Pharmacophore Modelling and Screening: Concepts, Recent Developments and Applications in Rational Drug Design

Chinmayee Choudhury and G. Narahari Sastry

Abstract Computational design of molecules with desired properties has become indispensable in many areas of research, particularly in the pharmaceutical industry and academia. Pharmacophore is one of the essential state-of-the-art techniques widely used in various ways in the computer-aided drug design projects. The pharmacophore modelling approaches have been an important part of many drug discovery strategies due to its simple yet diverse usage. It has been extensively applied for virtual screening, lead optimization, target identification, toxicity prediction and de novo lead design and has a huge scope for application in fragment-based drug design and lead design targeting protein–protein interaction interfaces and target-based classification of chemical space. In this chapter, we have briefly discussed the basic concepts and methods of generation of pharmacophore models. The diverse applications of the pharmacophore approaches have been discussed using number of case studies. We conclude with the limitations of the approaches and its wide scope for the future application depending on the research problem and the type of initial data available.

Keywords Computer-aided drug design · Pharmacophore mapping
Receptor-based pharmacophore · Ligand-based pharmacophore
Pharmacophore features · Pharmacophore fingerprints · Virtual screening
Pharmacophore searching · Docking · QSAR · De novo design

Abbreviations

ADMET	Absorption, distribution, metabolism, excretion, toxicity
CADD	Computer-aided drug design
CmaA1	Mycobacterial cyclopropane synthase

C. Choudhury · G. Narahari Sastry (✉)
Center for Molecular Modelling, Indian Institute of Chemical Technology, Hyderabad, India
e-mail: gnsastry@gmail.com

C. Choudhury
Department of Biochemistry, All India Institute of Medical Sciences, Basni, Jodhpur, Rajasthan, India

C. G. Mohan (ed.), *Structural Bioinformatics: Applications in Preclinical Drug Discovery Process*, Challenges and Advances in Computational Chemistry and Physics 27, https://doi.org/10.1007/978-3-030-05282-9_2

HHCPF	Hexadecahydro-1H-Cyclopenta[a]Phenanthrene Framework
HTS	High-throughput screening
MD	Molecular dynamics
Mtb	Mycobacterium tuberculosis
QSAR	Quantitative structure-activity relationship
TB	Tuberculosis

1 Introduction

Rational drug discovery is highly interdisciplinary and is one of the outstanding challenges, besides being highly arduous and expensive. The process of designing new medications requires investment of roughly 14 years [1] of time and cost as high as 1 billion USD [2]. Along with rapidly evolving HTS [3] and combinatorial chemistry technologies, computer-aided drug design (CADD) strategies are also effectively contributing to accelerate and economize the process of drug development [4–6]. A broad range of CADD applications are employed at almost all early stages of the drug discovery pipelines, starting from target identification, target structure prediction, screening of initial hits to prioritization and optimization of leads and understanding their structure-property relationships [7, 8]. We have been working in state-of-the-art CADD techniques such as homology modelling [9], molecular dynamics simulations [10–12], QSAR [13–15], molecular docking [16], pharmacophore modelling [17], virtual screening [18, 19] and cheminformatics [20] since more than a decade. One of the fundamental applications of cheminformatics is to develop programmes that store, manage and retrieve molecular structures in various formats, their calculated/experimental properties and bioactivities. Cheminformatics also involves computing molecular fingerprints and descriptors based on the molecular structures that label a physicochemical property and can be used as screening filters [21, 22]. These molecular descriptors of known active molecules can also be used to develop quantitative structure-activity/property relationship (QSAR/QSPR) models to predict the inhibitory activity or toxicity of novel compounds and preliminarily profile them in silico without performing expensive in vitro and in vivo assays [23–26]. Docking and simulations predict the three-dimensional binding mode of a given molecule in the binding site of a macromolecular receptor (protein/DNA), and their affinity is quantitatively assessed by a docking score. This technique has not only been proved enormously useful to study receptor–ligand interactions but also is used as a popular tool to virtually screen compound libraries to obtain a hit or to identify the target for a molecule by reverse engineering [27–29]. A large number of studies from our group have focused on application of these techniques to a plethora of drug targets such as phosphodiesterases [14], kinases [12, 30], HIV proteases [10, 13] and reverse transcriptase [31] and Mtb cyclopropane synthases [11, 17, 18]. We have also

initiated development of a disease (tuberculosis) specific Web portal, integrating all these techniques, which will be of tremendous help for researches working in the field of Mtb drug discovery [32].

Pharmacophore modelling is one of the enormously useful sub-areas of CADD with diverse structure and ligand-based applications [33, 34]. Like docking, one of the basic applications of pharmacophore models is virtual screening, but at a much faster speed as compared to docking [33]. This approach can also be implemented complementarily with docking and QSAR studies [18, 20]. Many studies use pharmacophore models for target/off-target identification as well [35, 36]. In this chapter, we basically focus on the in silico representation of the concept and the varieties of ways of application of pharmacophore models in drug discovery projects.

2 The Concept of Pharmacophore

The term 'pharmacophore' has gained immense popularity in the field of medicinal chemistry paralleled with computer-aided structure-activity relationship studies. In 1909, Ehrlich gave an introductory definition of pharmacophore [37, 38], by combining the words 'phoros' meaning carrying and 'pharmacon' meaning drug. Hence, a pharmacophore is 'the molecular framework carrying the crucial features accountable for a drug's biological activity'. Since then, many groups have attributed various definitions and meanings to this term based on their scientific background and research view. IUPAC has officially defined a pharmacophore model as [39]

> An ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response.

However, a century's research and development has expanded its circumstantial meaning and application considerably. Due to their simple way of capturing and representing the chemical features of compounds, pharmacophore models have drawn the attention of the medicinal chemistry community in last few years as a tool to screen the cig (chemistry) data [40]. Upon administration, when a drug/small molecule enters the human body, it comes across thousands of proteins (receptors, transporters, carriers, plasma proteins, etc.) to potentially interact with. But it chooses to bind to only those proteins (targets) where the protein's active site and drug have compatible shape/size and the protein–drug interactions are energetically favourable. Similarly, size/volume/shape and the chemical features of the residues lining the binding pocket determine which type of small molecules it is able to bind. Hence, the right size, correct shape and complementary chemical features are the key factors for the protein–drug recognition to instigate a biological effect. The central concept of pharmacophore is based on the perception that the molecular interaction pattern of a group of compounds with their biological target can be credited to a small set of common features complementary to the chemical features

present in the target's binding pocket. The general features include hydrogen-bond (HB) donors, HB acceptors, charged groups (positive and negative), hydrophobic sites and aromatic rings, which are used as chemical features in pharmacophore models by most of the programmes. Some programmes define a few additional features such as 'exclusion volumes' representing steric constraints. These features generally replicate the steric environment of the binding pocket to avoid clashes of the mapped of compounds with the protein surface. Pharmacophore models comprises distinct spatial arrangement of these features that denotes the chemical functionalities of active small molecules. Instead of real atoms/functional groups, a pharmacophore model emphasizes the chemical features of ligands/protein–ligand complexes, making it a better and fast tool to recognize molecular similarities.

3 A Typical Pharmacophore Model: Representation of Pharmacophoric Features

According to the definition, a pharmacophore model represents the binding patterns of bioactive molecules with the target binding site, by virtue of a distinct 3D arrangement of abstract interaction features accounting for different types of non-covalent interactions. These interaction types can be HB formation, columbic interactions, metal interactions, hydrophobic contacts, aromatic stacking or charge transfer interactions. Overall, a pharmacophore model characterizes a common binding mode of diverse ligands with a specific target. In pharmacophore modelling, the molecules are first segregated into a set of features, each representing a certain type of interaction with the binding site residues. Then, each feature is represented by points to be used for superimposition (least-squares fitting) of molecules with each other. Here we will be discussing features employed by most of the popular programmes [41–45].

HB donor (D): Hydroxyl groups, hydrogens bound to nitrogen, acetylenic CH groups and thiols (SH) are normally denoted as donors. However, the –CH and –SH groups are considered relatively weaker donors. Sometimes, along with acetylenes, other types of –CH such as the ones in nitrogen heterocycles of some kinase inhibitors are considered as donors. Keeping protonation in mind, basic amines such as $RCH_2N(Me)_2$ are considered as donors. Tautomeric and ionized states severely influence pharmacophore feature definition because they may amend the characteristic of a feature. Hence, molecules should be presented to the pharmacophore elucidation programmes in all possible protonation/ionization states.

HB acceptor (A): Generally, atoms with available lone pairs of electrons such as N, O, S are treated as acceptors. However, some programmes do not consider oxygen atoms present in furan/oxazole rings, as they are very weak acceptors according to theoretical and crystallographic evidence.

Along with defining the HB features, it is very essential to fix the positions of the complementary feature points to be overlapped in the resulting pharmacophore. That is why the pharmacophore modelling programmes link donor and acceptor features with the equivalent ligand atoms as well as the supposed locations of the corresponding complementary receptor atoms involved in the interaction.

Positive and negative features (P and N): In the molecules, atoms bearing formal charges are considered as positive or negative features provided they are not part of a dipole. Groups possessing net formal charges are also considered as positive/negative features. Centroid of the heteroatoms of a group is the region, where the positive/negative charged features are generally placed. Sometimes the positive and negative features are emphasized specifically based on their ionizability. For example, $R–NH_3^+$ is measured as positively ionizable feature, but $R–N(Me)_3^+$ is not as the interactions made by these two groups are significantly different.

Hydrophobic features (H): Choosing atoms/groups that should be measured as hydrophobic is neither easy nor straightforward. The most commonly used algorithm developed by Greene et al. [42] first allot a hydrophobicity score to each atom based on a set of empirical rules defined from medicinal chemists' perceptions and then atoms with amply large hydrophobicity values are grouped into clusters. Then a hydrophobic feature point is placed at the centroid of each such cluster. The order of hydrophobicity score is roughly rings/ring atoms > groups like $–CF_3$ > alkyl chains. Some simple algorithms [44] consider all non-donors/non-acceptor/non-charged atoms as steric groups (equivalent of hydrophobic groups), which also yield a depiction of molecular shape.

Aromatic rings (R): Aromatic rings are treated as a special type of hydrophobic feature represented by vectors instead of points so as to mimic the directionality of interactions like π–π stacking and cation–π interactions. Figure 1 shows an example of a typical pharmacophore model.

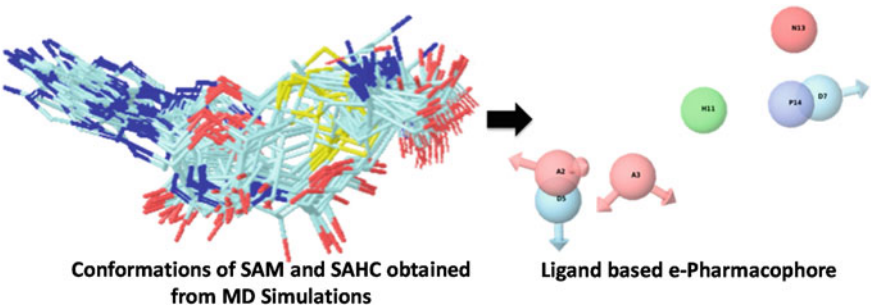

Fig. 1 An example of a pharmacophore model, generated from the conformations of S-adenosyl methionine (SAM) and S-adenosyl homocysteine (SAHC) [17] with Phase programme. Colour codes for the pharmacophoric features are as follows. Cyan: D, pink: A, red: N, blue: P, green: H and orange: R

Most recent pharmacophore modelling programmes define additional steric constraints features. These are called exclusion volumes (XVols), representing the steric effect of the binding pocket [46]. These features are required to avoid the clashes of the molecule with the protein surface while mapping. Feature generation not only facilitates the molecules to be aligned in an easy and rational way, but also can be used in scoring. The root mean square deviation (RMSD) between matched features gives quantitative account of the extent of overlay, which is often used as a fitness score [40]. Hence, the placement of feature points should be accurate, and one needs to be careful while deciding whether to consider all possible features or to choose few of them giving adequate information about the spatial orientation of a group of molecules. For example, sometimes there are huge number of hydrophobic features as compared to other features, which may bias the alignment and give a model with good score, but the model will be useless due to lack of specificity.

4 Evolution of the 'Pharmacophore' Concept: Historical Perspective

Paul Ehrlich first used the concept of pharmacophore in the end of nineteenth century, when he revealed the selective binding of methylene blue to nerve fibres. This realization ushered the beginning of pharmacophore concept as '*a molecular framework that carries (phoros) the essential features responsible for a drug's (pharmacon) biological activity*' [37, 38]. Based on this idea, Ehrlich improved the chemical structure of several compounds to yield efficacious drugs against syphilis (under the trade name Salvarsan), trypanosome and spirochete infections [37, 38], which made him win the Nobel prize in 1908 sharing with Ilya Metchnikoff. Although Ehrlich's early definition of pharmacophore is almost unchanged for over a century, Schueler proposed the first modern definition in his book 'Chemobiodynamics and Drug Design' in 1960 [47], where the 'chemical groups' were replaced by patterns of 'abstract features'. Beckett and co-workers [48] proposed the first pharmacophore model of muscarinic agents in 1963 that identified distance ranges between abstract features, and later in 1967, Kier developed the first 'computed' pharmacophore model for muscarinic receptor inhibitor binding pattern [49–51]. Simple pharmacophores were in application as tools for designing new drug molecules much before the dawn of a well-defined field like computer-aided drug design. In the 1940s, preliminary structure-activity relationship models were computed based on simple two-dimensional model structures utilizing the accessible information of the van der Waals sizes and bond lengths [52]. Eventually, in the 1960s, three-dimensional models could be built with the convenience of X-ray and conformational analysis techniques. Medicinal chemists could classify some common molecular frameworks that attributed to high biological activity more often as compared to other structures by retrospectively analysing the chemical structures of the various drugs. Evans et al. [52] named such frameworks as

'privileged structures', which offer the basic scaffold and the substituents at different positions impart receptor specificity. Dihydropyridines [53], Arylethylamines, N-arylpiperazines, diphenylmethane derivatives, biphenyls and pyridazines [52, 53], tricyclic psychotropics and sulphonamides, benzodiazepines [54] are among some popular examples of the privileged structures. Woods and Fildes [55] found that p-aminobenzoic acid (PABA) and *p*-aminobenzenesulphonamide have similar critical distances; hence, bind to the PABA target with similar efficacy and inhibits the biosynthesis of tetrahydrofolic acid. This was one of the examples of the early two-dimensional pharmacophore models. An early 3D pharmacophoric approach was the 'three-point contact model' proposed by Easson and Stedman [56] and Beckett [48] in the case of (R)-(−)-adrenaline [= (R)-(−)-epinephrine]. These models are based on a concept that when a chiral centre is present in a compound, the substituents on this asymmetric atom make three-point contacts with the binding pocket of the receptor, which can only be obtained for one of the two isomers of epinephrine (the more active natural (R)-(−)-epinephrine). Similarly, another three-dimensional approach was developed in the early 1970s, characterizing the activity of clonidine on the central norepinephrine receptor [57]. It was observed that the natural ligand norepinephrine fits into the binding pocket of its target by three main interactions [57], viz. ionic bond between an anion (carboxylate, phosphate) of the binding pocket and the protonated $–NH_2$ functional group, a HB between the NH–CO group of the binding site and the secondary alcoholic hydroxyl and a π-stacking between the protonated imidazole of a histidine residue of the binding pocket and the aromatic ring of the drug. It was also recognized that the cationic head must be light and the phenolic –OH groups are not important for the biological activity. Pullmann et al. [58] in their 3D pharmacophore model of the norepinephrine receptor computed the critical intramolecular distances for the above key interactions which could successfully explain the pharmacophoric similarity between clonidine and norepinephrine, which in turn enables clonidine to make the same kind of interactions as norepinephrine.

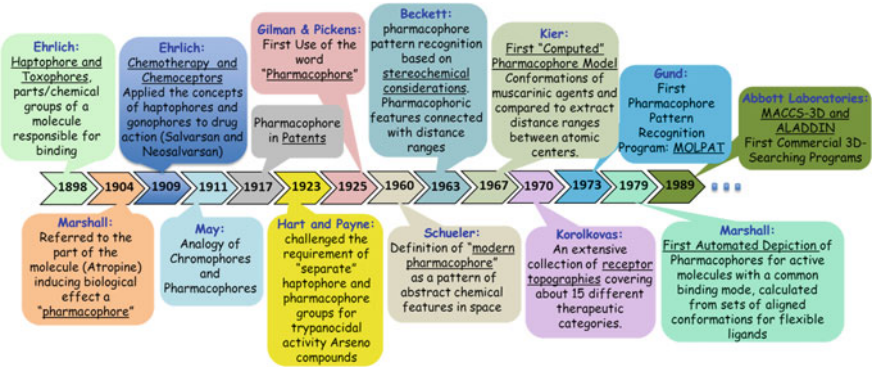

Fig. 2 Schematic presentation of timeline showing early developments in the field of pharmacophore modelling

These are some early efforts to explain pharmacophoric patterns that could act as key features for the design of new chemical entities. Figure 2 shows few early milestones in the field of emergence of pharmacophore modelling.

Nevertheless, in recent years, many effective pharmacophore modelling approaches and their contributions to drug discovery have been reported [59]. With the help of pharmacophoric insights and 3D searching tools, computer-aided drug design efforts are swiftly gaining efficiency since the 1990s. Still, this approach encounters many challenges that restrict its success. Pharmacophore approaches have been widely used in virtual screening, de novo ligand design, lead optimization and multi-target drug design. A range of automated pharmacophore modelling and screening tools have constantly appeared after the computational chemistry revolution witnessed in the past couple of decades [60]. Today, pharmacophore screening is one of the apt choices for researchers working in drug discovery and design.

5 Pharmacophore Model Generation

Pharmacophore models are typically generated either from a group of ligands, by aligning them and taking out the common interaction features indispensable for their biological activity. On the other hand, they can be constructed in a structure-based way, by probing probable interaction points in the receptor binding pocket, provided the 3D structure of the receptor is reported. The pharmacophore models can also be generated from a receptor–ligand complex by identifying the key interactions between the receptor and ligands.

5.1 Ligand-Based Pharmacophore Model Generation

Ligand-based pharmacophore modelling approach is used as a key strategy for facilitating screening compound databases when there is no three-dimensional structures are available for the target or receptor, but structure of a set of potent inhibitors are available. These active molecules are superimposed, and common pharmacophoric features representing crucial interactions between the ligands and the common target of these molecules are identified. Firstly, a conformational space of each of the active ligands is created corresponding to the flexibility of ligands, followed by their alignment and determination of the important common chemical features required for the creation of pharmacophore models. Currently, various automated pharmacophore generators are in use such as Phase [46] (Schrodinger Inc., http://www.schrodinger.com), HypoGen [61], HipHop [61] (Accelrys Inc., http://www.accelrys.com), GASP [62], DISCO [63], GALAHAD [64] (Tripos Inc., http://www.tripos.com) and MOE (Chemical Computing Group, http://www.chemcomp.com) [65]. Several academic programmes [40, 60, 66–68] are also

popularly being used. The key differences among these tools are mostly in the algorithms that are implemented for conformational search and alignment. This chapter is about the general steps followed by most of the programmes to recognize a pharmacophore pattern from a group of molecules that interact with a common receptor and the diverse applications of the pharmacophore concept.

5.1.1 Picking the Right Set of Compounds and Their Initial Structures

As the resulting pharmacophore models are highly inclined by the type, size and structural diversity of the participating ligands, it is imperative to choose the set of ligands that take part in the process of pharmacophore model generation. Some programmes like RAPID [69], HipHop [61] and the Crandell Smith method [70] assume all the compounds in the set as active, some other methods consider the information on the inactive molecules to be important as they give an idea about the structural features responsible for reducing the activities and the ones essential for enhancing activity. For example, DISCO [62, 71] and CLEW [72] provide an option to include or exclude inactive molecules in generating a model so that the user can identify the distinguishing features, while HypoGen [61] provides an option for including activity ranges of the set of ligands. As far as size of the dataset is concerned, most of the programmes are capable of handling up to 100 ligands in a set. If the dataset contains large number of molecules, then it can be sorted and categorized based on the activity value ranges. However, some programmes like SCAMPI [73] can handle up to a few thousand molecules but compromising the quality of the models. The high structural diversity of the dataset also is important to identify features that are most essential for target binding and produce high-quality models. Correct compound structures with correct atomic valencies, bond orders and properly defined aromaticity and the appropriate stereochemical flags are crucial for model generation.

5.1.2 Conformational Search

Ligands being flexible may have multiple possible conformations, and each conformation may bind to the binding site of the target in a particular fashion. Thus, it is crucial to consider the flexibilities of each molecule during pharmacophore development. Conformational search is considered as a separate stage in most of the pharmacophore modelling programmes like HipHop, DISCO and RAPID, where a large number of conformations are generated for each ligand. Systematic search, Monte Carlo sampling and molecular dynamics are the methods of choice for most of the software for conformation generation. As, the number of all possible conformers for molecules (especially when they have complex structures with a large number of rotatable bonds) is too large to handle and incorporate in the pharmacophore model building, energy minimization and clustering methods are used to

reduce the conformational space. The conformers with lowest energy or representatives from clusters of similar conformers are chosen to take part in model generation. In some other software, conformational search is parallelly performed along with pattern identification by retaining the conformers that possess certain features in a particular spatial arrangement. GASP [63] and GAMMA [74] use such an approach by the genetic algorithm (GA) techniques.

5.1.3 Feature Extraction and Representation

After conformational search, the molecules are subdivided into a set of features, each feature having the capability to form a particular type of non-covalent interaction with the receptor. There are three main levels of resolution for defining the features; (i) it may be atom based as implemented in MPHIL [75], GAMMA [74] and RAPID [69], where 3D atomic position related to the atom type is used as a feature; (ii) it can be atoms grouped into topological features such as a C = O group or a phenyl ring; or (iii) it may be function based, where the atoms are assembled into functional features describing the type of non-bonded interactions with the receptor. These features are HB acceptor (A), HB donor (D), base (+ve charge pH 7) (P), acid (−ve charge, pH 7) (N), aromatic moieties (rings) (R) and hydrophobic group (H). We have already discussed these features in Sect. 3 of this chapter. The third type of feature extraction method is immensely popular and is being used in many programmes like catalyst [43], Phase [46], HypoGen and HipHop [63]. Different topological features having the same chemical function can fall under same functional feature category. At the same time, the functional features are not assigned exclusively for any functional group. For instance, a –OH oxygen can act as both HB acceptor, a donor and at times may act as negatively charged feature. Commonly, the functional groups like a negatively/positively charged species, HB donor and acceptor are represented by their centres, which are nothing but the exact atom positions. Additionally, HB acceptors and donors are often represented by a vector that enforces a restriction of bond directionality between the feature on the binding site of the receptor and the complementary ligand feature. The centre of a hydrophobic site or an aromatic ring is defined as the centroid of the group.

After extracting the features, depiction of the whole molecule's structure is obtained by combining the selected features. These representations are generated mostly as: (i) 3D point set, where a ligand structure is represented as a group of categorized points in the 3D space, where each point is linked with a feature, (ii) a labelled graph, where nodes correspond to the features and the edges correspond to the relations, or (iii) a set of interpoint distances, where the ligand structure is represented as a collection of feature points, along with their interpoint distances. The third type of representation is commonly stored as a $n \times n$ distance matrix, n being the number of atoms.

5.1.4 Pattern Identification and Scoring

Once the features extracted for each ligand in the dataset, a pattern is identified as a set of relative positions in the 3D space, each linked to a feature. If a ligand holds a set of features in at least one of its conformations, the set of features can be aligned with the corresponding locations. Most of the methods are based on spatially overlaying conformations of various compounds with the pharmacophores points with minimal root mean square alignment errors. One can roughly classify the alignment methods as either point or property-based. In the first class of algorithms, pairs of pharmacophoric features are generally aligned using a least-squares fitting using clique detection methods [76, 77]. According to the graph-theoretical approach to molecular structures, a clique is a maximum completely connected sub-graph, which recognizes all imaginable combinations of atoms/functional groups to find out common substructures for the alignment. Property-based or field-based algorithms utilize grid or field descriptors, based on molecular properties such as volume, shape, charge distribution, electron density and electrostatic potentials of molecules. A 3D grid is generated about a ligand by computing the interaction energy components between the ligand and a probe placed at each grid point. Properties are calculated on a grid and later converted to a set of Gaussian representations. A number of either random or thoroughly sampled initial configurations are then generated followed by local optimizations with some similarity measure of the intermolecular overlap of the Gaussians.

After obtaining the pharmacophore candidates in the previous stages, they are generally scored and ranked. The basic obligation of a scoring scheme is implemented such that a high score implies higher chance of the ligands mapping to the pharmacophore model. Despite the great advances, molecular alignment handling ligand flexibility and proper selection of training set compounds are considered as the biggest challenges in ligand-based pharmacophore modelling.

5.2 Structure-Based Pharmacophore Model Generation

Structure-based pharmacophore modelling requires the 3D structure of the receptor or a receptor–ligand complex. The models are generated based on the spatial relationships of complementary interaction features of the binding pockets followed by selection and assembly of features to generate pharmacophore models.

5.2.1 Active Site Identification

The input for receptor-based pharmacophore modelling is the three-dimensional structure of a receptor usually in PDB format. The receptor binding pocket is identified using a spherical probe with customizable radius and location to include the binding site as well as the key interacting residues involved with ligands.

ThereĆare several programmes available for detection of clefts, crevices and binding pockets and to suggest possible active site locations based on the geometry of the surface [78, 79]. The key residues can be determined by user, deduced from studying the activity of the protein after mutation of a single residue. If mutation of a particular residue hampers function of the protein, then that residue may be part of the active site. Computational analyses such as multiple protein structural alignment techniques also help in identifying the active site of a protein by comparing it with a similar protein with known active site.

5.2.2 Complementary Image Construction

The receptor binding pocket is analysed to create an interaction map of features that the molecule is anticipated to satisfy for a reasonable interaction with the active site. In other words, a complement of the receptor binding site is created as the basis to create an input pharmacophore model. In particular, functional features like HB donors/acceptors and hydrophobic groups are identified in the binding site followed by rational placement of complementary features within the binding pockets in chemically acceptable positions [80, 81].

5.2.3 Generation of Queries, Searching and Hit Analysis

Once the active site is defined and chemically characterized, there is no straightforward single step to derive pharmacophore models from the binding site map. Since the receptor binding site has a potential to bind a variety of molecules in a variety of binding conformations, the interaction map often gives rise to huge number of features. To address this problem, adjacent features of the same type are clustered and the feature that lies nearest to the geometric centre of the cluster is retained as the cluster representative and all the other features are discarded. Sometimes, the number of the features is still very high even after the clustering, and all of them cannot be used as a single model because models possessing all such features would not be able to obtain any hits from the database. So, possible combinations of limited numbers of features are derived from the interaction map and multiple pharmacophore modes are composed. And then, these models are used by programmes like catalyst [43, 82] implemented in Accelrys Discovery Studio to search the compound database and test the validity of the models (also termed as pharmacophore ‘queries’ in catalyst) to screen or reject highly active compounds. It is always necessary to examine these models for how they interact with the binding site residues and how far the models extend within the binding pocket and if they fill specificity pockets and make the strongest interactions. Queries describing only the features present in an inhibitor might end up giving many false positive hits. At times, they screen compounds that are able to map to all the query features but also contain a bulky substituent causing steric hinderance and averting the compound

from fitting into the binding site. That is why inclusion of some excluded volume features is often recommended which penalizes the molecules' score if some atoms or group are placed in positions where they are likely to collide with the active site atoms.

5.3 *Generation of Pharmacophore Models from the Protein–Ligand Complexes*

Protein–ligand complexes produced by X-ray crystallography provide a detailed picture of the interactions between the ligand and the receptor, showing which atoms of the ligand are in contact with the receptor along with the atomic coordinates of those atoms. Also, the type of interactions can also be delineated from the atom types, distances and orientations of the ligand and receptor atoms. The major interaction that occurs in the receptor–ligand interface is hydrogen bonding. But other non-covalent interactions such as π–π and cation–π interactions are also obviously essential for protein–ligand complex formation apart from the hydrogen bonding. We have extensively looked at the importance of these interactions and the cooperativity existing among themselves to maintain supramolecular structures [83–86]. This information is of immense importance to establish a pharmacophore model from the complex. However, one needs to give attention to the facts that alternative pharmacophore models are possible within a single binding pocket owing to the flexibilities of both the active site and the ligands which are capable of rearranging themselves to accommodate different ligands and also there is a possibility of more than one active sites for a particular receptor. The programmes like 'LigandScout' developed by Wolber and Langer [87] and Phase [46] module of Schrodinger suite generate structure-based pharmacophore models from the protein–ligand complexes given as an input. We will be discussing the steps of generation of pharmacophore models from the protein–ligand complexes by the LigandScout and Phase, where the former characterizes the pharmacophoric features using kekule's patterns and the latter prioritizes the features based on the XP docking energy components.

5.3.1 Pharmacophore Model Generation with LigandScout

With the LigandScout [87] programme, as a first step, the correct molecular topology of rings and of hybridization state are assigned to the ligands by analysing the neighbouring atoms followed by assignment of double bonds and Kekule's patterns for functional groups such as carboxylic acids and esters, nitro groups, sulphonyl groups, thio acids, thio acetic esters guanidine-like groups, acetamidine and phosphinoyl groups functional groups. Next, the pharmacophoric features based on the hydrogen bonds, electrostatic interactions, charge transfer or

hydrophobic interactions between the ligand and the receptor are defined, and models are generated. Atoms belonging to nonacidic –OH groups (all –OHs excluding carboxylic, sulphinic, sulphonic, phosphonic or phosphinic acids), –SH groups, –C≡C– hydrogens and –NHs (barring trifluoromethyl sulphonamide hydrogens and tetrazoles) are recognized as HB donor atoms. When such an atom is found in the distance range of 2.5–3.8 Å from the heavy atom of a HB acceptor of the receptor molecule, a donor feature consisting of a donor point on the ligand side and a projected point on the macromolecule side is created. Atoms like –OH oxygen, –SH sulphur, –C≡C– carbon or –C≡N nitrogen are recognized as acceptor atoms, and an acceptor feature is placed with the initial point positioned on the acceptor atom and the projected point placed onto the heavy atom of the HB donor on the receptor within the distance range of 2.5–3.8 Å. The electrostatic interaction is represented as a vector resembling the definition of the H-bond acceptor. Hydrophobic areas are implemented in the form of spheres with a tolerance radius of 1.5 Å located in the centre of hydrophobic atom chains, branches or groups after testing a group of adjacent atoms to attain a sufficient overall hydrophobicity score.

5.3.2 e-pharmacophore Model Generation by Phase

The e-pharmacophores method of Phase module [46, 88] of Schrodinger suite is a new approach that utilizes the grid-based ligand docking with energetics (Glide) extra precision (XP) scoring function [89] to precisely quantify protein–ligand interactions. XP scoring function calculates enthalpic contribution of each interacting (pharmacophoric) site of a molecule towards the total score. Thus, each site gets a score based on the sum of enthalpic terms (such as HB, electrostatic, cation–π, π–π, hydrophobic and hydrophobically packed/associated HBs and other interactions) and is ranked. Then the e-pharmacophore models are generated from the top scoring features. The user can choose the number and type of features required to build a model. E-pharmacophores also include excluded volumes representing the regions of space occupied by the receptor where any portion of the ligand cannot be accommodated. E-pharmacophores have been shown to screen diverse set of bioactive molecules as compared to conventional structure-based methods, making it more useful.

5.4 Dynamic Pharmacophore Model Generation and Multicopy Simulations

The active sites of the drug targets being very flexible, structure-based pharmacophore models derived from a single conformational state of the protein may not satisfactorily account for all the possible potential drug–target interactions. In this situation, molecular dynamics simulation has been a very competent method to

tackle the target flexibility issues in SBDD. Dynamic pharmacophore models recognize compounds, which complementarily bind to the protein considering flexibility of their binding pockets, theoretically reducing the entropic penalties experienced by the protein due to ligand binding. MD simulation trajectories would give rise to multiple conformations of a protein active site, describing the targets' intrinsic flexibilities. Multiple copy minimization is also a regularly used exercise in computational drug design. The technique first fills the active sites of the receptors with multiple copies of probe molecules those do not react among themselves. Then, molecular dynamics, Monte Carlo/steepest descent minimizations are performed to minimize all these probes parallelly to obtain local minima. When the probes are clustered in the various regions of the active site in different orientations, the relative preferences of the binding regions can be estimated from the number of probes or the interaction energies.

Highly ordered and smaller clusters represent highly crucial prerequisite for favourable interactions, while the haphazardly spread larger ones indicate highly flexible sites. The MUSIC algorithm [80, 90] implemented with the BOSS programme uses similar strategy. It is capable of performing Monte Carlo simulations for a wide range of biomolecular systems in solvent clusters and mixtures and periodic solvent boxes with multiple solutes. It is able to calculate the interaction energies between solvent–solvent, solvent–solute and solute–solute. Usually, the probe or solvent are small molecules. For example, hydroxyl groups, aromatic groups and carbonyl groups are represented by small probes like $–CH_3OH$, C_6H_6 (Benzene) and $–CH_3CO$ (acetone), respectively. The probe molecules as well as the side chains of the receptor can be treated as rigid, partially/fully flexible or all-atom. The wide-ranging OPLS force field used in this programme is proven to be successfully handling the flexibilities of the receptor while generating pharmacophore models. Applications of the dynamic pharmacophore models will be discussed in the subsequent sections of the chapter.

6 Pharmacophore Finger Prints

The complex 3D structure of a molecule is reduced to an abstract collection of features in the pharmacophoric approach. Extending this concept, the structure of a molecule can be interpreted a as an exclusive data string by extracting all possible three-/four-point sets of pharmacophoric features. The inter-feature distances are assigned using distance binning or simply by bonds. These resulting unique strings describing the frequency of every possible combination at predefined loci of the string are known as pharmacophore fingerprints. Different types of molecular similarity analyses among libraries of molecules have been carried out using pharmacophore fingerprints [91, 92]. Also, the pharmacophoric fingerprint can be used to detect the common key features/groups contributing to the biological function of a group of active ligands.

7 Applications of Pharmacophore-Based Approaches

In this section, we discuss the diverse applications of the pharmacophore approaches under different scenarios.

7.1 Pharmacophore Approaches for Virtual Screening

Pharmacophore models being very simple by their definition can be used in a variety of ways depending on the research problem. This simplicity makes 'pharmacophore based search' a tool of choice for drug discovery scientists in the last decade [93]. When the structure of a set of molecules with similar or different scaffolds active on a particular target are known, then ligand-based pharmacophore models can be developed using their structures as described in Sect. 5.1. If the structures of some inactive derivatives are also known, then contribution of each feature towards the bioactivity can be compared between the positive and negative datasets to distinguish the wanted and unwanted features. The allowable steric arrangement of the ligands can also be mapped. When only the structure of the receptor or a receptor–ligand complex is available, then pharmacophore models are generated as described in Sects. 5.2 and 5.3 and can be utilized as queries to screen a database not only to screen compounds satisfying certain geometric and chemical restraints, but also to filter molecules with undesirable properties. For example, Voet and co-workers identified specific antagonists of human androgen receptor by applying two pharmacophoric filters back to back. One model is being generated from the available receptor-agonist complexes, while the other filter applied was a pharmacophore model generated from the receptor-antagonist complex. This approach enabled the authors to screen the compound that matches the antagonist-specific feature [94].

7.1.1 Dynamic E-pharmacophore Models: A Case Study with Mycobacterial CmaA1

We present here the summary of our recent work (Choudhury et al. [11, 17, 18]) on generation and application of dynamic structure and ligand-based pharmacophore models for screening a certain library against a mycobacterial target cyclopropane synthase (CmaA1). Mycolic acids are the characteristic constituents of Mtb cell wall which contribute towards the drug resistance, pathogenicity and persistence of the parasite. CmaA1 enzyme catalyses the cis-cyclopropanation of unsaturated mycolic acid chains at the distal position, which is an indispensable step in mycolic acid biosynthesis and maturation, thus making CmaA1 an important Mtb drug target. Five model systems of CmaA1 corresponding to different stages of cyclopropanation were studied using molecular dynamics (MD) simulations. A detailed picture of the structural changes in the two distinct binding sites, i.e. cofactor and

acyl substrate binding sites of CmaA1 during the cyclopropanation process was obtained by analysing the MD simulations trajectories. The apo-state of CmaA1 was observed to have a closed conformation where the cofactor binding site is inaccessible. Upon cofactor binding, H-bond between Pro202 of loop10 (L10) and Asn11 of N-terminal α1 helix disrupts making the cofactor binding pocket accessible. Upon cofactor binding, the non-polar side chains of the substrate binding site position towards the inner side of the pocket forming a hydrophobic environment for the substrate. In order to exchange the methyl group from the cofactor to the substrate, both the ligands tend to come close to each other facilitated by the upliftment of loop10. These observations prompted to think that the protein can remain in diverse conformations at different stages of its catalytic function and considering only one conformation for drug design would not be sufficient. So multiple structures obtained from the MD trajectories were used to generate, validate and use structure and ligand-based pharmacophore models.

7.1.2 Generation of Dynamic Structure-Based Pharmacophore Models

The molecular dynamics simulations on CmaA1 revealed that the binding sites of the enzyme exhibit huge conformational diversity, when bound to different ligands at various stages of its function. To use this conformational diversity of the binding sites in structure-based drug design, representative structures (snapshots) were extracted from all the five MD trajectories at a regular interval of 5 ns, thus obtaining a total of forty conformations of CmaA1 bound to different ligands in the two binding sites. The crystal structure of CmaA1 reported in PDB was also added to this pool. Now these 41 protein–ligand complexes were used to obtain e-pharmacophore models as described in Sect. 5.3.2. The first step used was evaluating the Glide energy terms. Active site of each CmaA1 structure was defined as a cubical box of 12 * 12 * 12 $Å^3$ dimension, and the Glide [89] energy grids were generated. Glide scores with XP descriptor information were obtained for the already bound ligands keeping their original conformations unchanged (unlike a typical docking where protein is held rigid while ligands are kept flexible). This exercise calculated all the interaction energy components between the receptor–ligand complexes, which were then submitted to the Phase module of Schrodinger to develop energy-based e-pharmacophore [88, 95] models. Figure 3 depicts the steps of the e-pharmacophore model generation and selection of best ones as virtual screening filters.

7.1.3 Pharmacophore Model Validation

To examine the capabilities of the dynamics-based e-pharmacophore models to successfully distinguish inhibitors and non-inhibitors of CmaA1, a set of 23 reported CmaA1 inhibitors (MIC:0.0125–12.5 μg/mL) [96] were used as a positive

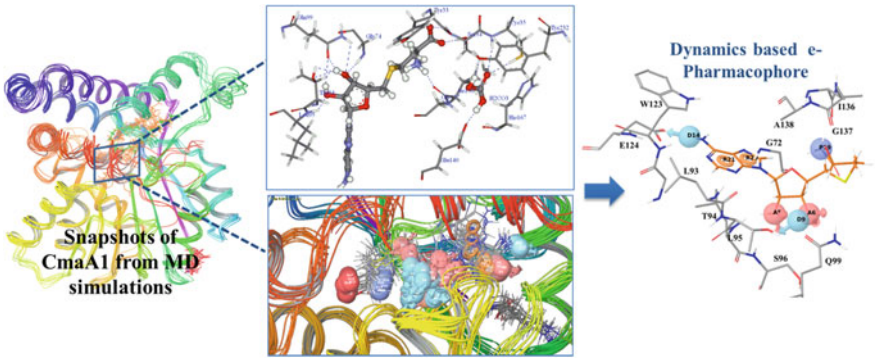

Fig. 3 Generation of dynamics-based e-pharmacophore models from the MD trajectory. The associated active site residues' interactions have been shown. The colour representations for the features are same as Fig. 1

dataset and 1398 Mtb inactive compounds reported in ChEMBL database (molecular weight ranging from 180 to 400, number of heavy atoms ranging from 12 to 27, similar to SAM/SAHC and the 23 inhibitors) were used as the negative dataset. Structures of these molecules were energy minimized and five lowest energy conformers were chosen for each of them. All these conformations were mapped to the 41 e-pharmacophore models using the 'advanced pharmacophore screening' option of Phase. Fast conformational sampling was used during pharmacophore screen, excluding molecules with >15 rotatable bonds. Molecules, which could be mapped to at least four pharmacophoric sites of each model were screened and among several conformers of a molecule the one with the best fitness score (S) given by the following equation [46] was retained for each compound. S is a measure of volume overlap and extent of match of chemical nature and directionalities of the pharmacophoric features with the corresponding complementary features of the molecules.

$$S = W_{\text{site}}\left(1 - S_{\text{align}}/C_{\text{align}}\right) + W_{\text{vec}}S_{\text{vec}} + W_{\text{vol}}S_{\text{vol}} + W_{\text{ivol}}S_{\text{ivol}}$$

where $W_{\text{site}} = (1 - S_{\text{align}}/C_{\text{align}})$, S_{align} = alignment score, C_{align} = alignment cut-off, S_{vec} = vector score, W_{vec} = weight of vector score, S_{vol} $(V_{\text{common}}/V_{\text{total}})$ = volume score, W_{vol} = weight of volume score, S_{ivol} = included volume score. Detailed explanations of the components of the fitness score are given in reference 47. Volumes were computed using van der Waals models of all atoms except non-polar hydrogens, and W_{ivol} is the weight of volume score. C_{align}, W_{site}, W_{vec}, W_{vol} and W_{ivol} are user-adjustable parameters, with default values of 1.20, 1.00, 1.00, 1.00 and 0.0, respectively.

Analysis of the hits obtained from these pharmacophore screening showed that most of the models developed from the CmaA1 complexes obtained from the MD trajectories were able to screen up to 17 reported inhibitors (out of 23), while the model developed from the crystal structure could screen only one inhibitor.

The fitness scores of the molecules with the dynamics-based models were also found to be higher. To further confirm our observation, a docking-based virtual screening was parallelly performed with the 41 CmaA1 snapshots and the reported inhibitors. Docking with the MD CmaA1 snapshots not only could bind the most active inhibitors as top scored hits, but also the docking scores were higher than the ones with the crystal structure. These results thus throw light on the effect of including multiple conformations of the targets on the screening abilities of the pharmacophore models. Five out of the 40 dynamic e-pharmacophore models were selected to be further used in our virtual screening study based on the consistency of docking and pharmacophore screening results.

7.1.4 Dynamic Ligand-Based Pharmacophore Models: Construction and Validation

Dynamic ligand-based pharmacophores were developed for the cofactors SAM and SAHC considering their conformational heterogeneity in CmaA1 binding sites as observed from MD trajectories of the respective model systems. Average structures of SAM/SAHC were created after superimposing the conformations obtained from each trajectory using uniform weighting method. Phase module of Schrodinger is used to build the ligand-based pharmacophore models, each comprising six types and 8–11 numbers of chemical features depending on the number and type of interactions with the CmaA1 binding sites. To verify the screening efficiencies of these models, a positive dataset of 23 CmaA1 inhibitors [96] and a negative dataset of 1398 non-inhibitors (the same dataset used to validate the structure-based models described in the previous section) were screened against each of the models. The ligand-based models created using multiple conformations of the cofactors obtained from the MD trajectories could screen up to 22 out of 23 CmaA1 active compounds when the condition for matching was minimum four features of a model. The fitness scores of the inhibitors matching the dynamic-ligand-based pharmacophore models were also higher as compared to the one developed from the conformation of SAHC bound to the crystal structure which was able to match to four CmaA1 inhibitors.

7.1.5 Pharmacophore-Based Virtual Screening

Once the best structure and ligand-based pharmacophore models were validated, they were employed as filters in a novel virtual screening workflow consisting of four different levels of screenings, viz. ligand-based pharmacophore mapping > structure-based pharmacophore mapping > docking > pharmacokinetic properties (ADMET) filters. A focused library of 18,239 molecules from three different sources was used for our virtual screening studies. As the first component of the dataset, 6583 drugs reported in DrugBank were chosen, targeting drug repurposing. The second component of the dataset was a set of 701 molecules which were already reported to be highly active (<1 μM activity) on Mtb cells/

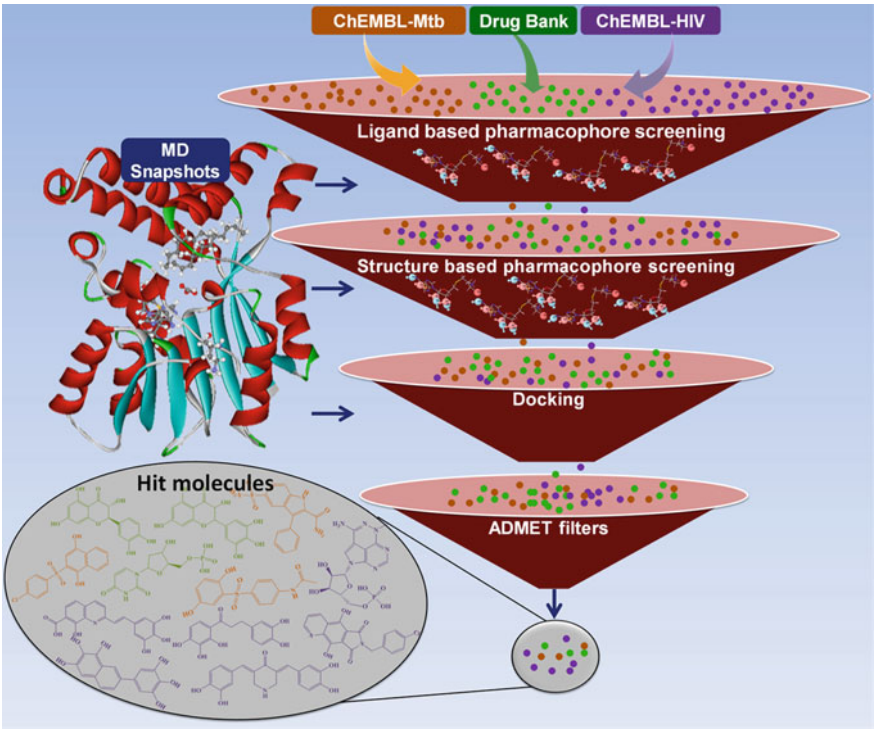

Fig. 4 Virtual screening workflow with structure and ligand-based pharmacophore models

targets and was considered to obtain molecules capable of acting on multiple Mtb targets including CmaA1. The third part of the dataset, i.e. a set of 11,089 highly active anti-HIV molecules (<1 μM activity on HIV cell lines/targets) was taken to screen molecules that can inhibit both Mtb-CmaA1 and HIV simultaneously. After subjecting these three subsets of molecules parallelly through the four screening filters, 12 compounds were obtained as potential anti-CmaA1 hits. As analysed from the Glide XP docking results, all of the identified hits made strong interactions with the important CmaA1 active site residues. Figure 4 shows virtual screening workflow with various levels of filters.

Virtual screening is usually a highly ordered approach combining diverse computational screening methods, where at each consecutive step, the filter criteria become more and more stringent, thus retaining the most promising compounds for experiments. As the steps proceed, the approaches used go on being more thorough and computationally expensive. So, being simple and fast by nature, pharmacophore models are usually implemented at the beginning of a hierarchical protocol to eliminate the compounds which do not even fulfil bare simple spatial and chemical requirements of the query, before subjecting the compound libraries to more complicated and computationally demanding docking calculations.

7.2 *Applications of Pharmacophores in Predicting Pharmacokinetic Properties*

Poor pharmacokinetic properties contribute majorly to failures of many drugs during development and clinical trials. Hence, these properties (also known as ADMET) must be profiled during the early drug discovery process so as to avoid failure at the later stages. Pharmacophore modelling approaches can be of great use for prediction of the ADMET properties. If one can identify the possible interactions made by a group of drug molecules having a well-defined ADMET profile with enzymes involved in drug metabolism, the common interacting features can be captured as pharmacophore models and equivalent features of the query molecules can be matched with the models. The cytochrome P450 (CYP) constitute the major group of enzymes involved in drug metabolism out of which isoenzymes 3A4, 2E1, 2D6, 2C19, 2C9 and 1A2 carry out 90% of the metabolism. Many recent studies report successful implementations [97, 98] of structure-based pharmacophore models trained from the known drugs CYP enzyme interactions to predict the suitability of query molecules to bind to a certain CYP. Also models to assess the probability of chemical alteration of the molecules by a CYP enzyme [99, 100] have been successfully developed and implemented. Inhibitors of the drug clearance enzymes such as the uridine 5′-diphospho-glucuronosyltransferases and transporters like P-glycoprotein/organic cation transporter have also been utilized to build pharmacophore models [101]. Pharmacophore models may also be employed to predict the possibilities of off-target binging of compounds accounting for the side effects, thereby helping design more target-specific compounds [102].

7.2.1 A Case Study with Hexadecahydro-1H-Cyclopenta[a] Phenanthrene Framework (HHCPF)

One of the recent studies from our group [20] reports implementation of ligand-based pharmacophore model features in combination with the QSAR techniques to establish a relationship between the number and type of pharmacophoric feature at a particular position of the core scaffold of a group of drugs with their drug-like properties and target binding affinities. A set of 110 FDA approved drugs containing the Hexadecahydro-1H-Cyclopenta[a]Phenanthrene Framework (HHCPF) (Fig. 5) was considered for the study to understand their structural and functional diversities and target specificities. Analyses of the target information collected from DrugBank, UniProt and PDB show the selectivity of the scaffolds for different targets and vice versa. The substituents present at 17 different positions of the scaffolds were classified as six pharmacophoric features, viz. H-bond donors, H-bond acceptors, aromatic rings, hydrophobic, charged and halogen groups. ADMET (human intestinal absorption, biodegradability, P-glycoprotein binding, carcinogenicity, Caco2 cell permeability, Ames test positivity, blood brain barrier permeability, hERG, CYP450 binding, Rat LD50, etc.)/physicochemical properties

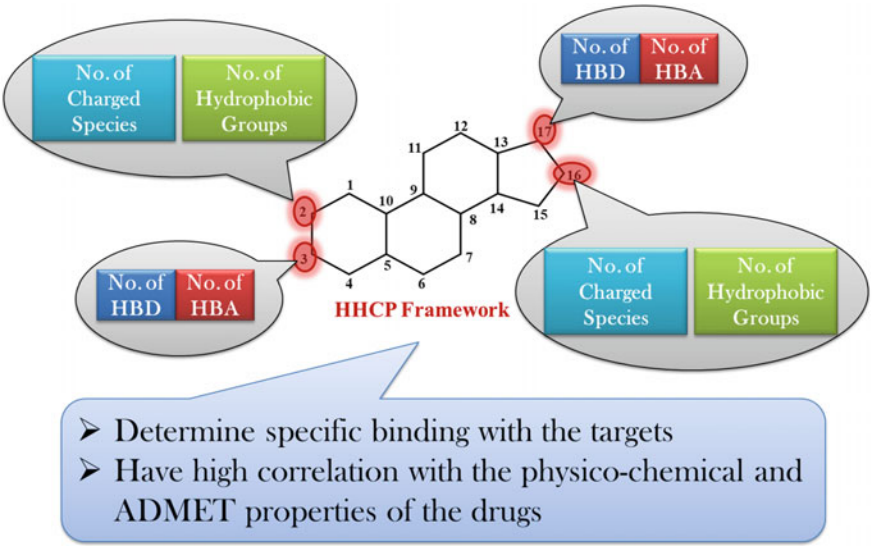

Fig. 5 Important substitution spots on the HHCPF, where number of different pharmacophoric features has a high correlation with target binding and ADMET properties

(polar surface area, polarizability, LogP, refractivity, etc., obtained from DrugBank) of the HHCPF drugs were observed to be highly correlated ($R > 0.8$) to the number and type of these pharmacophoric features at positions 3 and 17 of the framework. The chemical nature of the substitutions at different carbon atoms of the framework was observed to play extensive role in making specific interactions with the active site residues of their respective targets as revealed from analyses of the docking poses. The target binding was found to be greatly influenced by the presence/absence of aromatic rings, HB donors and HB acceptors as substitutions at different positions of the HHCPF scaffolds. Structure-based pharmacophore models were generated from the docked complexes of eight most important HHCPF drugs with their targets which can further be used to screen for new inhibitors. The general observation in the study was that the number and positions of double bonds in the framework regulate the preference of HHCPF drugs for a target class, and the substituents at particular carbon positions account for the target binding patterns and ADMET profiles.

7.3 Target Identification Using Pharmacophore Approaches

Pharmacophore models may also be employed to identify possible targets for active molecules, thereby facilitating the understanding of their mechanism of action. This approach is also proven to be helpful for studies that explore polypharmacology and

drug repositioning [103–105]. Firstly, pharmacophore-based fingerprints can be employed to search for similar molecules, whose mechanisms of action are already understood. In the other way around, pharmacophore models can be generated from the active sites of a group of probable proteins involved in the particular disease pathway and then the active molecules can be mapped to them to find out the best fit. The structures of these groups of proteins may be obtained from PDB or models generated using various techniques. The active site pharmacophore mapped with high scores can be proposed as potential targets for the compounds. A study on a group of plant metabolites and pharmacophore models of their possible targets was carried out by Rollinger et al. The best mapping targets were later proven to be accurate by experimental testing, thus validating the usefulness of the pharmacophore mapping approach [106].

7.4 De Novo Ligand Design with Pharmacophores

Apart from acting as a query to screen molecules with features at desired spatial locations and thus possibly prompting a desired biological response, pharmacophore models can also be employed for de novo design, of compounds, satisfying a specific physicochemical constrains. For example, the NEWLEAD method is able to create novel molecules from distinct disconnected fragments (mostly derived from known active ligands) that are consistent with the features of a pharmacophore model by using linkers. The linkers are small connecting fragment may be few atoms, chains or sometimes ring moieties [107]. Software packages like LUDI [108] or BUILDER [109] can grow such novel molecules when the receptor structures are also known. Many other packages also perform such de novo ligand design from the receptor-based pharmacophore features [110, 111]. Thus, pharmacophore models have versatile ways of application for lead generation. De novo design is meant to create entirely novel compounds, while pharmacophore searching screens the available chemical space. However, pharmacophore searching is faster and easier.

8 Limitations of Pharmacophore-Based Approaches

Though the literature is flooded a plenty of successful and reliable applications of pharmacophore-based approaches in rational drug design, its limitations should be cautiously considered as with any method [33, 112]. A systematic or straightforward way of constructing pharmacophore models is not available. This is the case especially with the receptor-based pharmacophore models where many different combinations of features are possible and each model may screen completely different set of molecules [113]. Lack of accuracy in pharmacophore scoring/fitness functions is one of the limitations of pharmacophore searching. So, quality of

mapping of a compound with a pharmacophore model which is often given by the RMSD between the feature of a model and atoms of the target molecule does not stand accurate as it does not take an account of similarity with the known active molecules [114]. Especially, the ligand-based pharmacophore models do not consider the overall compatibility with the receptor, thus sometimes end up with screening molecules those are very different from the other active compounds, with a completely different set of functional groups not complementary with the receptor. The pharmacophore-based searches against the compound databases lack fast conformation sampling as most of the programmes rely on conformer databases having only a limited number of energetically favourable conformations of molecules [115, 116]. There is a possibility of missing an active molecule if a suitable conformation is not available. So, it is desirable to generate as many low-energy conformers as possible for the database compounds, but again it would consume a lot of computational time. Especially for the rotatable bonds of small hydroxyl groups, it is difficult to sample all the different rotations.

9 Summary

Evolving from a simple concept to a well-validated and widely exploited method, the pharmacophore modelling approaches have been an essential part of many drug discovery strategies. The pharmacophore-based approaches are well known for their strength to propose a diverse set of molecules having diverse molecular frameworks but owing to a desired biological activity for one target. It has been extensively applied for virtual screening, lead optimization, target identification, toxicity prediction and de novo lead design, and it has ways to go [117]. Considering the strengths and limitations of the pharmacophore approaches, it can either be used alone to identify potential functional group substituents in molecules, design new molecules specific for a target by scaffold hopping keeping the substituents with certain pharmacophoric feature and orientation constant virtually screen for inhibitors, perform ADMET profiling of compounds, investigate possible off-targets or can be applied as a complementing approach along with other methods like docking and QSAR. The concept can be sensibly applied for fragment-based drug design, characterization of protein–protein interaction interfaces and target-based classification of chemical space. In this chapter, we touched upon the basic concepts and methods of generation of pharmacophore models. The diverse applications of the pharmacophore approaches exemplified though a number of case studies are believed to be useful for the readers. However, we believe that the choice and way of application of the method depends on the research problem and the type of initial data available.

Acknowledgements CC and GNS thank the Department of Science and Technology (DST), Government of India, for financial support in the forms of DST-INSPIRE Faculty Award [DST/INSPIRE/04/2016/000732] and JC Bose Fellowship, respectively.

References

1. Myers S, Baker A (2001) Drug discovery—an operating model for a new era. Nat Biotechnol 19:727–730
2. Moses H III, Dorsey ER, Matheson DH et al (2005) Financial anatomy of biomedical research. JAMA 294:1333–1342
3. Lahana R (1999) How many leads from HTS? Drug Discov Today 4:447–448
4. Veselovsky AV, Zharkova MS, Poroikov VV et al (2014) Computer-aided design and discovery of protein-protein interaction inhibitors as agents for anti-HIV therapy. SAR QSAR Environ Res 25:457–471
5. Song CM, Lim SJ, Tong JC (2009) Recent advances in computer-aided drug design. Brief Bioinform 10:579–591
6. Taft CA, Da Silva VB, Da Silva CH (2008) Current topics in computer-aided drug design. J Pharm Sci 97:1089–1098
7. Thiel KA (2004) Structure-aided drug design's next generation. Nat Biotechnol 22:513–519
8. Reddy AS, Amarnath HSD, Bapi RS et al (2008) Protein ligand intreraction database (PLID): datamining analysis of structure-function relationships. Comput Biol Chem 32:387–390
9. Reddy ChS, Vijayasarathy K, Srinivas E et al (2006) Homology modeling of membrane proteins: a critical assessment. Comput Biol Chem 30:120–126
10. Srivastava HK, Sastry GN (2012) A molecular dynamics investigation on a series of HIV protease inhibitors: assessing the performance of MM-PBSA and MM-GBSA approaches. J Chem Inf Model 52:3088–3098
11. Choudhury C, Priyakumar UD, Sastry GN (2014) Molecular dynamics investigation of the active site dynamics of mycobacterial cyclopropane synthase during various stages of the cyclopropanation process. J Struct Biol 187:38–48
12. Badrinarayan P, Sastry GN (2014) Specificity rendering 'hot-spots' for aurora kinase inhibitor design: the role of non-covalent interactions and conformational transitions. PLoS ONE 9:e113773
13. Srivastava HK, Choudhury C, Sastry GN (2012) The efficacy of conceptual DFT descriptors and docking scores on the QSAR models of HIV protease inhibitors. Med Chem 8:811–825
14. Srivani P, Srinivas E, Raghu R et al (2007) Molecular modeling studies of pyridopurinone derivatives—potential phosphodiesterase 5 inhibitors. J Mol Graph Model 26:378–390
15. Janardhan S, RamVivek M, Sastry GN (2016) Modeling the permeability of drug-like molecules through the cell wall of mycobacterium tuberculosis: an analogue based approach. Mol Bio Sys 12:3377–3384
16. Bohari MH, Sastry GN (2012) FDA approved drugs complexed to their targets: evaluating pose prediction accuracy of docking protocols. J Mol Model 18:4263–4274
17. Choudhury C, Priyakumar UD, Sastry GN (2015) Dynamics based pharmacophore models for screening potential inhibitors of mycobacterial cyclopropane synthase. J Chem Inf Model 55:848–860
18. Choudhury C, Priyakumar UD, Sastry GN (2016) Dynamic ligand-based pharmacophore modeling and virtual screening to identify mycobacterial cyclopropane synthase inhibitors. J Chem Sci 128:719–732
19. Reddy AS, Pati SP, Kumar PP et al (2007) Virtual screening in drug discovery—a computational perspective. Curr Protein Pept Sci 8:329–351
20. Choudhury C, Priyakumar UD, Sastry GN (2016) Structural and functional diversities of the hexadecahydro-1H-cyclopenta[a] phenanthrene framework, a ubiquitous scaffold in steroidal hormones. Mol Inform 35:145–157
21. Agrafiotis DK, Bandyopadhyay D, Wegner JK et al (2007) Recent advances in chemoinformatics. J Chem Inf Model 47:1279–1293
22. Vogt M, Bajorath J (2012) Chemoinformatics: a view of the field and current trends in method development. Bioorg Med Chem 20:5317–5323

23. Kapetanovic IM (2008) Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. Chem Biol Inter 171:165–176
24. Karelson M, Lobanov VS, Katritzky AR (1996) Quantum-chemical descriptors in QSAR/QSPR studies. Chem Rev 96:1027–1044
25. Gozalbes R, Doucet JP, Derouin F (2002) Application of topological descriptors in QSAR and drug design: history and new trends. Curr Drug Targets Infect Disord 2:93–102
26. Perkins R, Fang H, Tong W et al (2003) Quantitative structure-activity relationship methods: perspectives on drug discovery and toxicology. Environ Toxicol Chem 22:1666–1679
27. Kitchen DB, Decornez H, Furr JR (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 3:935–949
28. Paul N, Kellenberger E, Bret G et al (2004) Recovering the true targets of specific ligands by virtual screening of the protein data bank. Proteins 54:671–680
29. Kharkar PS, Warrier S, Gaud RS (2014) Reverse docking: a powerful tool for drug repositioning and drug rescue. Future Med Chem 6:333–342
30. Ravindra GK, Srivani P, Achaiah G et al (2007) Strategies to design pyrazolyl urea derivatives for p38 kinase inhibition: a molecular modeling study. J Comput Aided Mol Des 25:155–166
31. Srivastava HK, Bohari M, Sastry GN (2012) Modeling anti-HIV compounds: the role of analogue based approaches. Curr Comput Aided Drug Des 8:224–248
32. Gaur AS, Bhardwaj A, Sharma A et al (2017) Assessing therapeutic potential of molecules: molecular property diagnostic suite for tuberculosis ($MPDS^{TB}$). J Chem Sci 129:515
33. Yang SY (2010) Pharmacophore modeling and applications in drug discovery: challenges and recent advances. Drug Discov Today 15:444–450
34. Braga RC, Andrade CH (2013) Assessing the performance of 3D pharmacophore models in virtual screening: how good are they? Curr Topics Med Chem 13:1127–1138
35. Wolber G, Langer T (2005) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J Chem Inf Model 45:160–169
36. Koutsoukas A, Simms B, Kirchmair J et al (2011) From in silico target prediction to multi-target drug design: current databases, methods and applications. J Proteomics 74:2554–2574
37. Ehrlich P (1909) Ueber den jetzigen Stand der Chemotherapie. Ber Dtsch Chem Ges 42:17–47
38. Ehrlich P, Morgenroth J. Über Haemolysine (1900) Dritte Mitteilung. Berl Klin Wochnschr 37:453–457
39. Wermuth CG, Ganellin CR, Lindberg P et al (1998) Glossary of terms used in medicinal chemistry (IUPAC recommendations 1997). Annu Rep Med Chem 33:385–395
40. Güner OF (ed) (2000) Pharmacophore perception, development, and use in drug design, vol 2. Internat'l University Line
41. Bush B, Sheridan RJ (1993) PATTY: a programmable atom type and language for automatic classification of atoms in molecular databases. Chem Inf Comput Sci 33:756–762
42. Greene J, Kahn S, Savoj H et al (1994) Chemical function queries for 3D database search. J Chem Inf Comput Sci 34:1297–1308
43. Molecular Simulations Inc. (MSI), "Catalyst Software". http://www.accelrys.com/about/msi.html
44. Wang T, Zhou JJ (1998) 3DFS: a new 3D flexible searching system for use in drug design. Chem Inf Comput Sci 38:71–77
45. Pickett S, Mason J, McLay IJ (1996) Diversity profiling and design using 3D pharmacophores: pharmacophore-derived queries (PDQ). Chem Inf Comput Sci 36:1214–1223
46. Dixon SL, Smondyrev AM, Knoll EH et al (2006) PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. J Comput Aided Mol Des 20:647–671
47. Schueler FW (1946) Sex hormonal action and chemical constitution. Science 103:221–223
48. Beckett AH (1959) Stereochemical factors in biological activity. In: Fortschritte der Arzneimittel Forschung. Birkhäuser Verlag, Basel, pp 455–530

49. Kier LB (1967) Molecular orbital calculation of preferred conformations of acetylcholine, muscarine, and muscarone. Mol Pharmacol 3:487–494
50. Kier LB (1970) Receptor mapping using mo theory. In: Danielli JF, Moran JF, Triggle DJ (eds) Fundamental concepts in drug-receptor interactions, Academic Press: New York
51. Kier LB (ed) (1971) MO theory in drug research. Academic Press, New York, pp 164–169
52. Evans BE, Rittle KE, Bock MG et al (1988) Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists. J Med Chem 31:2235–2246
53. Thompson LA, Ellman JA (1966) Synthesis and applications of small molecule libraries. Chem Rev 96:555–600
54. Wermuth CG (1998) Search for new lead compounds: the example of the chemical and pharmacological dissection of aminopyridazines. J Heterocycl Chem 35:1091–1100
55. Woods DD, Fildes P (1940) The anti-sulphanilamide activity (in vitro) ofp-aminobenzoic acid and related compounds. Chem Ind 59:133–134
56. Easson LH, Stedman E (1933) Studies on the relationship between chemical constitution and physiological action. V. Molecular dissymmetry and physiological activity. Biochem J 27:1257–1266
57. Peroutka SJ, U'Prichard DC, Greenberg DA et al (1977) Neuroleptic drug interactions with norepinephrine alpha receptor binding sites in rat brain. Neuropharmacology 16:549–556
58. Pullmann B, Coubeils JL, Courrière P et al (1972) Quantum mechanical study of the conformational properties of phenethylamines of biochemical and medicinal interest. J Med Chem 15:17–23
59. Leach AR, Gillet VJ, Lewis RA (2010) Three-dimensional pharmacophore methods in drug discovery. J Med Chem 53:539–558
60. Güner OF (2002) History and evolution of the pharmacophore concept in computer-aided drug design. Curr Top Med Chem 2:1321–1332
61. Maynard AJ (2004) HypoGenRefine and HipHopRefine: pharmacophore refinement using steric information from inactive compounds. Presented at the ACS national meeting, Spring, 2004
62. Jones G, Willett P, Glen R (2000) GASP: genetic algorithm superposition program. In: Pharmacophore perception, development, and use in drug design, vol 2. International University Line, La Jolla, CA, USA, pp 85–106
63. Jones G, Willett P, Glen RC (1995) A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. J Comput Aided Mol Des 9:532–549
64. GALAHAD. Tripos, St. Louis, MO. http://www.tripos.com/
65. Lin A, Overview of pharmacophore applications in MOE. http://www.chemcomp.com/journal/ph4.htm
66. Vlachakis D, Fakourelis P, Makris C, Kossida S (2015) DrugOn: a fully integrated pharmacophore modeling and structure optimization toolkit. PeerJ 3:e725
67. Khedkar SA, Malde AK, Coutinho EC et al (2007) Pharmacophore modeling in drug discovery and development: an overview. Med Chem 3:187–197
68. Langer T, Hoffmann RD (eds) (2006) Pharmacophores and pharmacophore searches, pharmacophores and pharmacophore searches. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
69. Finn PW, Kavraki LE, Latombe JC et al (1997) Rapid: randomized pharmacophore identification for drug design. Comput Geom Theor Appl 10:263–272
70. Crandell C, Smith D (1983) Computer-assisted examination of compounds for common three-dimensional substructures. J Chem Inf Comp Sci 23:186–197
71. Martin YC (2000) DISCO: what we did right and what we missed. In: Pharmacophore perception, development, and use in drug design. International University Line, pp 49–68
72. Dolata D, Parrill A, Walters W (1998) CLEW: the generation of pharmacophore hypotheses through machine learning. SAR QSAR Environ Res 9:53–81
73. Chen X, Rusinko A III, Tropsha A et al (1999) Automated pharmacophore Identification for large chemical data sets. J Chem Inf Comput Sci 39:887–896

74. Handschuh S, Wagener M, Gasteiger JJ (1998) Superposition of three-dimensional chemical structures allowing for conformational flexibility by a hybrid method. J Chem Inf Comput Sci 38:220–232
75. Holliday J, Willet P (1997) Using a genetic algorithm to identify common structural features in sets of ligands. J Mol Graph Model 15:203–253
76. Bron C, Kerbosch J (1973) Algorithm 457: finding all cliques of an undirected graph. Commun ACM 16:575–577
77. Brint A, Willett P (1987) Algorithms for the identification of three-dimensional maximal common substructures. J Chem Inf Comp Sci 27:152–158
78. Guilloux VL, Schmidtke P, Tuffery P (2009) Fpocket: an open source platform for ligand pocket detection. BMC Bioinform 10:168–178
79. Schmidtke P, Bidon-Chanal A, Luque FJ et al (2011) MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. Bioinformatics 27:3276–3285
80. Carlson HA, Masukawa KM, Rubins K et al (2000) Developing a dynamic pharmacophore model for HIV-1 integrase. J Med Chem 43:2100–2114
81. Masukawa KM, Carlson HA, McCammon JA (2000) Technique for developing a pharmacophore model that accommodates inherent protein flexibility: an application to HIV-1 integrase. In: Guner OF (ed) Pharmacophore perception, development, and use in drug design. International University Line
82. Clark DE, Westhead DR, Sykes RA et al (1996) Active-site-directed 3D database searching: pharmacophore extraction and validation of hits. J Comput Aided Mol Des 10:397–416
83. Mahadevi AS, Sastry GN (2013) Cation-p interaction: its role and relevance in chemistry, biology and material science. Chem Rev 113:2100–2138
84. Chourasia M, Sastry GM, Sastry GN (2011) Aromatic—aromatic database, A2ID: an analysis of aromatic networks in proteins. Int J Biol Macromol 48:540–552
85. Saha S, Sastry GN (2015) Cooperative or anticooperative: how noncovalent interactions influence each other. J Phys Chem B 119:11121–11135
86. Badrinarayan P, Choudhury C, Sastry GN (2015) Molecular modeling. In: Dhar PK and Singh V (eds) Systems and synthetic biology (S2B2). Springer Press, pp 93–128
87. Wolber G, Langer T (2005) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J Chem Inf Model 45:160–169
88. Salam NK, Nuti R, Sherman W (2009) Novel method for generating structure-based pharmacophores using energetic analysis. J Chem Inf Model 49:2356–2368
89. Friesner R, Murphy RB, Repasky MP et al (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem 49:6177–6196
90. Carlson HA, Masukawa KM, McCammon JA (1999) Method for including the dynamic fluctuations of a protein in computer-aided drug design. J Phys Chem A 103:10213–10219
91. McGregor MJ, Muskal SM (1999) Pharmacophore fingerprinting. 1. Application to QSAR and focused library design. J Chem Inf Comput Sci 39:569–574
92. McGregor MJ, Muskal SM (2000) Pharmacophore fingerprinting. 2. Application to primary library design. J Chem Inf Comput Sci 40:117–125
93. Voet AR, Kumar A, Berenger F et al (2014) Combining in silico and in cerebro approaches for virtual screening and pose prediction in SAMPL4. J Comput Aided Mol Des 28:363–373
94. Voet A, Helsen C, Zhang KY et al (2013) The discovery of novel human androgen receptor antagonist chemotypes using a combined pharmacophore screening procedure. Chem Med Chem 8:644–651
95. Loving K, Salam NK, Sherman W (2009) Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. J Comput Aided Mol Des 23:541–554
96. Anuradha A, Trivelli X, Guérardel Y et al (2007) Thiacetazone, an antitubercular drug that inhibits cyclopropanation of cell wall mycolic acids in mycobacteria. PLoS ONE 12:e1343

97. Guner OF, Bowen JP (2013) Pharmacophore modeling for ADME. Curr Top Med Chem 13:1327–1342
98. Yamashita F, Hashida M (2004) In silico approaches for predicting ADME properties of drugs. Drug Metab Pharmacokinet 19:327–338
99. de Groot MJ, Ekins S (2002) Pharmacophore modeling of cytochromes P450. Adv Drug Deliv Rev 54:367–383
100. Ekins S, de Groot MJ, Jones JP (2001) Pharmacophore and three-dimensional quantitative structure activity relationship methods for modeling cytochrome p450 active sites. Drug Metab Dispos 29:936–944
101. Sorich MJ, Miners JO, McKinnon RA et al (2004) Multiple pharmacophores for the investigation of human UDP-glucuronosyltransferase isoform substrate selectivity. Mol Pharmacol 65:301–308
102. Hu Y, Bajorath J (2010) Polypharmacology directed compound data mining: identification of promiscuous chemotypes with different activity profiles and comparison to approved drugs. J Chem Inf Model 50:2112–2118
103. Keiser MJ, Roth BL, Armbruster BN (2007) Relating protein pharmacology by ligand chemistry. Nat Biotechnol 25:197–206
104. Koutsoukas A, Simms B, Kirchmair J et al (2011) From in silico target prediction to multi-target drug design: current databases, methods and applications. J Proteomics 74:2554–2574
105. Xu Y, Liu X, Li S (2013) Combinatorial pharmacophore modeling of organic cation transporter 2 (OCT2) inhibitors: insights into multiple inhibitory mechanisms. 10:4611–4619
106. Rollinger JM, Schuster D, Danzl B et al (2009) In silico target fishing for rationalized ligand discovery exemplified on constituents of ruta graveolens. Planta Med 75:195–204
107. Tschinke V, Cohen NJ (1993) The NEWLEAD program: a new method for the design of candidate structures from pharmacophoric hypotheses. Med Chem 36:3863–3870
108. Bohm HJ (1992) The computer program LUDI: a new method for the de novo design of enzyme inhibitors. J Comput Aided Mol Des 6:61–78
109. Roe D, Kuntz IJ (1995) BUILDER v.2: improving the chemistry of a de novo design strategy. J Comput Aided Mol Des 9:269–282
110. Joseph-McCarthy D (1999) Computational approaches to structure-based ligand design. Pharmacol Ther 84:179–191
111. Schneider G, Bohm HJ (2002) Virtual screening and fast automated docking methods. Drug Discov Today 7:64–70
112. Scior T, Bender A, Tresadern G et al (2012) Recognizing pitfalls in virtual screening: a critical review. J Chem Inf Model 52:867–881
113. Vancraenenbroeck R, De Raeymaecker J, Lobbestael E et al (2014) In silico, in vitro and cellular analysis with a kinome-wide inhibitor panel correlates cellular LRRK2 dephosphorylation to inhibitor activity on LRRK2. Front Mol Neurosci 7:51
114. Schomburg KT, Bietz S, Briem H et al (2014) Facing the challenges of structure-based target prediction by inverse virtual screening. J Chem Inf Model 54:1676–1686
115. Kirchmair J, Wolber G, Laggner C et al (2006) Comparative performance assessment of the conformational model generators omega and catalyst: a large-scale survey on the retrieval of protein-bound ligand conformations. J Chem Inf Model 46:1848–1861
116. Kirchmair J, Laggner C, Wolber G et al (2005) Comparative analysis of protein-bound ligand conformations with respect to catalyst's conformational space subsampling algorithms. J Chem Inf Model 45:422–430
117. Nagamani S, Gaur AS, Tanneeru K et al (2017) Molecular property diagnostic suite (MPDS): development of disease-specific open source web portals for drug discovery. SAR QSAR Environ Res https://doi.org/10.1080/1062936x.2017.1402819