termination criterion is reached. A few applications of decision trees in VS propose novel classification tools in QSAR [112, 113]. For instance, Jones-Hertzog et al. proposed a decision tree method for screening 23 000 compounds against 14 GPCR drug targets, which outperforms random selection and similarity searching in the majority of cases [112]. Yamakazi developed a single decision tree, which was trained on 130 PDE-5 inhibitors and 10 000 inactives, to screen 50 520 molecules in the SPECS database [113].

Ensemble methods denote a series of classifiers that combine the output of base classifiers to arrive at the final decision. Classical ensemble methods include bootstrapping, boosting, etc. Simple examples of the ensemble method in VS include the study of van Rhee et al., who adopted pairs of decision trees to screen 3000 molecules, which demonstrated a 13-fold enrichment over the hit rate of a 14 000 member HTS when screening against an ion channel target [114]. Random forest (RF) represents one of the most popular ensemble methods; the model consists of an ensemble of many randomly generated decision trees. Due to its ease of use, high accuracy, and robustness to adjustable parameters, RF has become a "gold standard" for QSAR method comparison [115]. Svetnik et al. have examined the usage of RF against several datasets, including the same HTS data which was used to evaluate boosting, and it was shown to have competitive performance [116]. RF has also been applied in conjunction with docking. For example, Teramoto and Fukunishi adopted a RF model to predict the root mean square deviation (RMSD) of a docked conformation relative to the bioactive conformation [117].

12.5.4.2 New Algorithms

Deep learning is a branch of machine learning consisting of a set of algorithms that attempt to model high-level abstractions in data by using multiple processing layers, with complex structures or otherwise, composed of multiple non-linear transformations. Recently, deep learning-based methods have witnessed impressive success in ligand-based virtual screening [118–122]. For instance, in 2012, Merck organized a challenge for the design of machine learning methods to model the bioactivities of ligands acting with target proteins, and methods using deep

learning achieved the best performance. Later, Ma et al. (2015) proposed a deep neural net model for determining QSARs, which demonstrated better performance than random forest models for most of the data they studied [115]. Most recently, we proposed a weighted deep learning algorithm that takes arbitrarily sized inputs and generates bioactivity predictions which are significantly more accurate than the control predictors with different molecular fingerprints and descriptors [123].

Applications of deep learning models in *de novo* ligand design have also been developed. There are multiple examples of models which adopt autoencoders and/or recurrent neural networks that can produce new molecules with ideal properties [124–127]. The usage of autoencoders also permit the representation of molecules as short, real-valued vectors, which are extracted from the bottleneck layers, in order to facilitate exploration of the chemical space [125].

Although deep learning is more readily applied in ligand-based VS, there are currently a few interesting examples for structure-based VS applications [128–130]. For instance, in AtomNet, the input of molecular complex is discretized to a 3D grid framework and directly fed into a convolutional neural network model [128]. Another similar work was created by Ragoza et al. (2016) where two independent classification tasks, i.e. activity and pose prediction, were trained and performed [130].

Multi-task learning is a class of machine learning approaches that learns a task together with other related tasks at the same time, with a shared representation. This can usually achieve a better model for the main task, because it allows the models to capture commonality among the tasks. Neural network models allow for the easy construction of multi-task classifiers and regression models, such as those for predicting binding activities against multiple targets at once. It has been shown that such QSAR models can perform better than single-task models [115, 121, 122, 131–133], because they can benefit from more training data, and share internal representations between tasks. In 2012, Merck & Co. hosted a Kaggle challenge where the ability of data science to improve predictive performance of QSAR methods was benchmarked. The winning team used multi-task deep networks, ensembled with other machine-learning techniques, to achieve a 15% relative improvement over the baseline method. The multi-task DNNs, which were called as "joint DNNs" in Ma et al. [115], can simultaneously model more than one molecular activity task. All tasks share the same input and hidden layers, but each task has its own output values [122, 132]. In 2017, the Pande group from Stanford University constructed a ligand-based virtual screening model through multi-task deep learning and established an excellent open source platform DeepChem [121]. In 2017, Xu et al. from Merck Pharmaceutical Company adopted multi-task deep learning to build ligand-based virtual screening models, and tried to analyze how it can improve the model performance [122].

12.5.5 Evaluation

12.5.5.1 Regression Models

In the Kaggle challenge organized by Merck in 2012, the correlation coefficient (r^2) was used to assess the performance of drug activity predictions. This metric is calculated as

$$r^2 = \frac{\left[\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})\right]^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2} \tag{12.3}$$

where y_i is the true activity, $\bar{y}$ is the mean of the true activity, $\hat{y}_i$ is the predicted activity, $\bar{\hat{y}}$ is the mean of the predicted activity, and n is the number of ligand molecules in the dataset. The larger the value of r^2, the better the prediction performance.

A common metric for evaluating regression models is the root mean square error (RMSE), given by

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{12.4}$$

where y_i and $\hat{y}_i$ are the true and predicted activity values, respectively, and n is the number of ligand molecules. The smaller the RMSE value, the better the prediction performance.

12.5.5.2 Classification Models

The overall prediction accuracy (Q), sensitivity (Sn), precision (P), specificity (Sp), and Matthew's correlation coefficient (CC) are commonly used for assessment of the classification system.

$$Q = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{12.5}$$

$$\text{Sn} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{12.6}$$

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{12.7}$$

$$\text{Sp} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{12.8}$$

$$\text{CC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TN} + \text{FN})}} \tag{12.9}$$

where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative, respectively.

The ROC curve is probably the most robust technique for evaluating classifiers and visualizing their performance. Classification machine learning models can

be validated by accuracy estimation techniques like the K-fold cross validation method, where the dataset is randomly partitioned into K subsets, and then K experiments are performed each respectively considering 1 subset for evaluation and the remaining $K - 1$ subsets for training the model.

12.6 Inverse Virtual Screening

The previously described procedure of virtual screening, in which a set of ligands is screened against a single protein target, can alternatively be described as "classical" or "forward" virtual screening. In "inverse" virtual screening (IVS), also known as "virtual target screening" or "target fishing," the roles assigned to the ligand and the protein are reversed: a single ligand of interest is screened against a set of proteins. While the underlying principles and methodologies of IVS are similar to typical virtual screening, a few challenges unique to IVS arise, which have been addressed in some unique ways.

12.6.1 Knowledge-Based Approaches

In order to perform an IVS, the most logically simple approach is to use methods that perform forward virtual screening but modify them so that the receptor is changed instead of the ligand. This means that most categories of forward virtual screening approaches (ligand-based, structure-based, etc.) also apply to IVS. Since many of these approaches determine interaction likelihood through searching of a database of experimentally observed binding events, such "knowledge-based" approaches have also populated the field of IVS.

One simple knowledge-based approach is the Similarity Ensemble Approach (SEA), which evaluates the biological similarity of targets by the chemical similarity of their respective ligands [134]. The chemical similarity is defined as the sum of Tanimoto Coefficients above a tuned threshold between all pairs of ligands, where each ligand is encoded by the Daylight fingerprint. Through this method, one can perform IVS by evaluating the chemical similarity of a query ligand against each target's respective ligand set. Accordingly, SEA has been applied in both retrospective and prospective drug-target prediction [135].

FINDSITE$^{comb2.0}$ is a unique approach that combines both protein structure and chemical similarity in order to predict protein–ligand interactions [79]. The method uses a combination of threading and structure comparison to identify ligand-bound protein structures (both experimental and modeled) from which a set of template ligands can be extracted. This set of template ligands is used to screen for active compounds in a compound library using chemical similarity. While the method is primarily developed and benchmarked as a forward

screening approach, it also can be used for IVS through evaluating the chemical similarity of the input ligand to the predicted active ligands for a given target.

Despite the relative speed and simplicity of these approaches, they are ultimately limited by the depth of the knowledgebase on which they are built. For example, a ligand-based approach cannot effectively screen a molecule that is too unlike any ligand in its database; such results would be based on spurious similarity and ultimately would not be trustworthy. Therefore, these methods are only as powerful as the databases on which they are built. This implies that they will only become more accurate over time as more data becomes available. However, there will always exist edge cases that are not properly addressed and documented by the databases of knowledge-based approaches, and so, methods that can make protein–ligand interaction predictions without explicit dependency on available data, such as docking, can potentially provide insight where none currently exists.

12.6.2 Docking Approaches

Given the popularity of protein–ligand docking approaches in forward virtual screening, it is no surprise that many methods for inverse virtual screening are also based on these methods. TarFisDock is one of the most simply constructed of these approaches [136]. It performs screening of a molecule of interest against a potential drug target database (PDTD) through a method based on DOCK 4.0 and returns a set of protein targets ranked by their predicted interaction energy. However, as noted by several other studies, while docking energy functions are reasonable to rank ligands given a single target, they are not necessarily comparable across targets. One way to address this problem is implemented in the first inverse docking program, INVDOCK, which requires that "hits" in the screen meet not only some minimum binding energy threshold, but also meet a threshold based on the target's binding affinity for its native ligand, ensuring that any predicted interactions would be competitive relative to the native interaction [137]. Another method of scoring targets is to implement an interaction fingerprinting technique, in which predicted interactions are evaluated based on the presence or absence of interactions between the ligand and each residue of the protein. Such a technique is implemented in IFPTarget [138], an approach that compares an interaction fingerprint generated from an AutoDock Vina docking result to a database of native interaction fingerprints and combines this comparison with a few more traditional scoring functions to identify likely ligand targets. This method can be seen as a hybrid between dependence on previously observed data and *a priori* docking predictions, where the docking scores can provide a prediction where no interaction is found and the fingerprint score can find near native binding modes that were predicted to be unfavorable by docking.

Since these approaches depend on protein–ligand docking, they inherit all of the shortcomings therein, such as the assumption of receptor rigidity, the lack of a scoring function that perfectly ranks docking results, the relative computational expense of docking, and the need for a high-resolution receptor structure with a well-defined binding pocket. This last point is particularly restrictive in IVS, as the majority of targets that one could logically screen against do not have solved structures, and even less have clearly defined binding pockets. In consideration of this problem, many docking-based IVS approaches only offer screening against a small database of proteins. If one wishes to use IVS to determine how a small molecule will impact a cell at a systems level, screening against such a limited database will not provide a complete view.

12.6.2.1 Applications of IVS

One of the most common applications of IVS is to increase the efficiency of computational drug discovery by assessing a molecule's ability to bind to proteins other than the intended therapeutic target. In these studies, the molecule of interest is typically identified through a previously performed forward virtual screen. When both forward and inverse virtual screens are used in combination, the forward screen can be viewed as a "sensitivity" screen in which a ligand that can tightly bind a given protein target is found, and the inverse virtual screen serves as a "specificity" screen which ensures that the ligand binds tightly only to the target protein and not to any other proteins within the biological context. Such off-target interactions can give rise to consequences other than the intended effect, such as side effects or compound toxicity, and therefore, prediction of these interactions can aid the efficiency of drug design studies by computationally identifying these problems before they are discovered *in vivo*. However, not all interactions other than the intended one are deleterious. In fact, the efficacy of a ligand may be enhanced through the interaction of several proteins at once, giving rise to the principle of "multi-target" design (a.k.a polypharmacology). Understanding the extent to which a compound will impact a biological context requires the application of systems biology models that take into account how the set of predicted binding events from an IVS result will perturb proteomic and metabolomic networks.

Another application of IVS is the discovery of therapeutic targets for a given active ligand. For example, there exist many drugs that are known to be clinically effective, yet their mechanism of action remains nebulous. Through IVS, one can identify a set of candidate proteins which might be the therapeutic target(s) of the drug molecule of interest. However, the mechanism of the drug need not be completely unknown for IVS to be helpful. In fact, IVS has also been shown to be effective for "drug repurposing," in which a clinically approved drug is used to treat some disease other than the one for which it was developed. Through

screening the drug to be repurposed against a library of potential therapeutic targets, high-affinity drug–target interactions can be discovered, which can lead to novel therapeutic action.

12.6.3 Challenges

While inverse virtual screening demonstrates promise for improving the drug design process through systems biology, a few challenges need to be overcome before the technique is widely accepted. One of the most pressing issues facing the field is the difficulty of constructing benchmark datasets due to publication bias. Since non-interacting protein–ligand pairs are frequently not reported, benchmarks lack confirmed non-interactions and instead rely on the assumption that if no interaction has been reported, it does not exist. Another challenge is the relative complexity of understanding of protein biochemistry relative to ligand chemistry. When constructing a forward virtual screen, one can choose a protein target that is sufficiently understood (i.e. solved protein structure, clearly defined binding pocket, a breadth of known binding partners, etc.), but in an IVS, no such luxury exists. If one is to gain a fully comprehensive view of how a molecule will impact all proteins within some biological context, all proteins therein must be addressed.

12.7 Conclusion

Hearkening back to its origins, virtual screening has come a long way and has evolved into a mélange of algorithms. Apart from the conventional ligand- and structure-based methods, the field has been burgeoning in recent years with machine learning-based approaches for the modeling of ligand bioactivity. Moreover, chemogenomics has been utilized in virtual screening to attempt to de-orphanize receptors, while inverse virtual screening is opening up new avenues for the prediction of off-target effects. In many ways, it feels as if the field has only just been born. Future developments are eagerly anticipated in the hopes that novel therapeutic compounds can be discovered for one of the largest protein families in human.

Acknowledgments

The study is supported in part by the National Institute of General Medical Sciences (GM070449, GM083107, GM116960), National Institute of Allergy and Infectious Diseases (AI134678), and the National Science Foundation (DBI1564756).

References

1 DeVree, B.T., Mahoney, J.P., Vélez-Ruiz, G.A. et al. (2016). Allosteric coupling from G protein to the agonist-binding pocket in GPCRs. *Nature* 535 (7610): 182.

2 Venter, J.C., Adams, M.D., Myers, E.W. et al. (2001). The sequence of the human genome. *Science* 291 (5507): 1304–1351.

3 O'Hayre, M., Vazquez-Prado, J., Kufareva, I. et al. (2013). The emerging mutational landscape of G proteins and G-protein-coupled receptors in cancer. *Nat. Rev. Cancer* 13 (6): 412–424.

4 Rompler, H., Staubert, C., Thor, D. et al. (2007). G Protein-coupled time travel: evolutionary aspects of GPCR research. *Mol. Interv.* 7 (1): 17–25.

5 Garland, S.L. (2013). Are GPCRs still a source of new targets? *J. Biomol. Screen.* 18 (9): 947–966.

6 Van Drie, J.H. (2007). Computer-aided drug design: the next 20 years. *J. Comput. Aided Mol. Des.* 21 (10, 11): 591–601.

7 O'Boyle, N.M. (2012). Towards a Universal SMILES representation – a standard method to generate canonical SMILES based on the InChI. *J. Cheminf.* 4 (1): 22.

8 Heller, S., McNaught, A., Stein, S. et al. (2013). InChI – the worldwide chemical structure identifier standard. *J. Cheminf.* 5 (1): 7.

9 Pletnev, I., Erin, A., McNaught, A. et al. (2012). InChIKey collision resistance: an experimental testing. *J. Cheminf.* 4 (1): 39.

10 Durant, J.L., Leland, B.A., Henry, D.R. et al. (2002). Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* 42 (6): 1273–1280.

11 O'Boyle, N.M. and Sayle, R.A. (2016). Comparing structural fingerprints using a literature-based similarity benchmark. *J. Cheminf.* 8 (1): 36.

12 Dalby, A., Nourse, J.G., Hounshell, W.D. et al. (1992). Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* 32 (3): 244–255.

13 Berman, H.M. (2008). The protein data bank: a historical perspective. *Acta Crystallogr. A* 64 (1): 88–95.

14 Consortium U (2016). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45 (D1): D158–D169.

15 Rose, P.W., Prlic, A., Altunkaya, A. et al. (2017). The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* 45 (D1): D271–D281.

16 Prilusky, J. (1996). OCA, a browser-database for protein structure/function. http://oca.weizmann.ac.il/oca-bin/ocamain.

17 Pándy-Szekeres, G., Munk, C., Tsonkov, T.M. et al. (2017). GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res.* 46 (D1): D440–D446.

18 Gaulton, A., Bellis, L.J., Bento, A.P. et al. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40 (Database issue): D1100–D1107.

19 Gilson, M.K., Liu, T., Baitaluk, M. et al. (2015). BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* 44 (D1): D1045–D1053.

20 Wassermann, A.M. and Bajorath, J. (2011). BindingDB and ChEMBL: online compound databases for drug discovery. *Expert Opin. Drug Discov.* 6 (7): 683–687.

21 Law, V., Knox, C., Djoumbou, Y. et al. (2014). DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 42 (Database issue): D1091–D1097.

22 Roth, B.L., Lopez, E., Patel, S. et al. (2000). The multiplicity of serotonin receptors: uselessly diverse molecules or an embarrassment of riches? *Neuroscientist* 6 (4): 252–262.

23 Alexander, S.P., Davenport, A.P., Kelly, E. et al. (2015). The Concise Guide to PHARMACOLOGY 2015/16: G protein-coupled receptors. *Br. J. Pharmacol.* 172 (24): 5744–5869.

24 Wang, Y., Xiao, J., Suzek, T.O. et al. (2009). PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* 37 (Web Server issue): W623–W633.

25 Okuno, Y., Yang, J., Taneishi, K. et al. (2006). GLIDA: GPCR-ligand database for chemical genomic drug discovery. *Nucleic Acids Res.* 34 (Suppl 1): D673–D677.

26 Okuno, Y., Tamon, A., Yabuuchi, H. et al. (2008). GLIDA: GPCR–ligand database for chemical genomics drug discovery – database and tools update. *Nucleic Acids Res.* 36 (Database issue): D907–D912.

27 Chan, W.K., Zhang, H., Yang, J. et al. (2015). GLASS: a comprehensive database for experimentally validated GPCR-ligand associations. *Bioinformatics* 31 (18): 3035–3042.

28 Zoete, V., Daina, A., Bovigny, C. et al. (2016). SwissSimilarity: a web tool for low to ultra high throughput ligand-based virtual screening. *J. Chem. Inf. Model.* 56 (8): 1399–1404.

29 Wu, J., Zhang, Q., Wu, W. et al. (2018). WDL-RF: predicting bioactivities of ligand molecules acting with G protein-coupled receptors by combining weighted deep learning and random forest. *Bioinformatics* 1: 12.

30 Huang, N., Shoichet, B.K., and Irwin, J.J. (2006). Benchmarking sets for molecular docking. *J. Med. Chem.* 49 (23): 6789–6801.

31 Mysinger, M.M., Carchia, M., Irwin, J.J. et al. (2012). Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* 55 (14): 6582–6594.

32 Weiss, D.R., Bortolato, A., Tehan, B. et al. (2016). GPCR-bench: a benchmarking set and practitioners' guide for G protein-coupled receptor docking. *J. Chem. Inf. Model.* 56 (4): 642–651.

33 Gatica, E.A. and Cavasotto, C.N. (2011). Ligand and decoy sets for docking to G protein-coupled receptors. *J. Chem. Inf. Model.* 52 (1): 1–6.

34 Sastry, G.M., Inakollu, V.S., and Sherman, W. (2013). Boosting virtual screening enrichments with data fusion: coalescing hits from two-dimensional fingerprints, shape, and docking. *J. Chem. Inf. Model.* 53 (7): 1531–1542.

35 Truchon, J.-F. and Bayly, C.I. (2007). Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. *J. Chem. Inf. Model.* 47 (2): 488–508.

36 Bemis, G.W. and Murcko, M.A. (1996). The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* 39 (15): 2887–2893.

37 Bajusz, D., Rácz, A., and Héberger, K. (2015). Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminf.* 7 (1): 20.

38 O'Boyle, N.M., Banck, M., James, C.A. et al. (2011). Open Babel: an open chemical toolbox. *J. Cheminf.* 3 (1): 33.

39 Landrum, G. (2012). RDKit: open-source cheminformatics. http://www rdkit.org (accessed 07 March 2022).

40 Grant, J.A. and Pickup, B. (1995). A Gaussian description of molecular shape. *J. Phys. Chem.* 99 (11): 3503–3510.

41 Hawkins, P.C.D., Skillman, A.G., and Nicholls, A. (2007). Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* 50 (1): 74–82.

42 Ballester, P.J. (2011). Ultrafast shape recognition: method and applications. *Future Med. Chem.* 3 (1): 65–78.

43 Roy, A. and Skolnick, J. (2014). LIGSIFT: an open-source tool for ligand structural alignment and virtual screening. *Bioinformatics* 31 (4): 539–544.

44 Hu, J., Liu, Z., Yu, D.-J. et al. (2018). LS-align: an atom-level, flexible ligand structural alignment algorithm for high-throughput virtual screening. *Bioinformatics* 34 (13): 2209–2218.

45 Morris, G.M., Huey, R., Lindstrom, W. et al. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30 (16): 2785–2791.

46 Jones, G., Willett, P., Glen, R.C. et al. (1997). Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* 267 (3): 727–748.

47 Trott, O. and Olson, A.J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31 (2): 455–461.

48 Korb, O., Stutzle, T., and Exner, T.E. (2009). Empirical scoring functions for advanced protein–ligand docking with PLANTS. *J. Chem. Inf. Model.* 49 (1): 84–96.

49 Allen, W.J., Balius, T.E., Mukherjee, S. et al. (2015). DOCK 6: impact of new features and current docking performance. *J. Comput. Chem.* 36 (15): 1132–1156.

50 Friesner, R.A., Banks, J.L., Murphy, R.B. et al. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47 (7): 1739–1749.

51 Halgren, T.A., Murphy, R.B., Friesner, R.A. et al. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* 47 (7): 1750–1759.

52 Meng, E.C., Shoichet, B.K., and Kuntz, I.D. (1992). Automated docking with grid-based energy evaluation. *J. Comput. Chem.* 13 (4): 505–524.

53 Huang, S.-Y., Grinter, S.Z., and Zou, X. (2010). Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. *Phys. Chem. Chem. Phys.* 12 (40): 12899–12908.

54 Liu, J. and Wang, R. (2015). Classification of current scoring functions. *J. Chem. Inf. Model.* 55 (3): 475–482.

55 Kuntz, I.D., Blaney, J.M., Oatley, S.J. et al. (1982). A geometric approach to macromolecule–ligand interactions. *J. Mol. Biol.* 161 (2): 269–288.

56 Basith, S., Cui, M., Macalino, S.J. et al. (2018). Exploring G protein-coupled receptors (GPCRs) ligand space via cheminformatics approaches: impact on rational drug design. *Front. Pharmacol.* 9: 128.

57 Yuan, X. and Xu, Y. (2018). Recent trends and applications of molecular modeling in GPCR–ligand recognition and structure-based drug design. *Int. J. Mol. Sci.* 19 (7): 2105.

58 Roth, B.L., Irwin, J.J., and Shoichet, B.K. (2017). Discovery of new GPCR ligands to illuminate new biology. *Nat. Chem. Biol.* 13 (11): 1143.

59 Manglik, A., Lin, H., Aryal, D.K. et al. (2016). Structure-based discovery of opioid analgesics with reduced side effects. *Nature* 537 (7619): 185.

60 Rodriguez, D., Brea, J., Loza, M.I. et al. (2014). Structure-based discovery of selective serotonin 5-HT(1B) receptor ligands. *Structure* 22 (8): 1140–1151.

61 Mysinger, M.M., Weiss, D.R., Ziarek, J.J. et al. (2012). Structure-based ligand discovery for the protein-protein interface of chemokine receptor CXCR4. *Proc. Natl. Acad. Sci. U.S.A.* 109 (14): 5517–5522.

62 Becker, O.M., Dhanoa, D.S., Marantz, Y. et al. (2006). An integrated in silico 3D model-driven discovery of a novel, potent, and selective amido-sulfonamide 5-HT1A agonist (PRX-00023) for the treatment of anxiety and depression. *J. Med. Chem.* 49 (11): 3116–3135.

63 Kirchhoff, V.D., Nguyen, H.T., Soczynska, J.K. et al. (2009). Discontinued psychiatric drugs in 2008. *Exp. Opin. Investig. Drugs* 18 (10): 1431–1443.

64 Saha, A.K., Becker, O.M., Noiman, S., et al. (eds.) (2006). PRX-03140: the discovery and development of a novel 5HT4 partial agonist for the treatment of Alzheimer's disease. Abstracts of Papers of the American Chemical Society.

65 Langmead, C.J., Andrews, S.P., Congreve, M. et al. (2012). Identification of novel adenosine A2A receptor antagonists by virtual screening. *J. Med. Chem.* 55 (5): 1904–1909.

66 Verma, S., Kumar, A., Tripathi, T. et al. (2018). Muscarinic and nicotinic acetylcholine receptor agonists: current scenario in Alzheimer's disease therapy. *J. Pharm. Pharmacol.* 70 (8): 985–993.

67 Sliwoski, G., Kothiwale, S., Meiler, J. et al. (2014). Computational methods in drug discovery. *Pharmacol. Rev.* 66 (1): 334–395.

68 Drwal, M.N. and Griffith, R. (2013). Combination of ligand- and structure-based methods in virtual screening. *Drug Discov. Today Technol.* 10 (3): e395–e401.

69 Khan, K.M., Wadood, A., Ali, M. et al. (2010). Identification of potent urease inhibitors via ligand-and structure-based virtual screening and in vitro assays. *J. Mol. Graphics Modell.* 28 (8): 792–798.

70 Weidlich, I.E., Dexheimer, T., Marchand, C. et al. (2010). Inhibitors of human tyrosyl-DNA phospodiesterase (hTdp1) developed by virtual screening using ligand-based pharmacophores. *Bioorg. Med. Chem.* 18 (1): 182–189.

71 Banoglu, E., Çalışkan, B., Luderer, S. et al. (2012). Identification of novel benzimidazole derivatives as inhibitors of leukotriene biosynthesis by virtual screening targeting 5-lipoxygenase-activating protein (FLAP). *Bioorg. Med. Chem.* 20 (12): 3728–3741.

72 Svensson, F., Karlén, A., and Sköld, C. (2011). Virtual screening data fusion using both structure- and ligand-based methods. *J. Chem. Inf. Model.* 52 (1): 225–232.

73 Swann, S.L., Brown, S.P., Muchmore, S.W. et al. (2011). A unified, probabilistic framework for structure- and ligand-based virtual screening. *J. Med. Chem.* 54 (5): 1223–1232.

74 Tan, L., Geppert, H., Sisay, M.T. et al. (2008). Integrating structure- and ligand-based virtual screening: comparison of individual, parallel, and fused molecular docking and similarity search calculations on multiple targets. *ChemMedChem* 3 (10): 1566–1571.

75 Klabunde, T. (2007). Chemogenomic approaches to drug discovery: similar receptors bind similar ligands. *Br. J. Pharmacol.* 152 (1): 5–7.

76 Brylinski, M. and Skolnick, J. (2008). A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proc. Natl. Acad. Sci. U.S.A.* 105 (1): 129–134.

77 Zhou, H.Y. and Skolnick, J. (2012). FINDSITEx: a structure-based, small molecule virtual screening approach with application to all identified human GPCRs. *Mol. Pharm.* 9 (6): 1775–1784.

78 Zhou, H. and Skolnick, J. (2012). FINDSITEcomb: a threading/structure-based, proteomic-scale virtual ligand screening approach. *J. Chem. Inf. Model.* 53 (1): 230–240.

79 Zhou, H., Cao, H., and Skolnick, J. (2018). FINDSITE$^{comb2.0}$: a new approach for virtual ligand screening of proteins and virtual target screening of biomolecules. *J. Chem. Inf. Model.* 58 (11): 2343–2354.

80 Roy, A., Srinivasan, B., and Skolnick, J. (2015). PoLi: a virtual screening pipeline based on template pocket and ligand similarity. *J. Chem. Inf. Model.* 55 (8): 1757–1770.

81 Yang, Y., Zhan, J., and Zhou, Y. (2016). SPOT-ligand: fast and effective structure-based virtual screening by binding homology search according to ligand and receptor similarity. *J. Comput. Chem.* 37 (18): 1734–1739.

82 Litfin, T., Zhou, Y., and Yang, Y. (2017). SPOT-ligand 2: improving structure-based virtual screening by binding-homology search on an expanded structural template library. *Bioinformatics* 33 (8): 1238–1240.

83 Chan, W.K. and Zhang, Y. (2020). Virtual screening of human class – a GPCRs using ligand profiles built on multiple ligand–receptor interactions. *J. Mol. Biol.* 432 (17): 4872–4890.

84 Mitchell, T., Buchanan, B., DeJong, G. et al. (1990). Machine learning. *Ann. Rev. Comput. Sci.* 4 (1): 417–433.

85 Wu, J.-S., Liu, B., Chan, W.K.B. et al. (2019). Precise modelling and interpretation of bioactivities of ligands targeting G protein-coupled receptors. *Bioinformatics* 35: i324–i332.

86 Hansch, C. and Fujita, T. (1964). *p*-σ-π Analysis. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* 86 (8): 1616–1626.

87 Cramer, R.D., Patterson, D.E., and Bunce, J.D. (1988). Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* 110 (18): 5959–5967.

88 Gedeck, P., Rohde, B., and Bartels, C. (2006). QSAR – how good is it in practice? Comparison of descriptor sets on an unbiased cross section of corporate data sets. *J. Chem. Inf. Model.* 46 (5): 1924–1936.

89 Evers, A., Hessler, G., Matter, H. et al. (2005). Virtual screening of biogenic amine-binding G-protein coupled receptors: comparative evaluation of protein- and ligand-based virtual screening protocols. *J. Med. Chem.* 48 (17): 5448–5465.

90 Klebe, G., Abraham, U., and Mietzner, T. (1994). Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* 37 (24): 4130–4146.

91 Baber, J.C., Shirley, W.A., Gao, Y. et al. (2006). The use of consensus scoring in ligand-based virtual screening. *J. Chem. Inf. Model.* 46 (1): 277–288.

92 Cherkasov, A., Ban, F., Li, Y. et al. (2006). Progressive docking: a hybrid QSAR/docking approach for accelerating in silico high throughput screening. *J. Med. Chem.* 49 (25): 7466–7478.

93 Jacobsson, M. and Karlén, A. (2006). Ligand bias of scoring functions in structure-based virtual screening. *J. Chem. Inf. Model.* 46 (3): 1334–1343.

94 Oloff, S., Zhang, S., Sukumar, N. et al. (2006). Chemometric analysis of ligand receptor complementarity: identifying Complementary Ligands Based on Receptor Information (CoLiBRI). *J. Chem. Inf. Model.* 46 (2): 844–851.

95 Zhang, S., Golbraikh, A., and Tropsha, A. (2006). Development of quantitative structure-binding affinity relationship models based on novel geometrical chemical descriptors of the protein–ligand interfaces. *J. Med. Chem.* 49 (9): 2713–2724.

96 Miller, D.W. (2001). Results of a new classification algorithm combining K nearest neighbors and recursive partitioning. *J. Chem. Inf. Comput. Sci.* 41 (1): 168–175.

97 Guha, R., Dutta, D., Jurs, P.C. et al. (2006). Local lazy regression: making use of the neighborhood to improve QSAR predictions. *J. Chem. Inf. Model.* 46 (4): 1836–1847.

98 Zhang, S., Golbraikh, A., Oloff, S. et al. (2006). A novel automated lazy learning QSAR (ALL-QSAR) approach: method development, applications, and virtual screening of chemical databases using validated ALL-QSAR models. *J. Chem. Inf. Model.* 46 (5): 1984–1995.

99 Labute, P. (1999). Binary QSAR: a new method for the determination of quantitative structure activity relationships. *Pac. Symp. Biocomput.* 444–455.

100 Klon, A.E., Glick, M., Thoma, M. et al. (2004). Finding more needles in the haystack: a simple and efficient method for improving high-throughput docking results. *J. Med. Chem.* 47 (11): 2743–2749.

101 Klon, A.E., Glick, M., and Davies, J.W. (2004). Application of machine learning to improve the results of high-throughput docking against the HIV-1 protease. *J. Chem. Inf. Comput. Sci.* 44 (6): 2216–2224.

102 Glick, M., Klon, A.E., Acklin, P. et al. (2004). Enrichment of extremely noisy high-throughput screening data using a naive Bayes classifier. *J. Biomol. Screen.* 9 (1): 32–36.

103 Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20 (3): 273–297.

104 Wilton, D., Willett, P., Lawson, K. et al. (2003). Comparison of ranking methods for virtual screening in lead-discovery programs. *J. Chem. Inf. Comput. Sci.* 43 (2): 469–474.

105 Wilton, D.J., Harrison, R.F., Willett, P. et al. (2006). Virtual screening using binary kernel discrimination: analysis of pesticide data. *J. Chem. Inf. Model.* 46 (2): 471–477.

106 Franke, L., Byvatov, E., Werz, O. et al. (2005). Extraction and visualization of potential pharmacophore points using support vector machines: application to ligand-based virtual screening for COX-2 inhibitors. *J. Med. Chem.* 48 (22): 6997–7004.

107 Lepp, Z., Kinoshita, T., and Chuman, H. (2006). Screening for new antidepressant leads of multiple activities by support vector machines. *J. Chem. Inf. Model.* 46 (1): 158–167.

108 Byvatov, E., Sasse, B.C., Stark, H. et al. (2005). From virtual to real screening for D3 dopamine receptor ligands. *ChemBioChem* 6 (6): 997–999.

109 Betzi, S., Suhre, K., Chétrit, B. et al. (2006). GFscore: a general nonlinear consensus scoring function for high-throughput docking. *J. Chem. Inf. Model.* 46 (4): 1704–1712.

110 Bazeley, P.S., Prithivi, S., Struble, C.A. et al. (2006). Synergistic use of compound properties and docking scores in neural network modeling of CYP2D6 binding: predicting affinity and conformational sampling. *J. Chem. Inf. Model.* 46 (6): 2698–2708.

111 Omata, K., Kobayashi, Y., and Yamada, M. (2007). Artificial neural network aided virtual screening of additives to a $Co/SrCO_3$ catalyst for preferential oxidation of CO in excess hydrogen. *Catal. Commun.* 8 (1): 1–5.

112 Jones-Hertzog, D.K., Mukhopadhyay, P., Keefer, C.E. et al. (1999). Use of recursive partitioning in the sequential screening of G-protein–coupled receptors. *J. Pharmacol. Toxicol. Methods* 42 (4): 207–215.

113 Yamazaki, K., Kusunose, N., Fujita, K. et al. (2006). Identification of phosphodiesterase-1 and 5 dual inhibitors by a ligand-based virtual screening optimized for lead evolution. *Bioorg. Med. Chem. Lett.* 16 (5): 1371–1379.

114 van Rhee, A.M. (2003). Use of recursion forests in the sequential screening process: consensus selection by multiple recursion trees. *J. Chem. Inf. Comput. Sci.* 43 (3): 941–948.

115 Ma, J., Sheridan, R.P., Liaw, A. et al. (2015). Deep neural nets as a method for quantitative structure–activity relationships. *J. Chem. Inf. Model.* 55 (2): 263–274.

116 Svetnik, V., Liaw, A., Tong, C. et al. (2003). Random forest: a classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* 43 (6): 1947–1958.

117 Teramoto, R. and Fukunishi, H. (2007). Supervised consensus scoring for docking and virtual screening. *J. Chem. Inf. Model.* 47 (2): 526–534.

118 Wallach, I., Dzamba, M., and Heifets, A. (2015). AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *Math. Z.* 47 (1): 34–46.

119 Winkler, D.A. and Le, T.C. (2017). Performance of deep and shallow neural networks, the universal approximation theorem, activity cliffs, and QSAR. *Mol. Inform.* 36 (1, 2).

120 Unterthiner, T., Mayr, A., Klambauer, G., et al. (eds.) (2014). Deep learning as an opportunity in virtual screening. *Proceedings of the Deep Learning and Representation Learning Workshop (NIPS 2014)*, Los Angeles, USA.

121 Ramsundar, B., Liu, B., Wu, Z. et al. (2017). Is multitask deep learning practical for pharma? *J. Chem. Inf. Model.* 57 (8): 2068–2076.

122 Xu, Y., Ma, J., Liaw, A. et al. (2017). Demystifying multitask deep neural networks for quantitative structure–activity relationships. *J. Chem. Inf. Model.* 57 (10): 2490–2504.

123 Wu, J., Zhang, Q., Wu, W. et al. (2018). WDL-RF: predicting bioactivities of ligand molecules acting with G protein-coupled receptors by combining weighted deep learning and random forest. *Bioinformatics* 34 (13): 2271–2282.

124 Ertl, P., Lewis, R., Martin, E. et al. (2017). In silico generation of novel, drug-like chemical matter using the LSTM neural network. *arXiv* preprint arXiv:171207449 .

125 Gómez-Bombarelli, R., Wei, J.N., Duvenaud, D. et al. (2018). Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* 4 (2): 268–276.

126 Olivecrona, M., Blaschke, T., Engkvist, O. et al. (2017). Molecular de-novo design through deep reinforcement learning. *J. Cheminf.* 9 (1): 48.

127 Segler, M.H., Kogej, T., Tyrchan, C. et al. (2017). Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent. Sci.* 4 (1): 120–131.

128 Wallach, I., Dzamba, M., and Heifets, A. (2015). AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv* preprint arXiv:151002855 .

129 Stepniewska-Dziubinska, M.M., Zielenkiewicz, P., and Siedlecki, P. (2018). Development and evaluation of a deep learning model for protein–ligand binding affinity prediction. *Bioinformatics* 34 (21): 3666–3674.

130 Ragoza, M., Hochuli, J., Idrobo, E. et al. (2017). Protein–ligand scoring with convolutional neural networks. *J. Chem. Inf. Model.* 57 (4): 942–957.

131 Rosenbaum, L., Dörr, A., Bauer, M.R. et al. (2013). Inferring multi-target QSAR models with taxonomy-based multi-task learning. *J. Cheminf.* 5 (1): 33.

132 Dahl, G.E., Jaitly, N., and Salakhutdinov, R. (2014). Multi-task neural networks for QSAR predictions. *arXiv* preprint arXiv:14061231 .

133 Ramsundar, B., Kearnes, S., Riley, P. et al. (2015). Massively multitask networks for drug discovery. *arXiv* preprint arXiv:150202072 .

134 Keiser, M.J., Roth, B.L., Armbruster, B.N. et al. (2007). Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* 25 (2): 197–206.

135 Keiser, M.J., Setola, V., Irwin, J.J. et al. (2009). Predicting new molecular targets for known drugs. *Nature* 462 (7270): 175–181.

136 Li, H., Gao, Z., Kang, L. et al. (2006). TarFisDock: a web server for identifying drug targets with docking approach. *Nucleic Acids Res.* 34 (Web Server issue): W219–W224.

137 Chen, Y.Z. and Zhi, D.G. (2001). Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins* 43 (2): 217–226.

138 Li, G.-B., Yu, Z.-J., Liu, S. et al. (2017). IFPTarget: a customized virtual target identification method based on protein–ligand interaction fingerprinting analyses. *J. Chem. Inf. Model.* 57 (7): 1640–1651.