Methods 71 (2015) 135–145

Contents lists available at ScienceDirect

Methods

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

Structure based virtual screening to discover putative drug candidates: Necessary considerations and successful case studies

Mohd Danishuddin, Asad U. Khan *

Interdisciplinary Biotechnology Unit, Aligarh Muslim University, Aligarh 202002, India

ARTICLE INFO

Article history:
Received 29 May 2014
Received in revised form 25 September 2014
Accepted 17 October 2014
Available online 27 October 2014

Keywords:
Virtual screening
Docking
Chemical databases
Target preparation
Scoring function

ABSTRACT

Drug discovery faces daunting challenges in the current economic situation, which is further exacerbated by resistance against a large group of available drugs. Development of a new drug with traditional approaches generally takes 12–15 years and may cost over $800 millions. Therefore, inexpensive and fast alternatives are required for new drug discovery. Various *in silico* approaches have shown potential for screening chemical databases against the desired biological targets for the development of new potential leads. Among them, the number of publications on structure based virtual screening has been rapidly mounting in recent years. This increase has led a need to evaluate and compare the performance of different virtual screening methodologies. In the present article, we describe some of the work and addresses the important issues for successful structure-based virtual screening. Moreover, few recent case studies are also discussed, where the virtual screening approaches have been applied successfully in designing putative drug candidates.

1. Introduction

The discovery and development of new drugs are highly complex, time consuming and expensive that has faced many changes and challenges over the past few decades [1]. Traditional approaches for drug development may take around 12–15 years from initiation of work on a target to final registration. High-throughput screening (HTS) adopted by many pharmaceutical industries and academia or other government aided agencies had provided substantial growth toward the discovering new therapeutic molecules [2]. The continuous development of combinatorial chemistry and high-throughput screening (HTS) has accelerated the drug discovery process by enabling huge libraries of compounds to be screened in a short time. However, due to the high cost and low hit rate of HTS, virtual screening is gaining prominence as a method to filter the very large chemical space that can be created by combinatorial synthesis into a manageable library of likely chemical ligands, or more generally, favorable scaffolds to be tested by HTS [3]. This field has become more popular, and experienced a rapid growth in pharma industries and academia.

Virtual screening is a method to distinguish molecules based on a desired property and can be applied for the identification of novel potential leads, their optimization and scaffold hopping. It can be divided into two broad categories: (1) structure based and (2) ligand based virtual screening [4]. Structure based virtual is believed to be more effective at detecting novel chemical scaffolds and more commonly use in academic labs. It is highly useful when the information about the target structure is available. Biological target structure can be selected from X-ray crystallography and NMR solved structures or can be determined through theoretical modeling (homology modeling) to sample the candidate molecules within its active site. Ligand based virtual screening approaches utilize the structure–activity data information from known active molecules with the aim of identifying structurally diverse compounds having similar bioactivity [3]. Further, this method includes approaches like; similarity, QSAR (Quantitative Structure Activity Relationships) and pharmacophore modeling. The disadvantage of LBVS (Ligand Based Virtual Screening) is that it tends to discover molecules by taking information from existing ligands as templates usually, resulting in more limited scaffold diversity in comparison to structure based virtual screening [5]. However, increases efficiency and reliability of computational tools have enabled them to become routine methods in pharmaceutical drug discovery, complementing *in vitro* high-throughput screening [3,6,7]. Survey on pubmed provides the data that shows the rapid increment and development in the quantum of publications on virtual screening in the last 10 years (Fig. 2A–C). Docking with >54% was the preferred method among the academic community for drug discovery throughout the last decade. In the past five years, the number of publications with experimental evaluation of virtual screening methods was also significantly increasing (Fig. 2B).

* Corresponding author.
E-mail address: asad.k@rediffmail.com (A.U. Khan).

http://dx.doi.org/10.1016/j.ymeth.2014.10.019
1046-2023/

However, despite the considerable progress achieved, there is a lack of common evaluation method which can be employed to predict higher accuracy and hence widely accepted by the academic researchers. Thus, in addition of frequent use, there are certain issues associated with the performance of different virtual screening approaches and their application potential. Therefore, in order to avoid such complications it has eventually become a challenge to choose a successful combination of methodologies for efficient drug discovery.

In this article we emphasized the recent progress of various structure based virtual screening methods with a particular focus on the target preparation, ligand preparation and pose prediction. Furthermore, we have also discussed some of the recent case studies where the virtual screening technology has been applied successfully.

2. Chemical databases

Computational chemical spaces are in the stage of a revolution. Such chemical spaces provide the reference frame, helpful in designing the compound data sets of finite size. Long dominated by a handful of established players, the field has rather suddenly opened up to a variety of innovative ideas from newcomers. Efforts of the NIH Molecular Libraries Roadmap Initiative made possible for the academic scientific community to perform large scale chemical screening against the hundred of biological targets. Online databases has revolutionised and offers a diverse array of free knowledge based chemistry resources for efficient virtual screening. These databases often contain the vast amount of small molecules, easier to optimize toward the identification of promising lead compounds. Pubchem (http://PubChem.ncbi.nlm.nih.gov) was initially introduced in 2004 as a scientific platform of the molecular libraries program. The main objective of this database is to enhance chemical biology efforts through HTS to identify small probes for the given biological target. The database holds >72 millions substances and >30 millions unique structures, whereas the pubchem bioassays have information on >500,000 records. ZINC is another free database of commercially-available compounds for virtual screening [8]. Compounds in this database are prepare in multiple protonation states and multiple tautomeric forms along with vendor information for purchase. Different subsets for compounds like; fragment, drug like and lead like have been catalogued in this database that can be accessed in four different formats (SDF, Mol2, 3D, SMILE). Similarly, DrugBank [9] holds the information on many chemicals and FDA approved drugs along with detailed biological targets information. This database has been used to facilitate the drug designing, drug interaction prediction and target discovery. Discovery of new drug targets is a crucial for both the pharmaceutical industry and academic research. Accounting this, Andersen et al. has studied the interactions of US Food and Drug Administration approved drugs with therapeutic targets encoded by human genome [10]. With this study they identified the 435 effect-mediating drug targets which are modulated by 989 unique drugs, through 2242 drug–target interactions. The story does not end here rather other chemical resources are available where information can be valuable for the users [11–13]. Some of the important databases, holding information on millions of compounds, were summarized in Table 1. This list has been updated with new information and new entries which can be useful for a future moves.

3. Structure based virtual screening

The discovery of novel small molecules with potential interaction against the targets of interest is critical in the early drug discovery process. During the past decade where high throughput screening (HTS) method has continued to generate disappointing results, structure based virtual screening become an alternative approach in early drug discovery process. To achieve this, one should necessarily have three dimensional structure of the desire target in hand. To date, two experimental techniques such as X-ray crystallography and NMR (Nuclear Magnetic Resonance) have been applied to determine the structure of target. X-ray crystallography is the preferred structural tool for the determination of three-dimensional structure of protein. The statistics of Protein Data Bank shows the three dimensional structures where X-ray crystallography count over 80% and NMR count over 16%. In cases where target structures are not available one can resort to techniques like homology modeling or threading to generate an approximate structure [14]. Active sites of a receptor are known from an experimentally solved protein–ligand complexes. However, in absence of these, sometime the active site or additional binding sites has to be predicted or determined theoretically. It is often assumed that the largest cleft on a protein surface corresponds to the active site, but this is not always the case. The annotation of binding sites allows for the identification of similar or identical binding sites between related targets, that is an important step for lead optimization. If a keyword, "protein binding site predictions" is given on Google search; several methods may be available for predicting the binding site. Among those successful methods, some of them are worth mentioning here like ConCavity [15], FINDSITE [16] and COFACTOR [17]. Computational tools for the identification and analysis of binding sites able to exploit protein structure data in order to facilitate the design of molecules. Results from these methods may provide the suggestions for the modulation of ligand properties. Prediction of the binding site is an important part of docking as blind docking over the entire protein surface results in much larger run time (limiting the size of the screening library) and may cause a substantial loss in accuracy, especially in grid based docking programs like DOCK [18]. Once selected, target structure can be passed to the docking protocol. Molecular docking is a computational technique that places a small molecule in the binding pocket of a receptor (DNA or Protein) and estimates its binding affinity. A docking method has two essential components: (1) Sampling: Refer to the generation of putative ligand binding orientations with the active site of the target. (2) Scoring: Refer to the evaluation of target–ligand complexes using different scoring functions [19,20]. Top ranked ligand with low binding energy is considered as binding mode. A key requirement for the success of structure based virtual screening or docking is the ability of the combination of a method and scoring function to rank active molecules early in a large set of compounds. Matrices like ROC, AUC and Enrichment Factor (EF) etc are currently used to evaluate the performance of the ranking methods in virtual screening. A ROC curve is a graphical plot created by plotting the fraction of the true positive rate (TPR or Sensitivity) versus the false positive rate (FPR, or 1− specificity) at various thresholds. The AUC–ROC (Area Under Curve–ROC) can be used to quantify the overall quality of the plot. It is a way to measure that how randomly selected active molecules are ranked compared to randomly chosen inactive molecules. A perfect value will result in an area under the curve of 1, while a random scoring function will have an ROC–AUC of 0.5 [21,22]. These matrices are very helpful in selecting the most promising compounds in a more objective way and concentrate their efforts on the synthesis of molecules that are more likely to be active against the investigated target. Despite these, success rate of the structure based virtual screening is also depended on data preparation and thus needed a necessary attention before performing the docking simulations [23]. The flowchart of structure based virtual screening is outlined in Fig. 1. As a process it can be divided into three basics steps. The

Table 1
Some of the chemical databases available for virtual screening.

Database	Records	Web address	Publications[a]
ZINC	80 million compounds, 1500 drugs	www.zinc.docking.org	62
PubChem	>72 million substances and >30 million unique structures	pubchem.ncbi.nlm.nih.gov	30
DrugBank	7678 drug entries including 1555 FDA-approved drugs	www.drugbank.ca	19
Maybridge	58,742	www.maybridge.com	29
NCI	>265,000	http://cactus.nci.nih.gov/	54
ChEMBL	1,520,172	www.ebi.ac.uk/chembl	13
DrugPort	1492 approved drugs	www.ebi.ac.uk/thornton-srv/databases/drugport/	ND
Pubchem	>500,000 descriptions	pubchem.ncbi.nlm.nih.gov/	ND
Bioassay	5000 protein targets >130 million bioactivity-outcomes	assay/assay.cgi	

ND: not determined.
[a] Five years record based on keyword search in pubmed. Search is depend on keyword "database name and virtual screening" in pubmed.

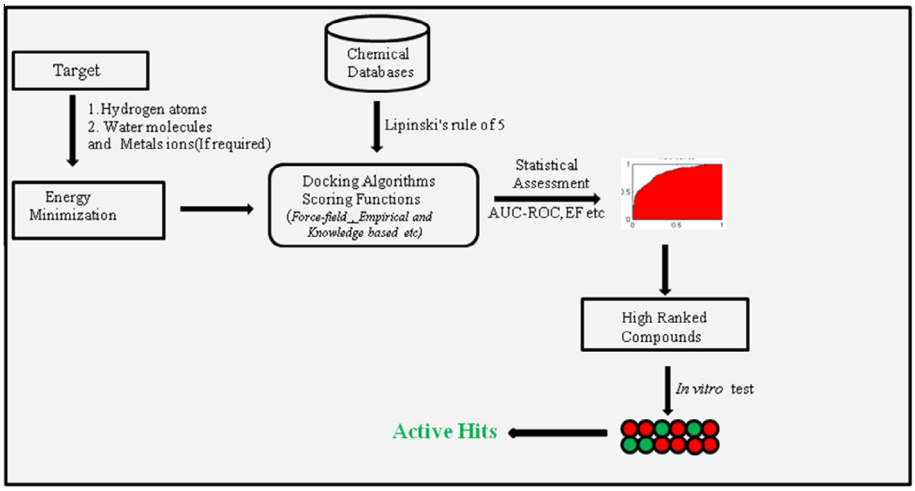

Fig. 1. Flow chart for structure based virtual screening.

first stage is preparation of the target structure and compounds library. On second step selected compounds set is docked to the target binding site. Finally, they are ranked and top hits are selected for their experimentally conformation.

4. Target preparation

Choice of the appropriate target structure plays an important role in structure based drug designing. Structures determined from crystal data with better resolution, *R* factor ~30% and over 90% backbone angles in most favoured region of Ramachandran plot are preferred for docking experiments. Moreover, the selected structure also capable to accommodate all relevant compound classes. In the absence of crystal structure one can use the protein structure prediction techniques i.e. homology modeling and threading, an important alternative approach to experimental structure determination techniques [24]. Comparative or homology modeling is known to be the most useful method to build the three dimensional structure of protein sequences that share >35% similarity with experimentally (X-ray or NMR) solved structures (served as a template). However, these theoretical models have some bad steric clashes and lack the proper strength for non-covalent bonding with the ligand, thus need to be removed prior to docking. Several energy minimization programs can be used to address this problem [25–27]. One can also use the option of the available target databases [9,28–30]. These databases are rich in diverse datasets and facilitate the identification and prioritization of drugs and drug targets in neglected disease pathogens. They can function both as a website, and as a tool for prioritization of the targets in whole genome. Others important factors address in structure based virtual screening are: target flexibility, incorporating water molecules or metals ions and selection of appropriate scoring function. Such challenging issues, focused on docking simulations have been reviewed in previous published studies [31,32] and also briefly explained here.

4.1. Target flexibility

Proteins are dynamics in nature that experience conformational changes during ligand binding due to induced fit effects and are vital to the protein own function [33]. Protein flexibility is essential to increased affinity between the target and drug. Several strategies are available to account the protein flexibility at a partial levels, already implemented in many protein–ligand docking programs. These strategies account for side chain flexibility only whereas inclusion of backbone flexibility is still a challenge. Considering target flexibility fully would lead to the optimal solution, however, it needs a high computational cost which may be one of the major practical challenges. A small deviation of 1 Å in protein backbone can significantly affect the protein–ligand interactions. Experimental techniques such as NMR and X-ray crystallography are commonly used to obtain conformations of the proteins. But these techniques usually provide snapshots of one or just some of the conformations of the proteins. To come up with this limitation, molecular dynamics simulation can be used to obtain a more complete set of protein conformers. The first molecular dynamics simulation was started with 10 ps [34], compared to the

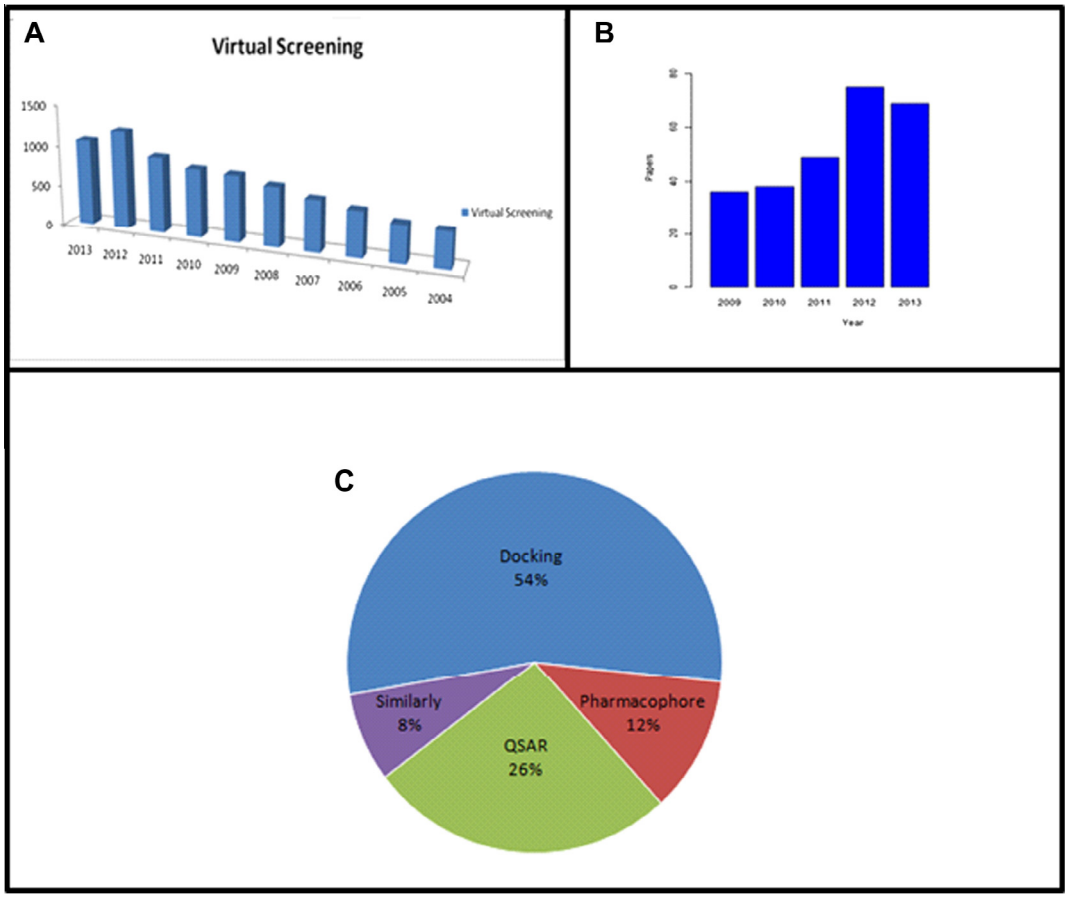

Fig. 2. (A) Depict the published papers on virtual screening in last 10 years. (B) Publications with experimental evaluation in last 5 years. (C) Pie chart depicts publications of different methods on virtual screening in last 10 years.

computational advancement of today's world where it can go up to >1000 ns [35]. Starting with the experimentally or theoretically solved structure, molecular dynamics simulation ended with large number of conformations at a given temperature and time scale. Sampling of these conformations provides a set of unbiased structures for analysis with molecular docking, virtual screening or other available *in silico* approaches. Major limitation is the accuracy of simulation, as the structure will tend to be driven away from the native structure during the simulation. In earlier attempts, an algorithm called the ICM-flexible receptor docking algorithm (IFREDA) was developed to account the protein flexibility during the virtual screening process [36]. It generates the discrete set of target conformations which were used for rigid receptor–ligand docking. This method was validated on four different protein kinases subfamilies (cAPK, CDK-2, P38 and LCK). With their analyses they found the reasonable increase in an enrichment factor as comparable to that obtained by using multiple experimental conformations. The advantage in using IFREDA approach is that it accounts both the side-chain rearrangements and essential backbone movements, thus sampling adequately the conformational space of the target, even in cases of large loop movements.

Despite of various progresses, target flexibility is still treated as less advance as compare to ligand flexibility. Versatile approaches such as Soft Docking [37], Monte Carlo calculations [38] and protein ensemble grids can be used for partial flexibility treatment. Soft docking allows a small degree of overlap between the protein and ligand, which is actually more than conventional docking [39]. The main advantage of this method is its simple implementation and computational efficiency. Unfortunately, soft docking can address only small conformational changes is its major disadvantage, whereas an ensemble method may consider the flexibility of the receptor upon binding the ligand by using multiple protein conformations instead of a single conformation. An ensemble of structures generated with molecular dynamics simulation or others experimental techniques (X-ray or NMR) has a significant role over the single conformation of a receptor. In one approach, Multi-CopyMD developed by Okamoto et al., was used to dock the multiple ligands in the binding site, simultaneously [40]. Using this approach they discovered highly potent and selective inhibitors for DAPK1 (death-associated protein kinase). Earlier, Rueda et al. [41] recommended the use of target conformer co-crystallized with ligand for better performance. They used the large and diverse benchmark data set of 1068 X-ray protein conformations and compare the performance of the ensemble and single-conformation docking for differentiating the binders from decoys. Finally, they concluded that the conformers co-crystallized with the largest ligands showed high selectivity for binders, and when combined in ensembles the results were improved as compared to randomly chosen protein conformations.

ReFlexIn (Receptor Flexibility by Interpolation), is a novel flexible docking method in which the potential grid for docking is done by linear interpolation of closely related structures [42]. Such methods allow the efficient and systematic evaluation of many conformations in docking by allowing a smooth and continuous deformation of the receptor structure and have been successfully tested on HIV-1 with both binders and non-binders [42]. This method resulted in good agreement while compared with the experimentally solved structure. In cases where best performing bound structure is not known in advance, it is beneficial to use the ReFlexIn method instead of docking to each individual receptor

structure. However, methods treating protein flexibility are computationally expensive and time consuming. To combat these costs, Paris and co-workers introduced a new method wFReDoW (web flexible receptor docking workflow) to handle molecular docking simulations of Fully Flexible Receptor (FFR) models [43]. The strength of this method is its cloud based web environment to handle the molecular docking simulations of fully flexible receptor models by using two approaches cooperatively, P-SaMIs (self-adaptive multiple instances) [44] and middleware built on Amazon Elastic Compute Cloud (Amazon EC2) instances. P-SaMI reduces the number of molecular docking simulations whereas the middleware is helpful in speeding the docking experiments using a high performance Computing (HPC) environment on the cloud. This method was found to reduce the computational time in the molecular docking simulations by selecting the promising protein conformation from FFR models.

4.2. Water molecules

Some proteins or enzymes contain significant water molecules in their binding sites. Such water molecules are resolved by crystallography and conserved between similar structures. These molecules in the active site often play an important role in target–ligand interaction and should therefore consider in computational drug design [45]. Water molecules mediate interaction by forming the hydrogen bonds between the protein and ligand. Unsuccessful docking, defined as the top scoring pose is greater than 2 Å from the experimentally solved structure, could in many cases be traced to neglect the bound water molecules in the docking run [46]. Appropriate placement of these water molecules is a non-trivial task as the network of bound waters can be appreciably different in the apo and holo-structures. Therefore, accounting water molecules accurately in docking is a long standing challenge. Several studies have highlighted this issue. It is necessary to retain these important water molecules within the target in order to achieve an appropriate binding orientation of the ligand. Docking programs like GOLD [47], FlexX [48] and AutoDock [49] implemented different procedure to position water molecules. Recently, Huggins and co-workers introduced the dead-end elimination method in order to place the water molecules in binding site [50]. This method places waters molecules correctly and create reasonable hydrogen bonds within the target–ligand complexes. WaterDock developed by Ross et al. [51] in combination with AutoDock Vina [52] predict 88% consensus water molecules correctly. This method has been validated from X-ray crystallography, neutron diffraction and molecular dynamics simulation data. The main advantage of their method is that it can also predict the water locations which are displaced by oxygen or nitrogen atoms. Amadasi and co-workers [53] combined methods HINT [54] and Rank score [55] and classified the water molecules in two categories; conserved displaced and sterically displaced. This hybrid approach correctly predicted the 76% and 87% of conserved and sterically disperse water molecules respectively. However, these water molecules were weakly bound which were found 4 Å away from the protein.

Rossato et al. proposed a directional approach, AcquaAlta and derived a geometric criterion defining the interactions of water molecules with protein and ligand [56]. Validating thoroughly on 20 crystal structures, method found to predict the 76% of crystallographic molecules correctly. Water molecules having weak interactions with the protein not consider in this analysis. Another important aspect of this approach is that it is independent of the quality of the crystal structure. Roberts's et al. observed the significance of water molecules in protein–ligand complexes [57]. Docking simulations were performed on a large set of experimentally solved structures contain water molecules in binding sites. They investigate the effect of prior optimization of the orientation of water molecules (both in the presence or absence of the bound ligand) on the accuracy of docking predictions. Their findings revealed that prior optimization of the orientation of water molecules are independent of outcome of the docking process. They also suggest that docking accuracy can be improved via inclusion of crystallographic water molecules in the binding site of target. Using the statistical formulas (for the energy and entropy) and fluid solvation theory, Li and Lazaridis [58] computed the contribution of the displaced water molecules upon binding of ligand in HIV-1 protease. They concluded that the displacement of water molecules is favorable for binding. The major advantage of this method is that it can estimate the bound water molecules which are most favorable to displace. For proper ligand binding orientation it is necessary to anticipate which water molecule should be considered or replaced during the docking simulations. WaterScore [59] and Consolv [60] have explored this concept and can be used to differentiate the water molecules that should be retained or displaced. A novel PPC (polarized protein-specific charge) model incorporated into MM/PBSA (molecular mechanism/Poisson–Boltzmann surface area) was investigated to check the effect of bridge water molecules mediating protein–ligand interactions. This function highlighted the importance of those water molecules which significantly improve the prediction accuracy [61]. Recently, adenosine A2A receptor was systematically studied to investigate the effect of explicit water molecules in virtual screening [62]. Decision trees (DT) approach was employed to select an ensemble of structures with different water molecule positions and orientations. Hydration sites for the docking analyses were extracted from the WaterMap [63] within a 6 Å radius of the ligand. The 299 high affinity A2A receptor antagonists and 17,337 decoys were used to validate the proposed protocol. The major success of this study was that the water molecules derived from a molecular dynamics simulation without any knowledge of crystallographic waters can improve enrichments to same extent as the crystallographic waters, which makes this strategy applies to structures without experimental knowledge of water positions.

4.3. Metals ions

Nearly 30% of the protein contain metal ions in their active site [64], crucial for structure stabilization, signal transduction and respiration. Such metal ions also play an important role for correct orientation of ligand in the binding site of the targets. The binding of ligands to these targets is quite challenging due to multiple coordination geometries and limited availability of accurate force field parameters for the metal–ligand interactions [65]. While considering these metal ions one should optimize the parameters such as radius, well depth and partial charge to enhance the accuracy of docking [66]. The significance of these parameters was investigated on 16 matrix metalloproteinases (MMPs)–ligands complexes. These optimizations showed the improvement in both binding energies and docking accuracy with less than 2 Å rmsd at the zinc binding site on 14 complexes [67]. Similarly, 95,000 small molecules from MDL Drug Data Report (MDDR) were screened against the five metalloenzymes [68]. These catalytic metals ions were defined as interaction center and optimized with van der Waals radii, well depth and charges. Performance of docking results was measured by reproducing experimental geometries and by the enrichment of known inhibitors from the majority of decoy molecules in large database screens. Finally, 15 compounds out of 50 top ranked were selected for experimental evaluation. Experimental analyses confirm the 5 compounds with Ki less than 120 μM in which two inhibitors have 2 μM. This study highlighted the significance of noncovalent scoring function to identify the

inhibitors; however such analyses restricted for some metalloenzymes only.

A model based on statistical analysis of 43,061 metal sites of the Protein Data Bank (PDB), containing magnesium, calcium, zinc, iron, manganese and copper was derived for the automatic calculation and definition of metal interaction geometries for molecular docking analyses [69]. The model was integrated in FlexX and includes the identification of the metal-coordinating ligands, calculation of the coordination geometry and the superposition of ideal polyhedra to identify the optimal position for free coordination sites. It was evaluated on dataset of 103 protein–ligand complexes from protein data bank which resulted in an improved docking performance on majority of the test set. In one approach, Pottel et al. [70] also discuss the issues, such as coordination geometry, atomic charge variability, and a potential proton transfer from small molecules to a neighboring basic residues which are generally ignored. Such important issues have been inadequately modeled by current available docking programs. While working on these aspects they developed the special functions and parameters to account for zinc–drug coordination, focusing on the above-listed phenomena and their impact on docking to zinc metalloenzymes. The newly developed functions were implemented in FITTED docking program and validated against the histone deacetylases (HDACs) target. The improved program was found to be more accurate in identification of active hits. Taking the accountability of atomic charge, Peters et al. [71] optimized the parameters for zinc-containing enzymes in which the zinc atomic charge varies depending on both the ligand (such as water or hydroxide) and coordinated protein residues. They also determined the diversity of bond distances (Zn–O bond is longer in the case of water binding while the Zn–N bonds are shorter due to the strength of the hydroxyl bond) and angles of zinc clusters containing histidine residues and water molecules. Moreover, they also observed the large variations of charge transfer between a neutral (water, charge transfer of 0.17) and negatively charged ligands (hydroxide, 0.41). They recommended the use of polarizable force fields in order to account for such charge transfer. However, due to the computational cost and time required to modify the atomic properties, its implementation in docking program is still a challenge.

Recently, Poongavanam et al. [72] used the quantum mechanical (QM) approach to predict the binding affinity of a set of compounds against the HIV-1 RT associated RNase H (RNH). This enzyme has magnesium ions in the catalytic site to exhibit their therapeutic effect. Three different types of optimizations were performed; (1) MP2 (Møller–Plesset perturbation method) based chelation was used for Mg^{2+} ions and the inhibitor, (2) QM/MM geometry optimization was performed on the whole system, (3) Several key residues along with two Mg^{2+} ions and inhibitor were added to QM and the rest of the system was treated with MM. In all of these analyses, quantum mechanical (QM) showed improved binding affinity prediction for the inhibitors. However, the limitation of this study is that model can only be used for HIV-1–RNH active site directed binders.

5. Ligand preparation

The accruing numbers of public and commercial chemical databases (Table 1) have further opened a door for structure based virtual screening by offering the selection of more diverse compounds against the targets [8–11]. It is well accepted that the content and quality of a compound library have pivotal effects on the success of a virtual screening, increase the false positive rate and often lead to waste of the computer resources. Therefore, before docking one should remove undesirable compounds and select only relevant ones. One common technique, Lipinski's rule of 5 [73] can used for pre-filtration to select compounds that are more likely to be bioavailable in later stages. Alternatively, ligand based virtual screening using known active compounds for similarity searching can be used to accomplish this task. The screening library however needs to be closely related to the objective of the particular virtual screening campaign. For example chemical diverse libraries are particularly attractive for identification of novel scaffolds for selected target. If the goal of the screening is directed at a specific target family, one may use target-oriented synthesis. Moreover, if the goal is lead optimization, chemical libraries with high intermolecular similarity are an attractive source. Table 1 summarizes the list of public and commercial chemical databases, commonly screened in real practice. As compounds have been selected, they are examined for proper geometry and may be minimized to remove the bad steric clashes. Protonation and tautomerization forms are other important factors that have influenced on ligand ranking in virtual screening. The effect of the protonation and tautomeric states on the docking was thoroughly studied [74,75]. We strongly suggest the reading of a comprehensive review by Elizabeth et al. wherein he has discussed the issues of ligand protonation, tautomerism, stereoisomerism and conformation on docking solutions [76].

Kalliokoski et al. [77] studied the effect of protonation and tautomerization on 19 targets and the publicly available DUD (Directory of Useful Decoy) set. They compared all protonated and tautomerized ligands forms with single reasonable ligand form. No major differences have been identified between the enumerated and the predicted set in terms of the enrichment metrics. The outcome of this study indicated that the single predicted state yield comparable improvement over tautomer form in structure-based virtual screening. Comparative study of four docking algorithms (eHiTS, GOLD, FlexX and Surflex) was conducted to investigate the performance of ligand protonation states on BACE1 target [78]. The outcome of this study revealed that FlexX (EF of 69) outperformed over the eHiTS and GOLD. Whereas, the performance of Surflex (EF of 58) was found to be comparable with the FlexX. Considering this and other factors they conclude that ligand protonation is highly depending on the characteristics of the docking algorithms and scoring functions used. FlexX-Pharm and GOLD were found to be affected by ligand protonation. Whereas, Surflex does not depend on ligand protonation.

6. Scoring functions

Scoring functions are widely applied in computing the fitness of receptor–ligand binding affinity [79]. Significant progress on scoring function has been made to develop a broad spectrum of methods for estimation of protein–ligand binding affinities. There are some basic principles that should followed by an ideal scoring function. The applied scoring function should be fast. It can easily differentiate the binders and decoys. Finally, the experimentally determined binding mode should be placed high by a scoring function. Ranking of ligands or compounds based on score during the database screening is a crucial step in structure based virtual screening. Despite of remarkable advancement in molecular docking programs, successful and rapid prediction of receptor–ligand interaction remain a significant challenge [80].Therefore, selection of an appropriate and reliable scoring function is still an open question. Several studies highlighted the importance of different types of scoring functions [81–83]. Generally, three types of scoring functions are currently applied. Complete presentation of these scoring functions is beyond the scope of this article; nevertheless herein, we have introduced some of their recent progress in brief. However, one can also find the in-depth description of these functions in the article by Hung et al. [84].

6.1. Force-field

Force field based scoring functions have been used to compute the binding affinity based on intermolecular interactions i.e. van der Waals and electrostatic interactions between all molecular complexes. These functions are derivative of both experimental and *ab initio* quantum mechanisms. The simplest method use a distance-dependent dielectric constant such as the force field, scoring function in DOCK [85]. The major limitation of this scoring function is their accountability of solvent effect. Most force field scoring functions omits the calculation of internal protein energy term by considering the single conformation of protein [86]. Despite the similar functional form, various force-field scoring functions i.e. *G*-Score [87] and *D*-Score [88] differ in parameter sets. Two energy terms i.e. Lennard-Jones VDM and an electrostatic were used for interaction analyses between protein and ligands. The optimized weights of the electrostatic, van der Waals, and hydrophobic contributions used to calculate the free energy of binding is highly consistent with the experimental affinity data. The parameters of the Lennard-Jones potential varies in different scoring function as *G*-score uses 8–4 Lennard-Jones potential where as *D*-score [88] have 12–6 Lennard-Jones potential. Electrostatic terms are accounted for a Columbic formulation with a distance-dependent dielectric function that lessens the contribution from charge–charge interactions. Yin et al. took the van der Waals, solvation and hydrogen bonding energies into account and developed a novel force field based scoring function; Medusa score, for evaluating protein–ligand binding affinity [89]. The effect of this scoring function has been examined by comparing with the outcomes of others methods. This function was shown to be superior as compared to others functions for both docking decoys and binding affinity.

6.2. Empirical

This scoring function is based on several empirical derived terms (VDW energy, electrostatic energy, hydrogen bonding energy, desolvation, entropy, hydrophobicity) known to be important in binding affinity [90,91]. The binding scores computed from this function is the sum of its unrelated terms. Due to the simple energy terms, empirical function is capable to predict the binding potential much faster as compared to others scoring functions. Experimentally determined binding energies and X-ray crystallography information are used to optimize the coefficients of various terms by using regression analyses. Regression analyses used is highly dependent on the molecular data, which is also known to be major disadvantage of these types of scoring functions. They also yield different weighting factors for the various terms. Empirical scoring function terms can be used different ways for non-bonded interaction. For example, LUDI separates hydrogen bonds into neutral and ionic hydrogen bonds. It also incorporates the entropy penalties by using the weighted sum of the number of rotatable bonds in ligand. On the other hands the ChemScore implements ligand rotational entropy in a more complicated form that describes the molecular environment surrounding each rotatable bond. A simple empirical scoring function was developed by Eldridge et al. [92] to estimate the binding free energy of known protein–ligand complexes. This function was found to reproduce the binding affinity of the complexes with cross-validated error of 8.68 kJ/mol. The main advantage of this scoring function is that it was validated on high-resolution protein–ligand three dimensional structure. In 2002, Wang et al. introduce the new scoring function, X-CScore derived from a larger data set which includes VDM, hydrogen bonds, hydrophobic interaction and rotatable bonds [93]. As compared to other force fields, this consensus scoring function is able to predict the binding energies with a standard deviation of ~2.2 kcal/mol. Its performances in identifying the correct bound conformation are considerably better. The major disadvantage of this scoring function is that it was derived from regression analyses, which characterized from common interactions depend on the training data set. Moreover, interactions such as cation–π interaction and π–π stacking, are not included. Water molecules crucial in protein–ligand interaction also not consider in this scoring function.

Krammer et al. [94] presented two empirical functions referred as Ligscore1 and Ligscore2. These functions consist of three distinct terms that describe the van der Waals interaction, the polar attraction between the receptor and ligand and the desolvation penalty attributed to the binding of the polar ligand atoms in the protein. The advantages of these functions are the attractive polar contact surface between the ligand and protein that results from close and favorable contacts between atoms of opposite polarity. The desolvation penalty term for Ligscore1 was derived from the sum of squared attractive and repulsive protein–ligand contact. However, the desolvation for Ligscore2 was derived from the sum of squared terms that includes the buried polar surface area of the ligand as well as the protein. In both the cases the contribution of desolvation tends to balance the favorable interactions established through polar contacts of the ligand binding to the protein receptor, vital in evaluating protein–ligand binding affinities. Evaluated on 118 protein–ligand complexes, Ligscore2 was performed well with correlation coefficient, R^2 of 0.75 and SD of 1.04. Assessment of different scoring functions on 195 high-quality protein–ligand complexes was performed by Li et al. [95]. Twenty scoring functions, implemented available commercial software were analyzed in terms of "scoring power" (binding affinity prediction),"ranking power" (relative ranking prediction), "docking power" (binding pose prediction), and "screening power". The outcome of this study showed that these scoring functions outperformed in the docking/screening power tests as compared to scoring or ranking power tests. The conception of ID score was comprehended by agglomerating the nine different protein–ligand interaction descriptors referred as van der Waals interaction, hydrogen-bonding interaction, electrostatic interaction, π–system interaction, metal–ligand interaction, desolvation effect, entropic loss effect, shape matching, and surface property matching [96]. This function was calibrated using the 2278 complexes and further tested on a larger independent test set. The performance of this model was found to be higher on different biological targets as compared to other existing scoring functions. The ballpark of ID-score development was tantamount of comprehensive protein–ligand interaction descriptors which have high sensitivity in predicting binding affinity with little structural difference.

A new empirical scoring function was reported for modeling the effect of hydrophobicity, ligand conformational entropy, van der Walls and hydrogen bonds interaction energy [97]. This function was tested on PDBbind dataset and also compared with other scoring functions. Cyscore showed the significant improvement in predicting protein–ligand score as compared to other well established scoring functions. The authors recommended its use in molecular modeling application, as lead optimization and binding pose prediction. However, this function is not optimized for virtual screening which is one of the major drawbacks of it. Moreover, it does not count the charge–charge interaction and water mediated interactions.

6.3. Knowledge based

Knowledge-based scoring function are derived from the statistical analysis of structural information embedded in experimentally determined atomic structures using inverse Boltzmann's method [98,99]. One of the popular knowledge-based scoring functions, potential of mean force (PMF) is known for their potential for

converting structural information into free energies without any knowledge of binding affinities [100]. This function also embraces solvent accessibility corrections to pair wise potential and therefore expected to be more applicable. Different knowledge-based scoring functions with diverse atom typing schemes along with various models have been developed [101–103]. The main drawback of these scoring functions is that they rely on atom random state for the approximation of the reference state without considering the effect of volume and interatomic connectivity [104]. The accuracy of this reference state is challenging task in knowledge-based scoring functions [105]. To combat this, Huang et al. reported an ITScore scoring function to bypass the accurate calculation of reference state [39]. Further, they improved this scoring function by including the solvation and entropy effect [106]. This new scoring function named as ITScore/SE, has witnessed the improvement in the performance of ITScore. Grinter and Zou proposed the novel approach to handle the sparse data problem in knowledge based function [107]. In this method, the inaccuracy estimation is used force-field-based potential (FFP) that does not rely on training data to weight the knowledge based scoring function. The proposed method was found to predict the 91% binding mode correctly on the set 100 complexes by Wang et al. [108], whereas binding affinity correlation was found to be 0.514 with the experimentally determined affinities in PDBbind. By calibrating with a diverse set of 219 protein–ligand complexes, Shen et al. reported a knowledge-based scoring function, referred to as IPMF by integrating additional experimental binding affinity information into the extracted potentials [109]. While comparing with seven others available functions, IPMF perform better with lowest mean error of 1.41 log Ki/Kd units from measured inhibition affinities and the highest Pearson's correlation coefficient of Rp^2 0.40 for the prediction test. Their analyses suggest that the additional binding affinity informational may have potential in developing scoring functions and predicting the binding affinity. In one approach, Neudert et al. introduced the novel knowledge-based scoring function, DSX (Drug Score eXtended), which includes distance-dependent pair potentials, novel torsion angle potentials, and newly defined solvent accessible surface-dependent potentials [110]. A further comparative evaluation of the Drug Score, DSX demonstrated improved results in recognition of near-native docking poses and ligand ranking. Cheng et al. improved the accuracy of binding affinity prediction by knowledge-guided strategy (KGS) [111]. Binding constant from appropriate reference complex is a main source to compute the binding affinity of a given complex. In combination with two *X*-score and PLP scoring function, the strategy was calibrated on three different datasets (HIV protease complexes [112], carbonic anhydrase complexes [44] and trypsin complexes [73]). KGS predict better result for both the crystal structures and computer generated poses as compared to *X*-score and PLP. The most important strength KGS of is that it can concert with any scoring method and does not require re-parameterization of the given scoring function.

7. Pose prediction

The aim of docking simulations is to search the appropriate binders for the selected target from the large chemical dataset. Correct docking is essential in structure based virtual screening as an incorrect orientation in the binding pocket will generally lead to false results, no matter how sophisticated the energy function. However, correct docking by itself does not guarantee a correct prediction of either binding affinity or ligand ranking and the protein–ligand complexes generated by docking, thus need a necessary careful assessment. Visual inspections for this task are not only tedious but also lead to inaccurate predictions. Clustering algorithm can be used to differentiate between the similar or dissimilar docked poses, thus improve pose prediction. The more accurate pose predictions obtained from clustering can increase the accuracy of scoring, as an inaccurate pose will be incorrectly ranked regardless of the scoring function. Several techniques like RMSD (root-mean-square deviation), *K*-means, Jarvis–Patrick and top-first approaches [112] are effective for clustering the dissimilar poses during the docking simulations. However, the major disadvantages of these methods are their dependency on pre-assigned cut-off value to determine the class of the poses which have a high influence on clustering result. To fill this gap, Peng et al. described a new approach by implementing the structural descriptors filtering and KGS-penalty based conformational clustering in docking without any biasness [113]. This function significantly increases the prediction accuracy from 53% to 78% while testing on 150 high-quality protein–ligand complex structures. It can be easily integrated with the selected docking program. Earlier, using the Kohonen self-organizing map (SOM), Bouvier et al. [114] proposed the method, AuPosSOM (Automatic analyses of poses using Self-Organizing MAP) for ligand conformation ranking with the analyses of interatomic contacts between protein–ligand complexes. Alternatively, consensus scoring with independent scoring functions can also be used for filtering the ligand conformations [115,116]. This method combines the set of pose predictions from multiple scoring functions into single predictions. However, it lack common threshold value for all the scoring functions, used for the docking experiments, which is its limitation.

In one approach, multipose binding concept has been implemented in molecular docking protocol [117]. The aim of this study was to analyse the influence of multipose as compared to single-pose on the binding affinity prediction. High-throughput docking study was performed on three datasets (PDBbind Refined has 2455 complexes, PDBbind Core has 216 complexes and CSAR-NRC has 343 complexes) with AutoDock and eHiTS. These two programs differ in search algorithm and scoring functions. Compared to single-pose binding affinity, multipose binding lead to the 30% improvement with high correlation for predicted binding affinity with experimental data and decrease of the residuals between predicted and experimental binding affinities. However, these outcomes were depended on the properties of the complex and the selection of the considered poses. In 2009, Kolb and Irwin [118] reviewed the docking literatures in which ligand pose predicted by docking was experimentally tested. The aim of this study was to analyze the strength of docking protocol to predict the binding mode correctly. Their study has highlighted that docking methods can often, but not always, succeed in predicting the ligand pose correctly. Features like small binding site, binding sites with strong orienting constraints and knowledge about the target have a high influence on predicting correct binding mode.

8. Molecular docking programs

Docking method is one of the most commonly used virtually screening approach in computational drug design. The docking process consists of sampling the ligand orientation within the binding site of the target to form a stable complex. The strength of any docking algorithm lies in between three basic principles; (1) capability of reproducing the X-ray pose, (2) scoring functions to rank the binding poses and (3) discriminating power for binders and decoys in virtual screening experiments. Numbers of docking programs are available, however, only few have been known for their reasonable success. Characteristics of some of these widely used programs are summarized below.

8.1. GOLD (Genetic Optimization of Ligand Docking)

GOLD [47] is automated docking program that mimics the evolutionary algorithm to explore the full range of ligand conformational flexibility with partial flexibility of the protein. Fitness score, based on three energy terms (hydrogen-bond energy, steric interaction energy, ligand internal energy) was used to evaluate the pose of ligand within the complex. The advantage of this method is that it can handle water and ions molecules in the active site of proteins or enzymes easily.

8.2. FlexX

FlexX uses an incremental buildup algorithm where ligands are break at rotatable bonds and start docking with a base fragment [48]. Base fragments are generated by serving all noncyclic bonds in a given ligand. The best solution is evaluated with entropy, hydrogen bonding, lipophilic, ionic and aromatic terms which is modified from Bohm's scoring function. It is known to predict potential locations of water molecules rather rely on crystallographic positions.

8.3. AutoDock

AutoDock [49] is a flexible automated and random search docking algorithm which uses Monti Carlo simulated annealing, evolutionary genetic and Lamarckian genetic algorithm methods. It is known to be the most widely used software among the researcher working in computational drug design area. This program can also be used with a graphical interface called the AutoDock Tools (ADT). The prediction of AutoDock has already shown a great success.

8.4. DOCK

The search strategies include incremental construction and random conformation search and utilize the existing Coulombic and Lennard-Jones grid-based scoring function. An update version of DOCK uses steric matching-scores along with electrostatic and molecular mechanics interaction energies to compute the efficacy of receptor–ligand complex [85].

8.5. GLIDE (Grid-based Ligand Docking with Energetic)

The Glide algorithm with hierarchical filters approximates a systematic search of conformational, orientation, and positional space of the ligand in the binding site of the target. Best docked pose was selected based on energy function that combines empirical and force-field-based terms [119]. The grid generation step requires input files of both ligand and active site, including hydrogen atoms.

8.6. FRED (Fast Rigid Exhaustive Docking)

FRED performs docking with two steps: (1) Shape fitting and (2) optimization. FRED requires the different conformers of single ligands. Ligand conformer libraries can be generated by using OMEGA (Open Eye Scientific Software). During docking simulation each ligand is placed into a specified grid box with all active site atoms using a smooth Gaussian potential. Scoring functions like; Gaussian shape scoring, ChemScore, PLP, ScreenScore, Chemical Gaussian Overlay (CGO) and Chemical Gaussian Tanimoto (CGT) were used in the optimization step to rank the ligands [120].

8.7. HADDOCK (High Ambiguity Driven biomolecular DOCKing)

HADDOCK is an information-driven flexible docking approach for the modeling of biomolecular complexes that makes use experimental information. Such information can be derived from biochemical or biophysical interaction data (chemical shift perturbation data resulting from NMR titration experiments), mutagenesis data or bioinformatics predictions. Moreover, it also allows conformational change of the molecules during complex formation both for the side chains and backbone [121]. In addition to protein–protein docking, HADDOCK has been applied to the modeling of protein–DNA, protein–RNA, protein–oligosaccharides and other protein–ligand complexes.

Despite the quality and quantity of the data, the outcome of docking highly depends on the predictive or discriminatory power of the underlying docking algorithm. Molecular properties such as molecular weight, number of rotatable bonds or polar atoms of ligands have high a influence on docking accuracy. It is believed that docking performance significantly decreases for ligands with large number of rotatable bonds [122]. GOLD being less sensitive in this respect and most recommended program whereas, in terms of enrichment factor, GLIDE turned out to be the most effective programs. Beside these, computational time is another important issue while selecting the docking program. Comparative study of different docking programs by Wandzik [122], showed that for high throughput screening, FlexX and FRED are considered to be quite fast programs as compared with GLIDE and AutoDock. The FRED and GLIDE scoring schemes have been investigated on cyclooxigenase-2, oestrogen receptor, mitogen-activated kinase, gyrase B, gelatinase-A and neuraminidase dataset in virtual screening experiment [123]. This study showed that the GLIDE scoring function performed better over the FRED scoring. However, for lipophilic binding sites, where hydrophobic effects outweighed electrostatic and hydrogen-bonding interactions performance of FRED was better. Kellenberger et al. [124] assessed eight different programs for their ability to find the conformation of near to the experimentally determined solution. The performance of GOLD, GLIDE, or QXP was found better in 61–63% cases when using a low cutoff value of RMSD (i.e.1 Å). However, when RMSD threshold was set to 2 Å, GLIDE and GOLD performance was found to improved up to 80–90%, whereas FlexX and FRED were found to be successful in 66% of the cases. An evaluation of different docking programs (DOCK, GLIDE, FlexX, ICM, Surflex) on the diverse dataset of DUD database was conducted by Cross and co-workers [125]. They analyzed that GLIDE outperformed with the average AUC of 0.72. Whereas, the average AUC was found to be 0.55–0.63 for DOCK, FlexX and ICM.

9. Successful studies

The number of applications of virtual screening has been increased rapidly during the past decade. Several successful cases have been reported which resorted to the synthesis of lead compounds by virtual screening methods. In this section we provide some of these studies for the year 2013 and 2014.

9.1. Case study 1

A new class of SENP2 inhibitors identified by a combination of structure based virtual screening and quantitative FRET based biological assay has been reported earlier by Kumar et al. [126]. In this study two round of virtual screening was performed by using the GLIDE version 5.7 interfaced with the Schrodinger suite. Glide-XP score was used to rank compounds. First round of virtual screening followed by FRET based assay has enabled the identification of

eight compounds with >40% SENP2 inhibition at 30 μM compound concentration. Out of eight, five compounds belong to two scaffolds containing a 1,2,5-oxadiazole core. These scaffolds represent a novel class of SENP2 inhibitors. Furthermore, to improve the inhibitory potency of these novel scaffolds and identification of structure related compounds another round of virtual screening was performed. The most potent compound of each scaffold showed an IC50 of 5.9 and 3.7.

9.2. Case study 2

A structure based virtual screening was carried to identify the novel inhibitors for KPC-2, a class A betalactamase [127]. GOLD (Genetic Optimization for Ligand Docking) 5.0 was used for virtual screening of the compound dataset. Moreover, *X*-score, a consensus scoring function was also used in order to carry out docking validation. The best docking results were used as the starting structure for MD simulation. The 5 ns calculation was performed with GROMACS 4.5.3 package using GROMOS 96 force field. Our analyses showed that high fitness value did not consistence with other scoring schemes. As compounds ranked among top with GOLD fitness score did not perform well in *X*-score and MD simulations. Two compounds on the basis of 5 ns molecular dynamics simulation trajectories were procured and their biological activity was experimentally validated. These compounds perform well with MIC value of 2 and 8 μg/ml as compared with known inhibitors.

9.3. Case study 3

A structure based virtual screening by using the GLIDE was performed by Lei et al. [128] to discover the novel inhibitors of Macrophage migration Inhibitory Factor (MIF). Top scored compounds based on Glide standard precision (SP) were redocked and scored by the Glide extra precision (XP) scoring mode. Top 2000 selected compounds were filtered for eliminating the toxic or other undesirable moieties. The 942 compounds on the basis of Tanimoto coefficient from FCFP_4 fingerprints were clustered into 150 clusters. Finally, 147 compounds from Specs and ChemBridge were purchased and experimentally evaluated. This hybrid approach identified the ten chemically diverse compounds with high potent inhibitory activity against MIF. Three compounds with IC50 values lower than 10 μM and one with an IC50 value below 1 μM (0.55 μM) possessed relatively good inhibitory activity against Macrophage migration Inhibitory Factor (MIF). These inhibitory activities were 26-fold higher as compared to reference compound ISO-1, used in this study.

9.4. Case study 4

Recently, over 3000 US Food and Drug Administration (FDA) approved drug molecules were computationally screened by using the Internal Coordinate Mechanics (ICM) method [ICM-Pro 3.7-2d (Molsoft L.L.C., San Diego, CA)] in order to identify the potential drugs against NEDD8-activating enzyme [129]. Four high-scoring compounds were selected for experimental evaluation. Initially, these compounds were tested in attenuating NAE activity in a cell-free system. NAE activity is evaluated by measuring the quantity of Ubc12–NEDD8 thioester product formed by the enzyme Piperacillin 1 at a concentration of 10 mM showed significant inhibition of NAE-mediated Ubc12–NEDD8 as compared to others tested compounds. Further, dose based experiment was also performed to investigate the ability of piperacillin 1 to inhibit NAE activity. The outcome of this experiment showed that piperacillin 1 dose-dependently inhibited Ubc12–NEDD8 conjugation with an estimated IC50 value of 1 mM. Thus, molecular modeling and kinetic studies revealed the Piperacillin 1 as a non-covalent ATP-competitive inhibitor of NAE.

9.5. Case study 5

Wang, et al. carried out virtual screening approach to discover novel PERK inhibitors [130]. They used two virtual screening approaches, ligand pharmacophore and docking for the identification of initial hits. Three million lead like compounds from ZINC database were selected. This library was used to generated a conformations of compound. Top 10,000 compounds were selected for docking analyses. The top 1000 compounds from this analyses were cluster for similarity analyses. Finally, fifty commercially available compounds were selected for biochemical analyses. Experiment analyses confirm 10 active compounds, two of which show IC50 values that are less than 10 μM in a dose–response assay.

10. Conclusions

Accurate identification of hits is highly important for the pharmaceutical industries since it limits cost and time-consuming experiments to synthesize them with optimized pharmacodynamic and pharmacokinetic properties. Structure based virtual screening becomes routine methods for early-stage drug discovery. Continuous improvement and development of computational based methods for drug designing have gained attraction for virtual screening in recent years. Effective consideration of crucial features during docking process, such as target flexibility, metal ions and water molecules, can give ideal docking solution. In the present article, we highlighted some of the recent progress and advancement in the features associated with the structure based virtual screening. Furthermore, some of the successful applications of these methods were also mentioned. We conclude that using the appropriate screening methods associated with the selected target one can achieve systematic success even within reasonable time and limited computer cost.

Acknowledgements

Author acknowledges DBT, Government of India for the support and internal facilities of the department. This work was supported by internal funds of Biotechnology Unit, AMU and DBT Grant, BT/PR11453/BID/07/296/2009 and BT/HRD/NBA/34/01/2012 to AUK.

References

[1] M. Dickson, J.P. Gagnon, Nat. Rev. Drug Discov. 3 (2004) 417–429.
[2] T. Cheng, Q. Li, Z. Zhou, Y. Wang, S.H. Bryant, AAPS J. 14 (2012) 133–141.
[3] W.P. Walters, M.T. Stahl, M.A. Murcko, Drug Discov. Today 3 (1998) 160–178.
[4] O. Dror, D.S. Duhovny, Y. Inbar, R. Nussinov, H.J. Wolfson, J. Chem. Inf. Model. 49 (2009) 2333–2343.
[5] G.B. McGaughey, R.P. Sheridan, C.I. Bayly, J.C. Culberson, C. Kreatsoulas, C.S. Lindsley, V. Maiorov, J.-F. Truchon, W.D. Cornell, J. Chem. Inf. Model. 47 (2007) 1504–1519.
[6] B.K. Shoichet, R.M. Stroud, D.V. Santi, I.D. Kuntz, K.M. Perry, Science 259 (1993) 1445–1450.
[7] V. Lounnas, T. Ritschel, J. Kelder, R. McGuire, R.P. Bywater, N. Foloppe, J. Comput. Struct. Biotechnol. 5 (2013) e201302011.
[8] J.J. Irwin, B.K. Shoichet, J. Chem. Inf. Model. 45 (2005) 177–182.
[9] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, Y. Djoumbou, R. Eisner, A.C. Guo, D.S. Wishart, Nucleic Acids Res. 39 (2011) D1035–D1041.
[10] M.R. Andersen, M.S. Almén, H.B. Schioth, Nat. Rev. Drug Discov. 10 (2011) 579–590.
[11] Y. Wang, E. Bolton, S. Dracheva, K. Karapetyan, B.A. Shoemaker, T.O. Suzek, J. Wang, J. Xiao, J. Zhang, S.H. Bryant, Nucleic Acids Res. 38 (2010) D255–D266.
[12] K.P. Seiler, G.A. George, M.P. Happ, N.E. Bodycombe, H.A. Carrinski, S. Norton, S. Brudz, J.P. Sullivan, J. Muhlich, M. Serrano, P. Ferraiolo, J.N. Tolliday, S.L. Schreiber, P.A. Clemons, Nucleic Acids Res. 36 (2008) D351–D359.
[13] P. Block, C.A. Sotriffer, I. Dramburg, G. Klebe, Nucleic Acids Res. 34 (2006) D522–D526.
[14] A. Hillisch, L.F. Pinedab, R. Hilgenfeldc, Drug Discov. Today 9 (2004) 659–669.

[15] J.A. Capra, R.A. Laskowski, J.M. Thornton, M. Singh, T.A. Funkhouser, PLoS Comput. Biol. 5 (2009) e1000585.
[16] M. Brylinski, J. Skolnick, Proc. Natl. Acad. Sci. 105 (2008) 129–134.
[17] A. Roy, J. Yang, Y. Zhang, Nucleic Acids Res. 40 (2012) W471–W477.
[18] D. Ghersi, R. Sanchez, Proteins 74 (2009) 417–424.
[19] M. Lape, C. Elam, S. Paula, Biophys. Chem. 150 (2010) 88–97.
[20] I. Halperin, B. Ma, H. Wolfson, R. Nussinov, Proteins 47 (2002) 409–443.
[21] J.F. Truchon, C.I. Bayly, J. Chem. Inf. Model. 47 (2007) 488–508.
[22] W. Zhao, K.E. Hevener, S.W. White, R.E. Lee, J.M. Boyett, BMC Bioinform. 10 (2009) 225.
[23] R.C. Christopher, W.I. Christopher, P. Labute, J. Comput. Aided Mol. Des. 26 (2012) 775–786.
[24] H. Wieman, K. Tøndel, E. Anderssen, F. Drabløs, Mini Rev. Med. Chem. 4 (2004) 793–804.
[25] E. Lindahl, B. Hess, D. van der Spoel, J. Mol. Model. 7 (2001) 306–317.
[26] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A.D. MacKerell, J. Comput. Chem. 31 (2009) 671–690.
[27] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten, J. Comput. Chem. 26 (2005) 1781–1802.
[28] Z. Gao, H. Li, H. Zhang, X. Liu, L. Kang, X. Luo, W. Zhu, K. Chen, X. Wang, H. Jiang, BMC Bioinform. 9 (2008) 104.
[29] X. Chen, Z.L. Ji, Y.Z. Chen, Nucleic Acids Res. 30 (2002) 412–415.
[30] M.P. Magarinos, S.J. Carmona, G.J. Crowther, S.A. Ralph, D.S. Roos, D. Shanmugam, W.C. Voorhis, F. Aguero, Nucleic Acids Res. 40 (2012) D1118–D1127.
[31] P. Cozzini, G.E. Kellogg, F. Spyrakis, D.J. Abraham, G. Costantino, A. Emerson, F. Fanelli, H. Gohlke, L.A. Kuhn, G.M. Morris, M. Orozco, T.A. Pertinhez, M. Rizzi, C.A. Sotriffer, J. Med. Chem. 51 (2008) 6237–6255.
[32] R. Santos, J. Hritz, C. Oostenbrink, J. Chem. Inf. Model. 50 (2010) 146–154.
[33] N.C. Shirai, M. Kikuchi, J. Chem. Phys. 139 (2013) 225103.
[34] J.A. McCammon, B.R. Gelin, M. Karplus, Nature 267 (1977) 585–590.
[35] J. Gu, F. Bai, H. Li, X. Wang, Int. J. Mol. Sci. 13 (2012) 14451–14469.
[36] C.N. Cavasotto, R.A. Abagyan, J. Mol. Biol. 337 (2004) 209–225.
[37] F. Jiang, S.H. Kim, J. Mol. Biol. 219 (1991) 79–102.
[38] J. Nilmeier, L. Hua, E. Coutsias, M. Jacobson, J. Chem. Theory Comput. 7 (2011) 1564–1574.
[39] S.Y. Huang, X. Zou, Int. J. Mol. Sci. 11 (2010) 3016–3034.
[40] M. Okamoto, K. Takayama, T. Shimizu, K. Ishida, O. Takahashi, T. Furuya, J. Med. Chem. 52 (2009) 7323–7327.
[41] M. Rueda, G. Bottegoni, R. Abagyan, J. Chem. Inf. Model. 50 (2010) 186–193.
[42] S. Leis, M. Zacharias, PLoS ONE 7 (2012) 1–13.
[43] R. de Paris, F.A. Frantz, O.N. de Souza, D.D. Ruiz, Biomed Res. Int. (2013) 1–12.
[44] P.N. Hübler, (Ph.D. thesis), PPGCC – PUCRS, Porto Alegre, RS, Brasil, 2010.
[45] R. Thilagavathi, R.L. Mancera, J. Chem. Inf. Model. 50 (2010) 415–421.
[46] S.B.A. de Beer, N.P.E. Vermeulen, C. Oostenbrink, Curr. Top. Med. Chem. 10 (2010) 55–66.
[47] G. Jones, P. Willett, R.C. Glen, J. Mol. Biol. 10 (1995) 43–53.
[48] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, J. Mol. Biol. 10 (1996) 470–489.
[49] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, J. Comput. Chem. 19 (1999) 1639–1662.
[50] D.J. Huggins, B. Tidor, Protein Eng. Des. Sel. 24 (2011) 777–789.
[51] G.A. Ross, G.M. Morris, P.C. Biggin, PLoS ONE 7 (2012) e32036.
[52] O. Trott, A. Olson, J. Comput. Chem. 31 (2010) 455–461.
[53] A. Amadasi, J.A. Surface, F. Spyrakis, P. Cozzini, A. Mozzarelli, G.E. Kellogg, J. Med. Chem. 51 (2008) 1063–1067.
[54] A. Amadasi, F. Spyrakis, P. Cozzini, D.J. Abraham, G.E. Kellogg, A. Mozzarelli, J. Mol. Biol. 358 (2006) 289–309.
[55] D.L. Chen, G.E. Kellogg, J. Comput. Aided Mol. Des. 19 (2005) 69–82.
[56] G. Rossato, B. Ernst, A. Vedani, M. Smieško, J. Chem. Inf. Model. 51 (2011) 1867–1881.
[57] B.C. Roberts, R.L. Mancera, J. Chem. Inf. Model. 48 (2008) 397–408.
[58] Z. Li, T. Lazaridis, J. Am. Chem. Soc. 125 (2003) 6636–6637.
[59] A.T. Garcia-Sosa, R.L. Mancera, P.M. Dean, J. Mol. Model. 9 (9) (2003) 172–182.
[60] M.L. Raymer, P.C. Sanschagrin, W.F. Punch, S. Venkataraman, E.D. Goodman, L.A. Kuhn, J. Mol. Biol. 265 (1997) 445–464.
[61] J. Liu, X. He, J.Z. Zhang, J. Chem. Inf. Model. 53 (2013) 1306–1314.
[62] E.B. Lenselink, T. Beuming, W. Sherman, H.W.T. van Vlijmen, A.P. IJzerman, J. Chem. Inf. Model. 54 (2014) 1737–1746.
[63] R. Abel, T. Young, R. Farid, B.J. Berne, R.A. Friesner, J. Am. Chem. Soc. 130 (2008) 2817–2831.
[64] M.E. Palm-Espling, M.S. Niemiec, P. Wittung-Stafshede, Biochim. Biophys. Acta – Mol. Cell Res. 1823 (2012) 1594–1603.
[65] T. Jain, B. Jayaram, Proteins Struct. Funct. Genet. 67 (2007) 1167–1178.
[66] J. Wang, P. Cieplak, J. Li, Q. Cai, M.J. Hsieh, R. Luo, Y. Duan, J. Phys. Chem. B 116 (2012) 7088–7101.
[67] X. Hu, W.H. Shelver, J. Mol. Graph. Model. 22 (2003) 115–126.
[68] J.J. Irwin, F.M. Raushel, B.K. Shoichet, Biochemistry 44 (2005) 12316–12328.
[69] B. Seebeck, I. Reulecke, A. Kämper, M. Rarey, Proteins 71 (2008) 1237–1254.
[70] J. Pottel, E. Therrien, J.L. Gleason, N. Moitessier, J. Chem. Inf. Model. 54 (2014) 254–265.
[71] M.B. Peters, Y. Yang, B. Wang, L.F. Molnár, M.N. Weaver, K.M.J. Merz, Chem. Theory Comput. 6 (2010) 2935–2947.
[72] V. Poongavanam, C. Steinmann, J. Kongsted, PLoS ONE 9 (2014) e98659.
[73] C.A. Lipinski, J. Pharmacol. Toxicol. Methods 44 (2000) 235–249.
[74] M. Petukh, S. Stefl, E. Alexov, Curr. Pharm. Des. 19 (2013) 4182–4190.
[75] Y.C. Martin, J. Comput. Aided Mol. Des. 23 (2009) 693–704.
[76] Y. Elizabeth, A. Mark, A.R. Paul, J. Mol. Recognit. 24 (2011) 149–164.
[77] T. Kalliokoski, H.S. Salo, M.L. Kakkonen, A. Poso, J. Chem. Inf. Model. 49 (2009) 2742–2748.
[78] T. Polgar, C. Magyar, I. Simon, G.M. Keserue, J. Chem. Inf. Model. 47 (2007) 2366–2373.
[79] D.B. Kitchen, H. Decornez, J.R. Furr, J. Bajorat, Nat. Rev. Drug Discov. 3 (2004) 935–949.
[80] G.L. Warren, C.W. Andrews, A. Capelli, B. Clarke, J. LaLonde, M.H. Lambert, M. Lindvall, N. Nevins, S.F. Semus, S. Senger, G. Tedesco, I.D. Wall, J.M. Woolven, C.E. Peishoff, M.S. Head, J. Med. Chem. 49 (2006) 5912–5931.
[81] N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, C.R. Corbeil, Br. J. Pharmacol. 153 (2008) S7–S26.
[82] O. Korb, T. Stutzle, T.E. Exner, J. Chem. Inf. Model. 49 (2009) 84–96.
[83] I.-L. Lu, H.D. Wang, J. Comput. Biol. (2012) 1215–1226.
[84] S.Y. Huang, S.Z. Grinter, X. Zou, Phys. Chem. Chem. Phys. 12 (2010) 12899–12908.
[85] T.J.A. Ewing, I.D. Kuntz, J. Comput. Chem. 10 (1997) 1175–1189.
[86] D.B. Kitchen, H. Decornez, J.R. Furr, J. Bajorath, Nat. Rev. Drug Discov. 3 (2004) 935–949.
[87] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Talor, J. Mol. Biol. 267 (1997) 727–748.
[88] B. Kramer, M. Rarey, T. Lengauer, Proteins 37 (1999) 228–241.
[89] S. Yin, L. Biedermannova, J. Vondrasek, N.V. Dokholyan, J. Chem. Inf. Model. 48 (2008) 1656–1662.
[90] C.A. Sotriffer, P. Sanschagrin, H. Matter, G. Klebe, Struct. Funct. Bioinform. 73 (2008) 395–419.
[91] H.J. Bohm, J. Comput. Aided Mol. Des. 8 (1994) 243–256.
[92] M.D. Eldridge, C.W. Murray, T.R. Auton, G.V. Paolini, R.P. Mee, J. Comput. Aided Mol. Des. 11 (1997) 425–445.
[93] R. Wang, L. Lai, S. Wang, J. Comput. Aided Mol. Des. 16 (2002) 11–26.
[94] A. Krammer, P.D. Kirchhoff, X. Jiang, C.M. Venkatachalam, M. Waldman, J. Mol. Graph. Model. 23 (2005) 395–407.
[95] Y. Li, Li Han, Z. Liu, R. Wang, J. Chem. Inf. Model. 54 (2014) 1717–1736.
[96] G.B. Li, L.L. Yang, W.J. Wang, L.L. Li, S.Y. Yang, J. Chem. Inf. Model. 53 (2013) 592–600.
[97] Y. Cao, L. Li, Bioinformatics 30 (2014) 1674–1680.
[98] M.J. Sippl, J. Mol. Biol. 213 (1990) 859–883.
[99] P.D. Thomas, K.A. Dill, J. Mol. Biol. 257 (1996) 457–469.
[100] I. Muegge, J. Med. Chem. 49 (2006) 5895–5902.
[101] C.Y. Yang, R. Wang, S. Wang, J. Med. Chem. 49 (2006) 5903–5911.
[102] I. Muegge, Y.C. Martin, J. Med. Chem. 42 (1999) 791–804.
[103] A.V. Ishchenko, E.I. Shakhnovich, J. Med. Chem. 45 (2002) 2770–2780.
[104] C. Zhang, S. Liu, Q. Zhu, Y. Zhou, J. Med. Chem. 48 (2005) 2325–2335.
[105] I. Muegge, Perspect. Drug Discov. Des. 20 (2000) 99–114.
[106] S.Y. Huang, X. Zou, J. Chem. Inf. Model. 50 (2010) 262–273.
[107] S.Z. Grinter, X. Zou, J. Comput. Chem. 35 (2014) 932–943.
[108] R. Wang, Y. Lu, S. Wang, J. Med. Chem. 46 (2003) 2287–2303.
[109] Q. Shen, B. Xiong, M. Zheng, X. Luo, C. Luo, X. Liu, Y. Du, J. Li, W. Zhu, J. Shen, H. Jiang, J. Chem. Inf. Model. 51 (2011) 386–397.
[110] G. Neudert, G. Klebe, J. Chem. Inf. Model. 51 (2011) 2731–2745.
[111] T. Cheng, Z. Liu, R. Wang, BMC Bioinform. 11 (2010) 193.
[112] O. Demir, R. Baronio, F. Salehi, C.D. Wassman, L. Hall, G.W. Hatfield, R. Chamberlin, P. Kaiser, R.H. Lathrop, R.E. Amaro, PLoS Comput. Biol. 7 (2011), http://dx.doi.org/10.1371/journal.pcbi.1002238.
[113] S.M. Peng, Y. Zhou, N. Huang, Chin. Chem. Lett. 24 (2013) 1001–1004.
[114] G. Bouvier, N.E. Todeschi, J.P. Girault, G. Bertho, Bioinformatics 26 (2010) 53–60.
[115] P.S. Charifson, J.J. Corkery, M.A. Murcko, W.P. Walters, J. Med. Chem. 42 (1999) 5100–5109.
[116] R. Wang, S. Wang, J. Chem. Inf. Comput. Sci. 41 (2001) 1422–1426.
[117] K. Atkovska, S.A. Samsonov, M.P. Rogacz, M.T. Pisabarro, Int. J. Mol. Sci. 15 (2014) 2622–2645.
[118] P. Kolb, J.J. Irwin, Curr. Top. Med. Chem. 9 (2009) 755–770.
[119] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, J. Med. Chem. 47 (2004) 1739–1749.
[120] M. McGann, J. Chem. Inf. Model. 51 (2011) 578–596.
[121] S.J. Vries, M. Dijk, A. Bonvin, Nat. Protoc. 5 (2010) 883–897.
[122] I. Wandzik, MATCH Commun. Math. Comput. Chem. 55 (2006) 271–278.
[123] T.S. Gasch, M. Stahl, J. Mol. Model. 9 (2003) 47–57.
[124] E. Kellenberger, J. Rodrigo, P. Muller, D. Rognan, Proteins: Struct. Funct. Bioinf. 57 (2004) 225–242.
[125] J.B. Cross, D.C. Thompson, B.K. Rai, J.C. Baber, K.Y. Fan, Y.B. Hu, C. Humblet, J. Chem. Inf. Model. 49 (2009) 1455–1474.
[126] A. Kumar, A. Ito, M. Takemoto, M. Yoshid, K.Y.J. Zhang, J. Chem. Inf. Model. 54 (2014) 870–880.
[127] M. Danishuddin, A. Khan, M. Faheem, P. Kalaiarasan, M.H. Baig, N. Subbarao, A.U. Khan, J. Biomol. Struct. Dyn. 32 (2014) 741–750.
[128] L. Xu, Y. Zhang, L. Zheng, C. Qiao, Y. Li, D. Li, X. Zhen, T. Hou, J. Med. Chem. 57 (2014) 3737–3745.
[129] H.J. Zhong, L.J. Liu, D.S.H. Chan, H.M. Wang, P.W. Chan, D.L. Mab, C.H. Leung, Biochimie 102 (2014) 211–215.
[130] Q. Wang, P. Jihyun, A.K. Devkota, E.J. Cho, K.N. Dalby, P. Ren, J. Chem. Inf. Model. 54 (2014) 1467–1475.