Manipal Journal of Science and Technology

Volume 2
Issue 1 *Issue 1*

Article 1

6-1-2017

Computational drug designing in health care

Rajanikant G K
National Institute of Technology Calicut, India, rajanikant@nitc.ac.in

Review Articles

Computational drug designing in health care

Fayaz S M and Rajanikant G K*

Email: rajanikant@nitc.ac.in

Abstract

Drug designing is an important step of defending health care. This is the central stage in the treatment of various diseases and ailments. In this paper, the process of drug designing has been discussed in detail. The meaning of drug designing and its traditional and rational methodologies has been briefly described. Further, the process of computer-aided drug design has been explained extensively. The different methods such as structure and ligand-based drug design and extraction of respective pharmacophores have been described. In addition, virtual screening methods such as docking, drug-likeness prediction and molecular dynamics (MD) simulation are discussed. Finally, the real-life applications and examples employing computational drug design approaches have been described.

Keywords: Computer-aided drug design, structure-based drug design, ligand-based drug design, virtual screening, pharmacophore, docking, molecular dynamics simulation

I. Drug designing

Drug design/drug discovery in general refers to the inventive process of finding new medications for a particular ailment or a disease. It involves identification of a molecule, which modulates the activity of a biological target, involved in a disease related pathway [1]. The drug is usually a small molecule that binds to a biomolecule, such as protein, and activates or inhibits its function, which therapeutically benefits the patient. A number of metabolic activities and processes occur inside our body. Various pathways that involve biomolecules such as proteins, DNA and RNA, mediate these activities. Disruption in the functioning of these biomolecules, in turn interrupts the normal pathways, leading to a different state that we generally refer to, as a disease. In such a condition, a few of the biomolecules are either inhibited or activated. Thus, in order to avoid the disease effects, it is desirable to modulate these biomolecules by activating the inhibited ones and inhibiting the activated ones. These biomolecules are referred as biomolecular targets/drug targets. Therefore, the biomolecular targets are the key molecules involved in a specific pathway related to a disease. Hence, the small molecular modulators, known as drugs, are designed to bind to these drug targets and modify their action.

Traditional drug design approaches comprised of testing the small molecules on cultured cells (*in vitro*) or animals (*in vivo*), and comparing their effects to treatments. Initially, drug discovery involved the identification of the active constituents from traditional therapies. Later, the small molecular libraries comprising synthetic molecules, natural products or extracts were tested on cultured cells (*in vitro*) or animals (*in vivo*) to identify the therapeutic substances and comparing their effects to treatments. This method is known as classical pharmacology [2] or phenotypic drug discovery [3]. It does not require the knowledge of the mechanism of action of the drug or its biological target and is often successful

Fayaz S M
Department of Biotechnology, Manipal Institute of Technology, Manipal University, Manipal – 576104, India

Rajanikant G K
Associate Professor and Former Head, School of Biotechnology, National Institute of Technology Calicut, Calicut – 673601, India

* Corresponding Author

How to cite this article: Fayaz S. M., Rajanikant G. K, "Computational Drug Designing in Health Care", *Manipal J. Sci. Tech.*, vol.2(1), 1-14, 2017.

due to serendipity. However, it requires a lot of time, resources, and labour to identify a drug. These issues led the researchers towards a more focused approach known as the rational drug design.

II. Rational drug design

This approach involves modulation of a specific biological target that is involved in a disease. In a process known as reverse pharmacology, high throughput screening (HTS) of large compound libraries is carried out against isolated biological targets [4]. A biomolecule can be considered as a drug target, if its modulation by a small molecule, which binds to it, exhibits therapeutic value during experimental studies. Then, this target is considered "druggable."

Once a suitable target is identified, it is cloned and expressed for employing it in a screening assay. Moreover, the three-dimensional structure of this molecule may be determined through X-ray crystallography or Nuclear magnetic resonance (NMR) spectroscopy. Target specific small molecules can be identified by screening libraries of potential drug compounds against the target molecule. Compounds that bind to the desired target are known as hits. The efficiency of these hits is further tested through *in vitro* or *in vivo* assays. The effective hits are considered "drug-like," if they possess desirable pharmacokinetic properties such as ADMET (Absorption, Distribution, Metabolism, Excretion, and minimal Toxicity). The drug-likeness of a compound can be predicted using Lipinski's Rule of Five. However, the rational drug design process remained complex due to the huge requirement of investments, resources, and time for initial screening processes.

Therefore, in order to minimize at least the initial screening process and to simplify the drug designing method, Computer-Aided Drug Design (CADD) or *in silico* drug design was adopted.

III. Computer-Aided Drug Design (CADD)

Drug discovery remains a time consuming, costly, challenging, and ineffective process with low success rate [5]. In 2010, the cost of discovery of a new molecular entity (NME) was approximately $ 1.8 billion [6]. In order to minimize the failure of compounds at Phase I, II, and III clinical trials, it is essential to adopt novel strategies at the initial stages itself.

With the development of information and software technologies, the computers have been successfully employed in the drug design and discovery process. The CADD uses computational strategies to predict whether any compound will bind to a biological target molecule and if so, how strongly (binding affinity). Here, it is hypothesized that the strong binding of a compound can modulate a biological target. Only the compounds that are predicted to possess good affinity are synthesized for further testing, thereby saving enormous time and cost. Thus, the CADD has hastened the drug discovery process by minimizing the required trials [7, 8].

This review majorly focuses on the computational methods that are adopted for drug designing in health care. The CADD involves two major types of strategies known as structure-based drug design (SBDD) (Figure 1) and ligand-based drug design (LBDD) (Figure 2).

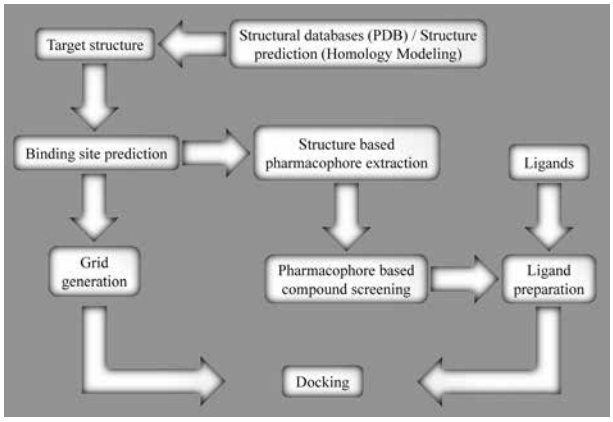

Figure 1: Work flow of structure-based drug design

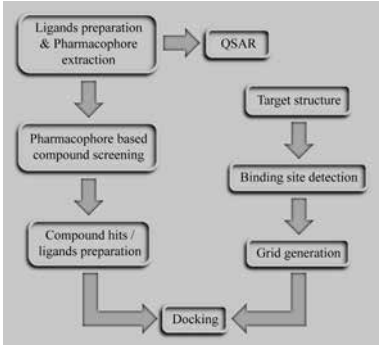

Figure 2: Work flow of ligand-based drug design

IV. Structure-based drug design (SBDD)

A Structure of the Target

This method involves the three-dimensional structure of the biological target molecule and hence, known as SBDD. The target molecule may be a protein, DNA or RNA. The structure of the target is usually determined through experimental methods such as X-ray crystallography or NMR spectroscopy [9]. The structure obtained through X-ray crystallography gives information that is more reliable over NMR spectroscopy that contains multiple conformers, which make the structural analysis difficult. In general, X-ray structures with good resolution (resolution at less than 2 Å) are considered for SBDD, since they contain more reliable structural information.

Sometimes, it may not be possible to determine the structure of the few biological molecules due to their size or difficulty in crystallization. In such cases, the structure of a target could be predicted. Protein structure prediction, in general, refers to the prediction of its secondary, tertiary, and quaternary structures from the primary structure. The primary structure of a protein is its sequence itself. The secondary structure involves α-helices, β- sheets, and loops. The tertiary structure is a three-dimensional and is made up of secondary structural elements, which define the protein function. The tertiary structure of a protein may contain only α-helices, only β-sheets, or a combination of both. The quaternary structure is the highest level of a protein structure that defines its biological function. The structural prediction could be done chiefly through three methods viz., homology modelling, threading, and *ab initio*.

The homology modelling method involves the structural prediction of a protein (target) using a known homologous protein structure (template). The first step is to identify the template with a similar sequence. This is done through algorithms such as basic local alignment search tool (BLAST). The target sequence is blasted against the sequences of all the protein structures in Protein Data Bank (PDB). The protein with the highest similarity is selected as the template. Further, the secondary structure of the target is compared with the template to see, if both the proteins share similar secondary structural elements. The co-ordinates of the template are copied and the structure of the target is built from them. The homology modelling is relatively simple and easy. It requires only two inputs viz., the target sequence and the template protein structure.

If there is no template for a target protein, then the homology modelling cannot be carried out. In such cases, the threading/fold recognition is adopted to identify the structure of the target protein. It is estimated that there are only around 1,000 to 10,000 stable protein folds in nature. The threading folds the target sequence through all known folds and finds the best fit of the sequence to a set of candidate folds. For each fold, the probability that the target sequence can have that fold is estimated through sequence-fold alignment using a fitness scoring function. The folds with the highest fitness scores are used to build the target protein structure.

The *Ab initio*/free modelling method is adopted to find the fold of target protein by simulating the biological process of protein folding. This is a very difficult task because a protein chain can fold into millions of different conformations. This method is carried out only when no detectable homologues can be found and when there are templates with extremely low homology. So far, this method has very limited practical applications. There is no requirement of the known structures in this method. Only the basic physical and chemical principles are applied. We need to scan all the possible conformations of the target protein in order to detect its correct fold/structure. This method is still at the experimental level.

Whatever may be the method adopted to identify the target protein structure, the predicted structure should be optimized and validated. It is a known fact that the biomolecules always tend to be in their stable state with minimum energy. Thus, the predicted structure is considered correct, if it is stable with the minimum energy.

In order to achieve a more reliable structure, the predicted structure is optimized through energy minimization. Finally, the optimized structure is validated through various parameters such as Ramachandran plot.

The structures of other biomolecules such as DNA and RNA have been identified and deposited in various structural databases such as PDB. These structures could be used as the starting points for the SBDD. However, if the structure of a target DNA/RNA is not available, it could be predicted using various DNA/RNA structure prediction tools (Table 1). The target structure is predicted based on the similar nucleotide sequences of the known DNA/RNA structures. Finally, the predicted structures are optimized through energy minimization to obtain reliably the realistic structures. However, at present, the knowledge regarding structural prediction of DNA/RNA is limited when compared to protein and hopefully, a number of strategies, methods, and tools would be available for the same in the near future.

Table 1: Computational tools to predict/build the structure of biomolecules

Sl No	Biomolecule	Software/Module
1	Protein	Modeller
2	Protein	Prime (Schrodinger)
3	Protein	Swiss-Model
4	Protein	RaptorX
5	Protein	Yasara
6	Protein	I-Tasser
7	Protein	Phyre
8	Protein	ROBETTA
9	DNA	3D-DART
10	DNA	Graphite-LifeExplorer
11	DNA	DNAtools
12	DNA	3DNA
13	RNA	Mfold and MSym
14	RNA	RNAfold
15	RNA	RNA123
16	RNA	RNAstructure
17	RNA	BARNACLE
18	RNA	FARNA
19	RNA	iFoldRNA

B. Binding site

Drugs do not simply bind at any place on their targets. They need enough space to settle and charge with to share. There are certain places on the biomolecular target that contain these features and could accommodate small molecule drugs. These places are known as binding sites. Drugs always bind to their targets at the binding sites only. If the target is a protein enzyme, then its active site is mostly considered its binding site. Hence, the drugs will be complementary in shape and charge to the binding site of the target.

Binding site identification is the first step in the SBDD. The target, for instance the protein target, contains information regarding the key interaction sites that are necessary for its interaction with the ligands within the binding pocket. The key features in the binding site are:

H-bond donors: Polar hydrogen atoms in the amino acids that act as electron donors.

H-bond acceptors: Electronegative atoms in the amino acids like oxygen and nitrogen, with lone electron pair(s), that act as electron donors.

Hydrophobic groups: Carbons in the amino acids with hydrophobic side chains. Since the binding site is present deep from the surface of the protein, it is mostly made up of hydrophobic amino acids.

Knowledge regarding the binding site is indispensable for the SBDD. Generally, the binding site information is obtained through literature or from the target structure. Mutational studies would have been carried out on the target and the information such as the location of the binding site in the protein, key amino acids/nucleotides present in the binding site, etc. is reported in the literature. Based on this information, the binding site of the target is considered. Another way to find the binding site is to check whether only the target structure is identified or is it along with any ligand/ inhibitor. If this target-ligand complex structure is available, then we could consider the binding site of this ligand for the SBDD.

Occasionally, there may not be any information regarding the binding site. In such cases, the binding site of the target is predicted using the computational tools (Table 2). For a protein, the binding site is a pit/pocket at its surface. Thus, all the probable sites on a protein are predicted and the site with the highest volume is generally considered the binding site. Through the experimental data, it has been inferred that the pocket with the highest volume, compared to the other pockets on a protein surface, is the binding site.

Table 2: Computational tools to predict the binding sites of a protein

Sl No	Server/Software/Module
1	SiteMap (Schrodinger)
2	CASTp
3	Pocket-Finder
4	Discovery Studio
5	ConSurf
6	3DLigandSite
7	FINDSITE
8	LIGSITE
9	metaPocket
10	Q-SiteFinder

However, prediction of the binding site is not easy in the case of DNA/RNA. At least, some literature inputs regarding this aspect is very much needed. For a DNA target, the binding site is mostly present in one of its grooves, since it is a linear double helical structure. This is true in the case of pre-miRNA also. However, finding the actual location of the binding site is difficult. It could not be predicted as surface pits, as in the case of a protein. In the case of DNA, the region with which it interacts with other biomolecules is generally considered the binding site. Similarly, for miRNA, the dicer-binding site is generally selected as its binding site. Whatever may be the biomolecule, the binding site should be selected based on the mechanism of action of the biomolecule that we intend to target and the drug should be able to modulate its action by binding to this site.

C. Structure-based pharmacophore

SBDD mainly involves screening a database of ligands for a given target molecule, known as docking, or *in situ* building of ligands at the binding site using small fragments. However, the disadvantages of both these categories is that screening a ligand database against the target structure might prove to be computationally time consuming and tedious and the ligand building may lead to the design of novel chemical entities that might not be feasible for the synthesis. To overcome these complications, the concept of structure-based pharmacophore was introduced.

A pharmacophore is a set of molecular features that are necessary for the recognition of a ligand by a biological target molecule. The description of the features is specific to the binding site of a target molecule, which a ligand should possess in order to bind to this molecule (Figure 3) [13]. A pharmacophore usually contains molecular features such as H-bond donors, H-bond

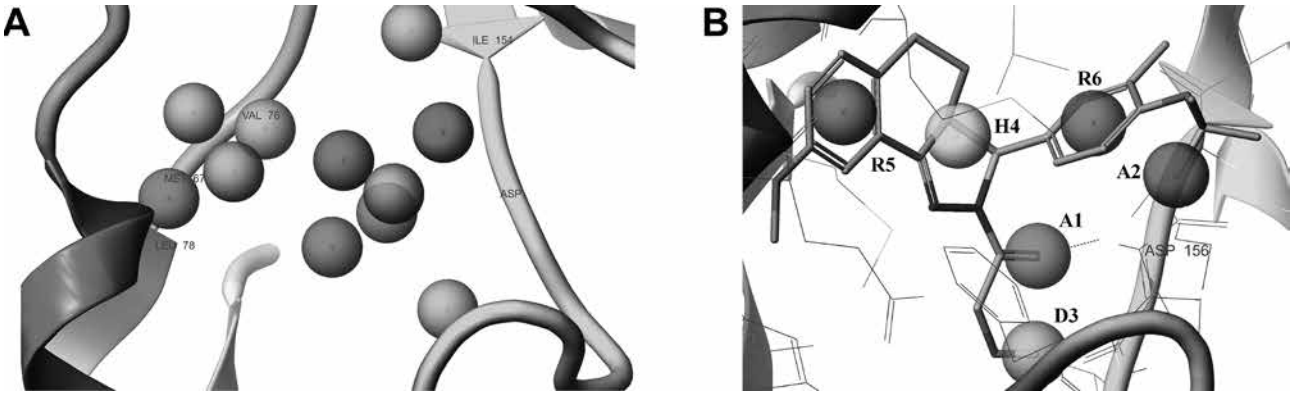

Figure 3: Structure-based pharmacophore derived from the (A) Binding site of a protein and (B) Protein-Ligand complex (A – Acceptor, D – Donor, R – Aromatic ring and H – Hydrophobic group).

acceptors, hydrophobic groups, aromatic rings, metal ligation, positive charge, and negative charge centres. A ligand binds to its receptor by establishing interactions such as hydrogen bonds (H-bonds). Therefore, the H-bond donors and the H-bond acceptors are important. Aromatic rings are necessary for π-π contacts. Similarly, hydrophobic centres are involved in hydrophobic contacts of ligand with its receptor. The positive charge centres and the negative charge centres act as the H-bond donors and the acceptors, respectively.

The pharmacophore features present on a ligand will be complementary to the features present in its receptor/biomolecular target-binding site and vice versa. Due to this complementarity, a ligand binds to its receptor. A pharmacophore model identified and derived from the receptor-binding site is known as the structure-based pharmacophore. These pharmacophore models can be used for virtual screening of compound libraries to identify novel ligands that might bind to a target structure from which the pharmacophore model has been derived.

Structure-based pharmacophore is the preferred choice, if the geometry/structure of the biomolecular target is known. If, the structure of the target is identified along with its ligand (target-ligand complex), then the ligand interactions with the binding site is known. If only the structure of the target is available, the pharmacophore model/hypothesis is identified from its binding site. If a target-ligand complex is used for deriving pharmacophore model, it contains features of the ligand with which it interacts with the receptor-binding site. Thus, the compounds containing these features could bind to this receptor. If the binding site of the target is used for pharmacophore model derivation, all the possible features, with which a ligand could interact, will be identified in the pharmacophore model. Thus, the compounds containing any combination of these features will be able to bind to this receptor-binding site.

One of the recent advancements in the structure-based pharmacophore extraction is the energy-optimized pharmacophore (e-pharmacophore), which is based on the energetic terms involved in the receptor-ligand interactions. In order to extract this pharmacophore, a fragment library will be docked initially to the receptor-binding site. The energetic descriptors of the docking score will be extracted and assigned to the pharmacophoric features. This pharmacophore hypothesis not only predicts all the available features present in the binding site, but also envisages the most favourable features in the pharmacophore hypothesis.

V. Ligand-based drug design

A. Importance

Ligand-based drug design relies on the knowledge of ligands that bind to a desired biological target molecule. In general, it is sensible to use target/receptor structure and the binding site information for drug discovery. However, in many cases, the target structure might not be available and could not be computationally modelled. In such circumstances, the structure-based drug design is not feasible. Hence, the ligand-based drug design is carried out in these cases. The ligand-based drug design chiefly involves the studies on ligand-based pharmacophore extraction and quantitative structure activity relationship (QSAR).

In some cases, the available target structures may not be complexed with a ligand or there might be only the information of ligands that module a biological target. In such cases, the ligand-based drug design is carried out to identify the key features of the ligands that are necessary for their binding to the target.

B. Ligand-based pharmacophore

The pharmacophore model/hypothesis derived from a set of ligands that bind to a biological target is known as ligand-based pharmacophore. The key advantage of this is that neither the receptor structure nor the information of the binding site is required. Here, it is assumed that since the ligands are binding to a common target receptor, they should possess common

features. Hence, the ligands that bind to a target are superposed and the common features among them are derived as a pharmacophore (Figure 4). However, in practice, the ligand-based pharmacophore extraction proves to be somewhat difficult. This is due to the reason that it is not known which conformer of a ligand is the active conformer. Further, if all the known ligands are derivatives of the same base structure/ scaffold, then they might contain many common features that may not be involved in binding with the receptor. For these reasons, it is best to derive ligand-based pharmacophore from a set of known active ligands that are derivatives of different structure/scaffold.

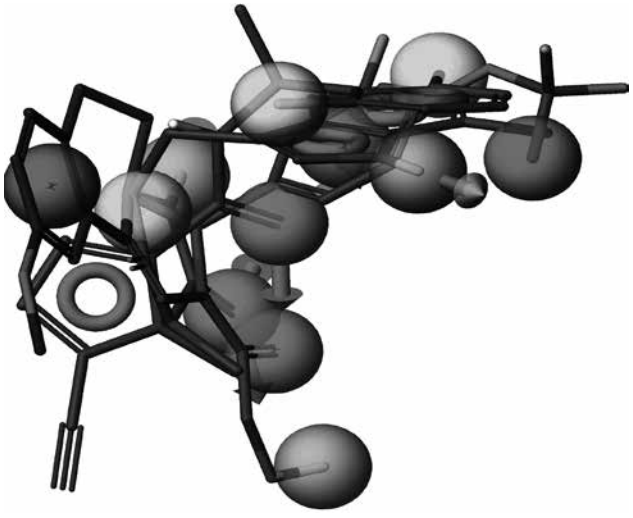

Figure 4: Ligand-based pharmacophore derived from the alignment of ligands of the same target.

The process of developing a ligand-based pharmacophore for a set of active ligands that bind to a target molecule involves generation of low energy conformations that might contain the bioactive conformations, superimposing these conformations, extracting similar functional groups that are common to all the ligands and representing these groups/features as a pharmacophore model. Finally, the model is validated by employing it to screen a set of compounds that contain both the active and inactive ligands. The pharmacophore model should be able to pick only the active compounds.

In general, the active compounds are initially divided in to two groups. The first group is considered the training set to derive the pharmacophore model and the second group is treated as a test set to test and validate the derived pharmacophore model.

C. QSAR

QSAR stands for "quantitative structure-activity relationship." The QSAR is used to quantitatively predict the activity of a set of given compounds from their structures. Through ligand-based pharmacophore, we could predict novel ligands that may bind to the desired receptor. However, we may not know the definite activity of these novel ligands towards their target. In such cases, experimental studies that determine the biological activities, such as Inhibitory Constant (IC_{50}) values, are required. Nevertheless, CDD has made this problem relatively simple by providing an important tool viz., QSAR. If we know the structures of the ligands that bind to a receptor and their biological activities, then we could predict the probable activities of novel ligands, whose experimental activity is unknown. If the structures and the biological activities of the compounds are known, then we can derive QSAR by correlating the properties of the compounds and their activities. These QSAR models might be used to predict the activity of the new analogs (Figure 5).

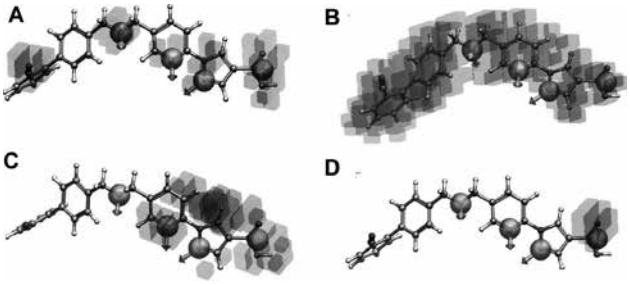

Figure 5: The QSAR visualization of various substituents of a molecule. (A) Electron withdrawing feature, (B) Hydrophobic features, (C) H-bond donor and (D) Negative ionic.

The QSAR is of two types, viz., 2D-QSAR and 3D-QSAR. The 2D-QSAR considers the structures of compounds in a planar axis. It derives only the structural components of the compounds from the X and Y-axes and correlates them with their activity. However, 3D-QSAR considers the third dimension viz., Z-axis also. It derives the structural components of the compounds from their conformations present in three

dimensions. Since the information regarding the conformations of the compounds is involved, 3D-QSAR is mainly used in biological applications to identify the activity of novel ligands.

Since the activity of novel ligands is predicted from the activities of already known ligands, the starting point to derive a QSAR model/ equation is to take the ligands that bind to a common receptor, along with their experimental activities. These ligands are then divided in to two parts viz., the training set and the test set. The training set is used to build the QSAR model and the test set is employed to test/validate the model.

VI Virtual screening

A. *Importance*

Experimental screening of a large number of compounds to identify the compounds that bind to a specific target is known as high-throughput screening (HTS). It involves screening of large chemical libraries to test their ability to modify a target molecule. The HTS is the most important source of new hits. Pharmaceutical companies have screening libraries up to a few million compounds. However, screening such a huge number of compounds experimentally is costly and time consuming. In this context, a more robust and convincing computational method known as virtual screening (VS) was introduced.

The VS involves *in silico* screening of compound libraries through methods such as docking [17]. In other words, it is a computational technique employed in drug discovery to identify small molecules that might bind to a target receptor [18-20]. These molecules will be less in number so that they can be synthesized, purchased, and tested. Therefore, the VS has been essentially involved in the drug discovery process [21]. The VS can be broadly categorized into ligand and structure-based [22-32]. Pharmacophore-based VS is both the ligand-based and the structure-based, whereas the docking-based VS is structure-based.

B. *Pharmacophore-based screening*

In the previous sections, the extraction and building of the ligand-based and the structure-based pharmacophore models have been described. Since drug designing involves the identification of novel target specific ligands through pharmacophore extraction, identifying the features that a ligand should possess in order to bind to the target of our interest is imperative. Now, the next step is to identify the compounds that possess these features. This is achieved through pharmacophore-based screening of a large set of compounds. The developed pharmacophore model is employed to search a database of compounds and the compounds that fit the pharmacophore model are ranked based on their fit. Finally, these compounds could be docked to study their binding affinities, provided the receptor structure is available. The different computational tools that could be used to derive pharmacophore models are given in the Table 3.

Table 3: Computational tools to identify/extract pharmacophore models

Sl No	Software/Company	Module
1	Open Eye Scientific Software	ROCS
2	LIQUID	LIQUID
3	Accelrys	Catalyst
4	Chemical Computing Group	MOE
5	Schrodinger	Phase
6	Tripos	DISCOtech
7	Inte: Ligand	LigandScout

C. *Docking*

In the SBDD, the binding affinity of the ligands with the target receptor is identified through docking, a robust and widely used *in silico* method [33]. The molecular docking could be considered "lock-and-key" phenomenon. Here, the protein is thought to be the "lock" and the ligand as a "key." The binding site of the target is considered a key hole. Through docking, one could find the correct relative orientation of the "key" that will open up the "lock" [34–36]. The output of docking involves scoring of ligands based on their probable binding affinities that are predicted through linear regression, machine learning, neural networks, semi-empirical, and *ab initio* quantum chemistry methods, or other statistical techniques [37, 38].

Based on the rate of flexibility involved, the docking method is generally categorized in to two types viz., rigid docking and flexible docking. In rigid docking, both the ligand and the receptor are rigid, whereas during flexible docking, both of them are flexible. However, incorporating flexibility for the receptor molecule during docking would be computationally robust. Therefore, the receptor is usually considered rigid. However, neglecting the flexibility of the receptor might lead to the poor docking results [39]. Hence, in order to simulate the reality as far as possible, few programs allow partial flexibility of the receptor. Such a docking process is known as induced fit docking, where the binding site region of the receptor is flexible [40–42]. Employing multiple structures, with varying conformations of the same protein is adopted in the recent docking methodologies and this process is known as the ensemble docking [43]. Thus, during most of the docking processes, the ligand would be flexible and the receptor would be completely rigid or partially flexible (induced fit docking).

To perform a docking, the three-dimensional structure of the biological target is required. Typically, the structure might have been determined and deposited in the structural databases such as PDB. If the structure of the receptor is not available, then it could be predicted through the methods described in section 4.1. The ligand molecules could be the compounds of our interest or could be taken from the databases such as Zinc database. If the compounds are from the databases, they might be initially subjected to pharmacophore-based screening and the resultant compounds may be used for docking. The ligand compounds might also be screened for their drug-like property and those compounds that pass this screening may be used for docking.

After considering the receptor and the ligand molecules, the next step is to prepare their structures to check, if there is any problem with the structures and to add missing atoms and polar hydrogen. Different possible conformations of the ligands could be generated. The next step is to generate a bounding box/grid around the binding site of the receptor, in order to guide the program where exactly it has to dock the ligands. The information regarding the selection of binding site is described in section 4.2. Then the ligand molecules are subjected to docking with the receptor molecule. If a large number of ligands are screened, this process is known as docking-based virtual screening. A number of free and commercial *in silico* tools are available for docking (Table 4). During the docking process, all the possible orientations/poses of the ligands may be searched and the best pose may be predicted through the search algorithm. The interactions of all these ligand poses with the receptor molecule are calculated through scoring function. Finally, the ligand poses with the best scores are given as the docking output. The success of a docking program depends mainly on two components viz., the search algorithm and the scoring function [44–47].

Table 4: Computational tools for Protein-Ligand docking

Sl No	Software/Module	License
1	AutoDock	Free
2	DOCK	Free for academics
3	eHITS	Free for academics
4	FlexX	Commercial
5	FRED	Free for academics
6	Glide	Commercial
7	GOLD	Commercial
8	ArgusLab	Free

VII. ADMET and drug-likeness

Drug-likeness is a qualitative concept used in drug discovery process to identify how a "drug-like" ligand/compound is with respect to bio-availability. Through pharmacophore or docking-based VS, the potential ligands are identified. However, not all these ligands will be translated to drugs, the reason being that certain properties, known as drug-like properties, are required for these compounds to act as the drugs. When a compound/drug is administered, it undergoes absorption (A), distribution (D), metabolism (M), excretion (E), and may even induce toxicity (T). The profiles that are required

for a drug are collectively known as ADMET properties. After VS, using ADMET properties, the ligand molecules could be predicted for their drug-likeness. Lipinski's rule of five (RO5) is commonly used to predict the drug-likeness of a compound [48–53].

Compounds that are obtained through VS are subjected to drug-likeness prediction by studying their ADMET properties. The compounds that satisfy the RO3 may be used for *in vitro*, *in vivo,* and in clinical studies.

VIII. Molecular dynamics simulation

The prediction of binding affinity of a ligand, through docking studies alone, has limitations since it does not reflect the actual receptor-ligand interactions inside the biological systems. This may be because the receptor and the ligand molecules are flexible and are surrounded by the solvent molecules (e.g. water) that may influence their interactions. Hence, the ligands that are identified through docking studies, for a particular biological target, might fail in further studies. In order to overcome the limitations of docking, an *in silico* method known as molecular dynamics (MD) simulation was introduced.

The MD simulations are in many respects very similar to the real experiments, since they involve the study of dynamics or flexibility of molecules in a real cell-like environment. Computational tools such as Gromacs, Amber, Charmm, and NAMD are available for MD simulation studies. At the beginning of simulation, the receptor-ligand complex is placed in a box of solvent. Then salt is added to neutralize the charge of molecules. This reflects the environment inside the biological cells. Then the positions and velocities of all the particles in the system are set to their initial values. As the simulation progresses, the relative positions and velocities of particles are predicted through the Newton's laws of motion. Force fields are used to calculate the different forces acting on each of the particles. A typical

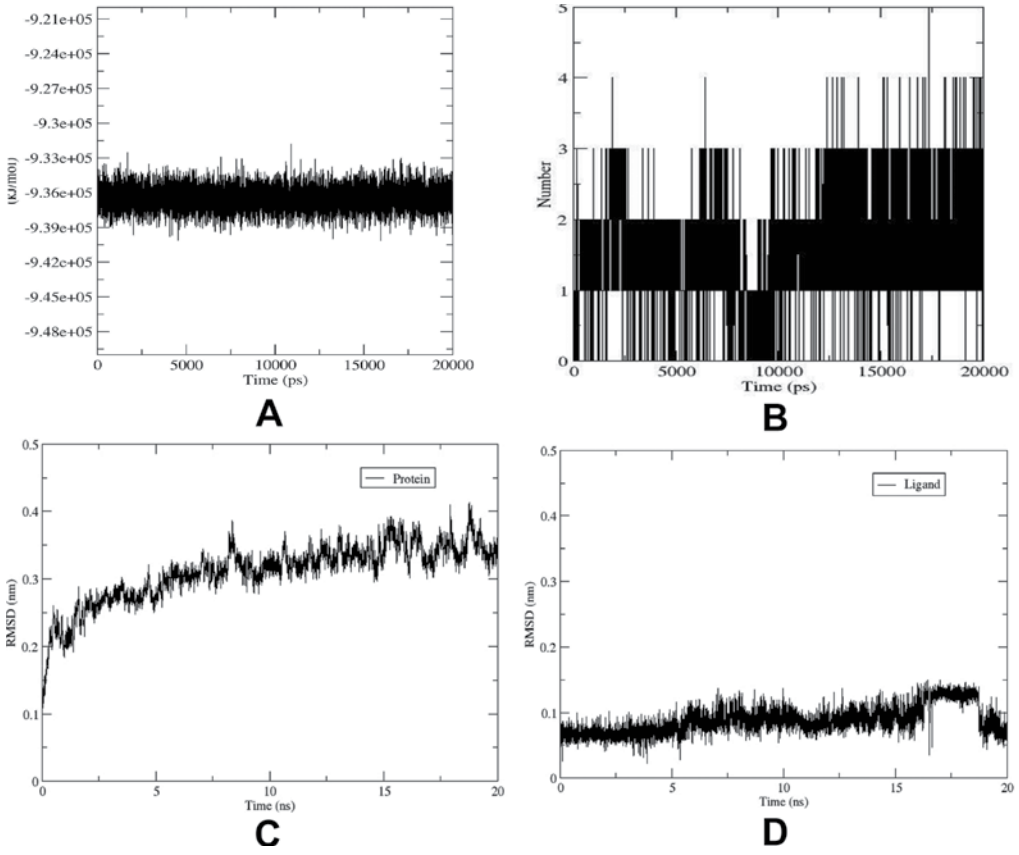

Figure 6: MD Simulation results showing (A) Potential energy of the Protein-Ligand complex, (B) H-bonds formed by the Ligand with the Protein, (C) RMSD of the Protein backbone during the course of simulation and (D) RMSD of the Ligand during the course of simulation.

MD simulation involves the following steps: The force field for simulation is selected and the topologies for receptor and the ligand molecules are generated. The system is neutralized by the addition of required number of Na^+ and Cl^- counter ions and a required concentration of NaCl salt. Further, it is equilibrated with NVT [moles (N), volume (V) and temperature (T)] and NPT [moles (N), pressure (P) and temperature (T)] ensemble equilibration protocol. Finally, extensive MD simulation is performed for N-ns (ns-nanoseconds).

The MD simulation results may be used to find whether the ligand binds to the desired receptor target or not. From the MD simulation, it may be analyzed if the ligands bind to the receptor with similar conformational modes as obtained through docking. Further, it can be employed to check if the ligand remained in the binding site throughout the simulation. The parameters such as potential energy and the Root Mean Square Deviation (RMSD) of the receptor backbone and ligand in the complexes must be steady and they must remain stable during the course of simulation (Figure 6).

IX. Real life applications and examples

Computational drug design techniques are extensively used in the real world. The QSAR is used in diverse studies other than drug design [54, 55, 56]. Docking is also used in diverse fields such as bioremediation [57]. It is mainly being used in the field of drug design. This was used to identify drugs against Angiotensin-converting enzyme (ACE), human immunodeficiency virus-1 protease (HIV-1 PR), the rennin-angiotensin system (RAS), and the NS3-NS4A serine protease of hepatitis C virus (HCV) [58–65].

X. Conclusions

Drug designing is a continuous ongoing process for the treatment of various diseases and for a better healthcare. It is an integral tool to enable the development of various drug candidates. Several biotechnology companies have clearly documented the success of adopting computational drug design in the discovery process. Many pharmaceutical companies have moved from traditional drug discovery approaches to computational strategies. The ligand and the structure-based drug designing approaches could be employed to reduce the wastage of time and the resources. Further, the prediction of early stage ADMET properties might contribute to reduction in the late-stage failure of the drug candidates. Therefore, computational drug design approaches would enormously improve the field of healthcare.

References

[1] M. Ulf, P. Krogsgaard-Larsen and L. Tommy, "Textbook of Drug Design and Discovery," Washington, DC: *Taylor & Francis*, 2002.

[2] T. Takenaka, "Classical vs reverse pharmacology in drug discovery," *BJU Int*, Vol. 88, pp. 7–10, 2001.

[3] J. A. Lee, M. T. Uhlik, C. M. Moxham, D. Tomandl and D. J. Sall, "Modern phenotypic drug discovery is a viable, neoclassic pharma strategy," *J Med Chem,* Vol. 55, pp. 4527–4538, 2012.

[4] D. C. Swinney, J. Anthony, "How were new medicines discovered?" *Nat Rev Drug Discov*, Vol. 10, pp. 507–19, 2011.

[5] D. Anson, J. Ma and J. Q. He, "Identifying Cardiotoxic Compounds," *Genetic Engineering & Biotechnology News*, Vol. 29, pp. 34–35, 2011.

[6] S. M. Paul et al., "How to improve R & D productivity: The pharmaceutical industry's grand challenge," *Nat Rev Drug Discov*., Vol. 9, pp. 203–214, 2010.

[7] J. Singh, C. E. Chuaqui and P. A. Boriack-Sjodin, "Successful shape-based virtual screening: the discovery of a potent inhibitor of the type I TGF-beta receptor kinase (TbetaRI)," *Bioorg Med Chem Lett*., Vol. 13, pp. 4355–59, 2003.

[8] O. M. Becker, D. S. Dhanoa and Y. Marantz, "An integrated in silico 3D model-driven discovery of a novel, potent, and selective amidosulfonamide 5-HT1A agonist (PRX-00023) for the treatment of anxiety and depression," *J Med Chem.,* Vol. 49, pp. 3116–3135, 2006.

[9] A. R. Leach and J. Harren, Structure-based Drug Discovery. Berlin: *Springer*, 2007.

[10] R. Wang, Y. Gao, L. Lai, "LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design," *J Mol Model,* Vol. 6, pp. 498–516, 2000.

[11] G. Schneider and U. Fechner, "Computer-based de novo design of drug-like molecules," *Nat Rev Drug Discov*, Vol. 4, pp. 649–663, 2005.

[12] W. L. Jorgensen, "The many roles of computation in drug discovery," *Science*, Vol. 303, pp.1813–8, 2004.

[13] O. F. Guner, "Pharmacophore Perception, Development, and use in Drug Design," *La Jolla, Calif: International University Line*.

[14] C. Tim, "The Identification of Bioisosteres as Drug Development Candidates," *Cresset Biomolecular Discovery*.

[15] A. R. Leach, "Molecular modelling: Principles and applications," Englewood Cliffs, *N.J: Prentice Hall*, 2001.

[16] F. Ballante and R. Ragno, "3-D Q S Autogrid/R: an alternative procedure to build 3-D QSAR models. Methodologies and applications," *J Chem Inf Model*, Vol. 52, 1674–1685, 2012.

[17] W. P. Walters, M. T. Stahl and M. A. Murcko, "Virtual screening – an overview," *Drug Discov Today*, Vol.3, pp.160–178, 1998.

[18] U. Rester, "From virtuality to reality - Virtual screening in lead discovery and lead optimization: A medicinal chemistry perspective," *Curr Opin Drug Discov Devel*, Vol.11, pp. 559–568, 2008.

[19] J. M. Rollinger, H. Stuppner and T. Langer, "Virtual screening for the discovery of bioactive natural products," *Prog Drug Res*, Vol. 65, pp. 213–249, 2008.

[20] R. S. Bohacek, C. Mc Martin and W. C. Guida, "The art and practice of structure-based drug design: A molecular modelling perspective," *Med Res. Rev.*, Vol.16, pp. 3–50, 1996.

[21] M. J. McGregor, Z. Luo and X. Jiang, "Virtual screening in drug discovery In Huang, Ziwei," *Drug Discovery Research. New Frontiers in the Post-Genomic Era.* Wiley-VCH: Weinheim, Germany, pp. 63–88, 2007.

[22] C. McInnes, "Virtual screening strategies in drug discovery," *Curr Opin Chem Biol*, Vol.11, pp. 494–502, 2007.

[23] T. S. Rush, J. A. Grant, L. Mosyak and L, A. Nicholls, "A Shape-Based 3-D Scaffold Hopping Method and Its Application to a Bacterial Protein–Protein Interaction," *J Med Chem*, Vol. 48, pp. 1489–1495, 2005.

[24] P. J. Ballester, I. Westwood, N. Laurieri, E. Sim and W. G. Richards, "Prospective virtual screening with Ultrafast Shape Recognition: the identification of novel inhibitors of arylamine N-acetyltransferases," *J R Soc Interface*, Vol. 7, pp. 335–342, 2010.

[25] P. Willet, J. M. Barnard and G. M. Downs, "*Chemical similarity searching,*" *J Chem Inf Comput Science*, Vol. 38, pp. 983–996, 1998.

[26] H. Sun, "Pharmacophore-based virtual screening," *Curr Med Chem.*, Vol.15, pp.1018–1024, 2008.

[27] R. T. Kroemer, "Structure-based drug design: Docking and scoring," *Curr Protein Pept Sci*., Vol. 8, pp. 312–328, 2007.

[28] C. N. Cavasotto and A. J. Orry, "Ligand docking and structure-based virtual screening in drug discovery," *Curr Top Med Chem*., Vol.7, pp. 1006–1014, 2007.

[29] J. Irwin, "Community benchmarks for virtual screening", *J Comput. Aided Mol. Des*., Vol.2, pp. 193–199, 2008.

[30] A. C. Good and T. I. Oprea, "Optimization of CAMD techniques 3. Virtual screening enrichment studies: A help or hindrance in tool selection?" *J Comput. Aided Mol. Des*., Vol. 22, pp. 169–178, 2008.

[31] G. Schneider, "Virtual screening: an endless staircase?" *Nat Rev Drug Discov*., Vol. 9 pp. 273–276, 2010.

[32] P. J. Ballester, "Ultrafast shape recognition: Method and applications" *Future Med Chem*, Vol. 3, pp. 65–78, 2011.

[33] D. B. Kitchen, H. Decornez, J. R. Furr and J. Bajorath, "Docking and scoring in virtual screening for drug discovery: Methods and applications," *Nat. Rev. Drug Discov.*, Vol.3, pp. 935–949, 2004.

[34] W. L. Jorgensen, "Rusting of the lock and key model for protein-ligand binding," *Science*, vol. 254, pp. 954–5, 1991.

[35] B. Q. Wei, L. H. Weaver, A. M. Ferrari, B.W. Matthews and B. K. Shoichet, "Testing a flexible-receptor docking algorithm in a model binding site," *J Mol. Biol*., Vol. 337, pp. 1161–1182, 2004.

[36] T. Lengauer and M. Rarey, "Computational methods for biomolecular docking". *Curr. Opin. Struct. Biol.*, Vol. 6, pp. 402–6, 1996.

[37] R. Rajamani and A. C. Good, "Ranking poses in structure-based lead discovery and optimization: Current trends in scoring function development," *Curr. Opin. Drug Discov. Devel*, Vol.10, pp. 308–315, 2007.

[38] W. F. de Azevedo and R. Dias, "Computational methods for calculation of ligand-binding affinity," *Curr. Drug Targets*, Vol.9, pp.1031–9, 2008.

[39] N. M. Cerqueira, N. F. Bras, P. A. Fernandes and M. J. Ramos, "MADAMM: A multistaged docking with an automated molecular modelling protocol," *Proteins,* Vol. 74, pp. 192–206, 2009.

[40] C. Hartmann, I. Antes and T. Lengauer, "Docking and scoring with alternative side-chain conformations," *Proteins*, Vol.74, pp.712–726, 2009.

[41] R. D. Taylor, P. J. Jewsbury and J. W. Essex, "FDS: Flexible ligand and receptor docking with a continuum solvent model and soft-core energy function," *J. Comput. Chem.*, Vol. 24, pp. 1637–1656, 2003.

[42] M. Totrov and R. Abagyan, "Flexible ligand docking to multiple receptor conformations: A practical alternative," *Curr Opin Struct Biol*, Vol.18, pp.178–84, 2008.

[43] S. M. Fayaz and G. K. Rajanikant, "Ensemble pharmacophore meets ensemble docking: A novel screening strategy for the identification of RIPK1 inhibitors", *J. Comput. Aided. Mol. Des.,* Vol. 28, pp.779–94, 2014.

[44] B. B. Goldman and W. T. Wipke, "QSD quadratic shape descriptors. 2. Molecular docking using quadratic shape descriptors (Q S Dock)," *Proteins,* Vol. 38, pp.79–94, 2000.

[45] E. C. Meng, B. K. Shoichet and I. D. Kuntz, "Automated docking with grid-based energy evaluation," *J. Comput. Chem*., Vol. 13, pp. 505–524, 2004.

[46] G. M. Morris et al., "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function," *J. Comput. Chem*., Vol.19, pp.1639–1662, 1998.

[47] M. Feig, A. Onufriev, M. S. Lee, W. Im, D. A. Case and C. L. Brooks, "Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures," *J. Comput. Chem.*, Vol. 25, pp. 265–284, 2004.

[48] C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Adv. Drug. Deliv. Rev*., Vol. 46, pp. 3–26, 2001.

[49] C. A. Lipinski, "Lead and drug-like compounds: The rule-of-five revolution," *Drug Discov Today Technol*., Vol. 1, pp. 337–341, 2004.

[50] T. I. Oprea, A. M. Davis, S. J. Teague and P. D. Leeson, "Is there a difference between leads and drugs? A historical perspective," *J. Chem. Inf. Comput. Sci*., Vol. 41, pp. 308–15, 2001.

[51] P. D. Leeson and B. Springthorpe, "The influence of drug-like concepts on decision-making in medicinal chemistry," *Nat Rev Drug Discov.*, Vol. 6, pp. 881–890, 2007.

[52] A. K. Ghose, V. N. Viswanadhan and J. J. Wendoloski, "A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery 1. A qualitative and quantitative characterization of known drug databases," *J. Comb. Chem*., Vol.1, pp. 55–68, 1999.

[53] M. Congreve, R. Carr, C. Murray and H. Jhoti, "A 'rule of three' for fragment-based lead discovery?" *Drug. Discov. Today*, Vol. 8, pp. 876–7, 2003.

[54] W. Tong, H. Hong, Q. Xie, L. Shi, H. Fang and R. Perkins, "Assessing QSAR Limitations – A Regulatory Perspective," *Curr. Comput. Aided Drug. Des.* Vol.1, pp.195–205, 2005.

[55] J. C. Dearden, "In silico prediction of drug toxicity," *J. Comput. Aided Mol. Des.,* Vol.17, pp.119 – 127, 2003.

[56] E. K. Freyhult, K. Andersson and M. G. Gustafsson, "Structural Modelling Extends QSAR Analysis of Antibody-Lysozyme Interactions to 3D-QSAR," *Biophys.* J., Vol. 84, 2264–2272, 2003.

[57] P. S. Suresh, A. Kumar, R. Kumar and V. P. Singh, "An *in silico* approach to bioremediation: Laccase as a case study," *J. Mol. Graph Model*, Vol. 26, pp. 845–849, 2008.

[58] J. Greer, J. W. Erickson, J. J. Baldwin and M. D. Varney, "Application of the three-dimensional structures of protein target molecules in structure-based drug design," *J. Med. Chem.,* Vol. 37, pp.1035–1054,1994.

[59] T. Hendrik, G. Klaus, B. Hans-Joachim, M. Raimund and K. Hugo, "Structure-based Ligand Design" (Methods and Principles in Medicinal Chemistry). *Weinheim: Wiley-VCH.* 1998.

[60] D. W. Cushman, H. S. Cheung, E. F. Sabo and M. A. Ondetti, "Design of potent competitive inhibitors of angiotensin converting enzyme. Carboxyalkanoyl and mercaptoalkanoyl amino acids," *Biochemistry*, Vol.16, pp. 5484–5491, 1977.

[61] M. A. Ondetti, B. Rubin and D. W. Cushman, "Design of specific inhibitors of angiotensin-converting enzyme: New class of orally active antihypertensive agents," *Science*, Vol. 196, pp. 441– 444, 1997.

[62] M. Von Itzstein et al., "Rational design of potent sialidase-based inhibitors of influenza virus replication," *Nature*, Vol. 363, pp. 418–423, 1993.

[63] P. J. Goodford, "A computational procedure for determining energetically favorable binding sites on biologically important macromolecules," *J. Med. Chem.*, Vol. 28, pp. 849–857, 1985.

[64] J. M. Wood et al., "Structure-based design of aliskiren, a novel orally effective renin inhibitor," *Biochem. Biophys. Res. Commun.*, Vol. 308, pp. 698–705, 2004.

[65] F. G. Njoroge, K. X. Chen, N. Y. Shih and J. J. Piwinski, "Challenges in modern drug discovery: A case study of boceprevir, an HCV protease inhibitor for the treatment of hepatitis C virus infection," *Acc, Chem, Res.,* Vol. 41, pp. 50–59, 2008.