

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.

For more information visit [www.intechopen.com](http://www.intechopen.com)



# Molecular Docking in Modern Drug Discovery: Principles and Recent Applications

*Aaftaab Sethi, Khusbhoo Joshi, K. Sasikala  
and Mallika Alvala*

## Abstract

The process of hunt of a lead molecule is a long and a tedious process and one is often demoralized by the endless possibilities one has to search through. Fortunately, computational tools have come to the rescue and have undoubtedly played a pivotal role in rationalizing the path to drug discovery. Of all techniques, molecular docking has played a crucial role in computer aided drug design and has swiftly gained ranks to secure a valuable position in the modern scenario of structure-based drug design. In this chapter, the principle, sampling algorithms, scoring functions and diverse available software's for molecular docking have been summarized. We demonstrate the interplay of docking, classical techniques of structure-based design and X-ray crystallography in the process of drug discovery. In addition, we dwell upon some of the limitations faced in docking studies. Finally, several success stories of molecular docking approaches in drug discovery have been highlighted, concluding with remarks on molecular docking for the future.

**Keywords:** molecular docking, virtual screening, drug discovery, computer aided drug design, conformational sampling, scoring functions

## 1. Introduction

The path to drug discovery is a long, challenging & arduous task not to mention the overburdening finances it demands. As of 2014, the average cost of developing a new drug from scratch was found to be an estimated \$2.5 billion, an increase of 145% from the previous study done by the same organization in 2003. The major reasons for this drastic increase in the cost is mainly attributed to high failure rate of drugs among others [1]. Understanding of the drug discovery process is important to handle the challenges faced by the pharma companies in terms of cost and innovation.

The process of identifying a target, synthesizing an active compound with suitable characteristics like minimal toxicity, high bioavailability, cost-effective synthesis, etc., and finally developing it to introduce in the market is a time-consuming, extremely complex and risky endeavor [2]. Initially, a target is identified which plays a key role in progress of the disease. Once a link between the target and the disease has been established, the next step is to identify potential candidates which can stop or reverse the progress of the disease [3]. This process starts with the

discovery of molecules that show efficacy in a simple screen, called “hits.” Screening is a process in which normally a large number of compounds from natural products and online databases are examined for biological activity in high-throughput assays. This step in the drug discovery process is very crucial and demands maintaining huge molecular libraries and carrying out thousands or millions of assays, which leaves the academicians and small pharmaceutical companies at a disadvantage and also shoots up the cost for larger industries. Next, the “hits” found are chemically modified to give improved pharmaceutical properties, such compounds are often called “leads.” But, it is quite apparent that the method stated above for discovery of a drug has a number of pitfalls. From an academic point of view, carrying out high throughput screens (HTS) is costly, time-consuming and not feasible; while, from an industrial perspective, it does nothing to improve the eminent danger of market saturation.

Truly innovative and blockbuster drugs are what drive the pharmaceutical industry forward but, over the past few years introduction of new molecular entities (NMEs) has vastly reduced. For example, in 2007 only 19 NMEs were approved by the US Food and Drug Administration (FDA), the least since 1983 [4]. Currently, and even in the future it is expected that only slight modifications of the existing blockbuster drugs would be carried out which would further aggravate this problem [5]. HTS would not help in either curbing the rising costs of discovering hits or the problem of finding truly innovative and blockbuster NMEs, the two major hurdles that the pharmaceutical industry faces now-a-days.

To overcome these challenges, molecular docking is an exemplary tool. During the first step to find hits from existing chemicals for a drug discovery and development project, virtual screening (VS) is a perfectly viable and an alternative or complementary approach to HTS for fulfilling the screening of thousands or millions of compounds within a few days. In addition, the speed of VS helps in kick-starting projects for newer targets for which no leads are available [6]. Molecular docking is one of the most applied virtual screening methods and has become increasingly useful overtime on account of immense growth in 3D X-ray and NMR structures and their improved resolution (physics and knowledge based docking algorithms depend on it) reported in the Protein Data Bank (PDB). As an example, in total 46,541 X-ray structures were reported at the end of 2008 in PDB, but by the end of 2018 it had grown to a staggering figure of 131,993 [7]. In addition, it is a resource saving technique which provides accessibility of screening to academia and small industries which were earlier limited to large pharmaceutical giants.

In this chapter, we will discuss a particular class of molecular design, i.e., “Docking” along with the various algorithms, techniques, success stories and limitations related to it. In the end, we will conclude with its scope in the near future.

## **2. Molecular docking**

Two molecules can interact in a number of ways let alone the interaction of a protein and protein or a protein and small molecule. Molecular docking helps us in predicting the intermolecular framework formed between a protein and a small molecule or a protein and protein and suggest the binding modes responsible for inhibition of the protein. To accurately carry out docking studies one requires the high-resolution X-ray, NMR or homology-modeled structure with known/predicted binding site in the biomolecule. To date, 148,827 are available in the database (PDB) [3]. Docking methods fit a ligand into a binding site by combining and optimizing variables like steric, hydrophobic and electrostatic complementarity and also estimating the free energy of binding (scoring) [8].

There are two basic components which distinguish the variety of docking softwares available to choose from—One is sampling algorithm and the other is scoring function, these are discussed in detail.

## 2.1 Sampling algorithm

As pointed out earlier, there are a huge number of modes of binding between two molecules and even with advances in parallel computing and higher clock speeds of modern computers it would be expensive and time-consuming to generate all the possible modes. Therefore, algorithms were needed which could fish out valuable conformations from the fruitless ones.

Various algorithms were developed in this regard and can be classified by the number of degrees of freedom they ignore. The simplest of the algorithms introduced treated the molecules as two rigid bodies thereby reducing the degree of freedom to just six (three translational and three rotational). A very well cited example of a program using this algorithm is DOCK [9]. This program was designed to find molecules which had a huge extent of shape similarity to the pockets/grooves or binding sites. It derives an image of suspected binding sites present on the surface of the protein. This image consists of several overlapping spheres of varying radii which touch the molecular surface of the macromolecule at just two points. The ligand molecule is also considered as a set of spheres which approximately fill the space occupied by the ligand. Once the respective representations of the protein surface and the ligand in terms of sphere are complete, the pairing rule is applied. The pairing rule is based on the principle that ligand sphere can be paired with a protein sphere if the internal distances of all the spheres in the ligand set match all the internal distances within the protein set, allowing some user specified tolerance. Thus, it allows the program to identify geometrically similar cluster of spheres on the protein site and the ligand. Many other programs were developed later which make use of such matching algorithm (MA) which include LibDock [8], LIDAEUS [10], PhDOCK [11], Ph4DOCK [12], Q-fit [13], SANDOCK [14], etc. All these programs based on MA have the advantage of speed but have several limitations such as prior need for detailed receptor geometry and lack of molecular flexibility which does not accurately define many aspects of ligand-protein interactions.

The second algorithm is that of incremental construction (IC), wherein the ligand is fragmented from rotatable bonds into various segments. One of the segments is anchored to the receptor surface. The anchor is generally considered to be the fragment which shows maximum interactions with the receptor surface, has minimum number of alternate conformations and fairly rigid such as the ring system. Once the base/anchor has been established, the next step is to add each of the fragments step by step. Ideally, those fragments are added first which have a greater chance of showing interactions like hydrogen bonding since they are directional in nature and are responsible for specificity of the ligand. In addition, hydrogen bonds lead to more accurate prediction of geometry. Once a particular fragment is added, the poses with the least energies are considered for the next iteration, making the algorithm extremely fast and robust [15]. IC has been used in programs like DOCK 4.0 [16], FlexX [15], Hammerhead [17], SLIDE [18] and eHiTS [19], SKELGEN [20], ProPose [21], PatchDock [22], MacDock [23], FLOG [24], etc. One major limitation of this program is that it is restricted to medium sized ligands and is not feasible for large size ligands where the number of fragments generated pose a problem.

Another useful algorithm is the Monte Carlo (MC) technique. In this approach, a ligand is modified gradually using bond rotation and translation or rotation of the



entire ligand. More than one parameter can also be changed at a time to get a particular conformation. That conformation is then evaluated at the binding site based on energy calculation using molecular mechanics and is then rejected or accepted for the next iteration based on Boltzmann's probability constant. Acceptance or rejection of the conformation is a function of the change in energy with respect to a parameter  $T$ , which can be physically interpreted as temperature (simulated annealing). This criterion of acceptance or rejection makes this method strikingly different than the others. Whereas the other algorithm favor decrease in energy, in MC method increases are also possible. For higher values of  $T$  increases are likely. If one starts at a high value of  $T$ , then small energy barriers can be jumped and the configuration can move beyond local minima and is therefore particularly useful in situations where a global minimum is sought among many local minima [25]. An interesting spin-off of the MC approach is the Tabu search, which maintains a record of the search space of the binding site which has already been visited and thus ensures that the binding site is explored to the maximum [26]. MC approach has been made use of in programs like DockVision 1.0.3 [25], FDS [27], GlamDock [28], ICM [29], MCDOCK [30], PRODOCK [31], QXP [32], ROSETTALIGAND [33], RiboDock [34], Yucca [35], AutoDock [36], etc. One of the major concerns with MC approach is the uncertainty of convergence, which can be improved by performing multiple independent runs.

Genetic algorithm (GA) is quite similar to MC method and is basically used to find the global minima [37]. These are much inspired by the Darwin's Theory of Evolution [38]. GA maintains a population of ligands with an associated fitness determined by the scoring function. Each ligand represents a potential hit. The GA alters the ligands of the population by mutation or crossover. In the first stage, a new population is created by accessing and then selecting the more fit ligands from the previous step. The members of the populations are then transformed in the alteration step. The mutation operator creates new ligands from a single ligand by randomly changing a fragment in its representation while the crossover operator exchanges information between two (occasionally more) members of the population [39–41]. GA has been incorporated in programs like Autodock 4.0 [42], DARWIN [43], DIVALI [39], FITTED [44], FLIPDOCK [45], GAMBLER [46], GAsDock [47], GOLD 3.1 [48], PSI-DOCK [49]. GA has a similar limitation like that of MC method wherein the uncertainty of convergence is a major drawback.

Another approach is the hierarchical method. In this approach, the low energy conformations of the ligand are pre-computed and aligned. The populations of the pre-generated ligand conformations are merged into a hierarchy such that similar conformations are positioned adjacent to each other within the hierarchy. Afterwards, on carrying out rotation or translation of the ligand, the docking program will make use of this hierarchical data structure and thus minimize the outcomes. Let us understand with a simple example—if an atom near the rigid center of the ligand is found to clash with the protein in a given rotation/translation, then this approach can reject all of the conformations lying below in the hierarchy to that of the conformation under scrutiny, because the descendants must contain the same clash as well [50]. GLIDE software makes use of the hierarchical method [51, 52].

## **2.2 Scoring functions**

Sampling changes among varying degrees of freedom must be performed with sufficient accuracy to identify a conformation that best matches the receptor structure, and also must be fast enough to permit the evaluation of millions of compounds in a set computational time. This is taken care by the variety of algorithms discussed above. Algorithms are further complemented by scoring functions.

The evaluation and ranking of predicted ligand conformations is a crucial aspect of VS. When we are interested in only how a single ligand binds to a biomolecule, then the scoring function needs to predict the docked orientation which most accurately represents the “true” structure of the intermolecular complex. On the other hand, if we are interested to evaluate multiple ligands, in that scenario the scoring function should not only identify the “true” docking pose but also be able to rank one ligand relative to another. Therefore, the design of reliable scoring functions and schemes which can rank different poses is of fundamental importance [53].

The scoring functions usually estimate binding energy of complex using many assumptions and simplifications to arrive as close as possible to actual binding energy in minimum time. Popular scoring functions have an adequate balance between accurate estimation of binding energy and computational cost in terms of time. There have been a number of scoring functions developed over the past many years and can be classified into three main categories—force field, empirical and knowledge based [54].

Force field functions: force field (FF) scoring functions are developed based on physical atomic interactions like van der Waals interactions, electrostatic interactions and bond lengths, bond angles and torsions [55]. Force field functions and parameters are usually derived from both experimental data and ab initio quantum mechanical calculations according to the principles of physics.

$$E = \sum_i \sum_j \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \right) \quad (1)$$

Here,  $r_{ij}$  stands for the distance between protein atom  $i$  and ligand atom  $j$ ,  $A_{ij}$  and  $B_{ij}$  are the van der Waal parameters,  $q_i$  and  $q_j$  are the atomic charges and  $\epsilon(r_{ij})$  is the distance dependent dielectric constant.

One common example of a FF scoring function is that of the program DOCK [56] represented in Eq. (1), where, the effect of solvent is indirectly considered by the distance dependent dielectric constant  $\epsilon(r_{ij})$  seen in the Coulombic potential. One major drawback of this function is that it does not consider an important solvent effect that charged groups favor aqueous environments whereas non-polar groups tend to stay in non-aqueous environments, commonly referred to as the desolvation effect [57]. Ignorance could lead to biased results as the function would now be totally dependent on the coulombic interactions and would thus favor highly charged ligands. In other words, it only takes into account the interaction of protein and ligand, which is inadequate. To build a more robust function one needs to also evaluate how both interact with water before the formation of the complex and how water mediates this process.

Later the Shoichet group [58] improved upon the existing function by adding the effects of the solvent on protein-ligand interactions using implicit solvent models. They employed the Poisson-Boltzmann approach to model the electrostatic potential of the protein. The van der Waals interactions were calculated using the Lennard-Jones potential; the electrostatic interaction between the ligand and the protein was estimated using a precomputed receptor potential determined with DelPhi [59]. Ligand desolvation penalties were calculated with HYDREN [60]. The solvent-corrected scores were found to be closer to experimental binding free energies than the DOCK program scores, but were still too favorable. The overestimation of complex stability could be due to the neglect of solute entropic terms [58].

There a few scoring functions which be classified in this category such as DockScore [56], GoldScore [61], HADDOCK Score [62], ICM SF [29], QXP SF [32], etc.

Empirical scoring functions: the basis of this scoring function is that the binding energies of a complex can be approximated by a sum of individual uncorrelated terms. The coefficients of the various terms involved in calculation of binding energy are obtained from regression analysis using experimentally determined binding energies or potentially from X-ray structural information. Empirical functions have simpler energy terms to evaluate when compared to force field scoring functions and thus are much faster in binding score calculations.

The first empirical scoring function developed to predict binding free energies was implemented in LUDI, credited to the pioneering work of Bohm [63]. The energy was derived using experimental binding free energies and protein-ligand crystal structures for 45 complexes.

$$\Delta G_{bind} = \Delta G_O + \Delta G_{hb} \sum_{h-bonds} f(\Delta R, \Delta \alpha) + \Delta G_{ionic} \sum_{ionic\ int.} f(\Delta R, \Delta \alpha) + \Delta G_{lipo} |A_{lipo}| + \Delta G_{rot} NROT. \quad (2)$$

Here,  $\Delta G_O$  is the binding energy independent of protein interactions,  $\Delta G_{hb}$  describes contribution to binding energy from hydrogen bonds,  $\Delta G_{ionic}$  denotes contribution to binding energy from unperturbed ionic interactions,  $\Delta G_{lipo}$  considers contribution to binding energy through lipophilic interactions while  $A_{lipo}$  is the lipophilic contact surface between the protein and the ligand,  $\Delta G_{rot}$  describes the loss of binding energy due to freezing of internal degrees of freedom in the ligand while NROT represents number of rotatable bonds and  $f(\Delta R, \Delta \alpha)$  is a penalty function that accounts for large deviations from ideal hydrogen bond and salt bridge geometry.

As shown in Eq. (2), the binding free energy is modeled using hydrogen bonds, salt bridges, the hydrophobic effect, and solute entropy terms. The hydrogen bond and salt bridge terms are modified by a penalty function which accounts for deviation from ideal geometry. Entropy loss of the ligand upon complex formation is based on the Number of ROTatable bonds (NROT) in the ligand [64, 65]. Eldridge et al. presented an empirical scoring function referred to as ChemScore by taking into account different energetic parameters like hydrogen bonds, the lipophilic effects of atoms, the effective number of rotatable bonds in the ligand among others. The scoring function was calibrated using 82 ligand-receptor complexes with known binding affinities [66].

By including different empirical energy terms, many different empirical scoring functions have been developed such as SCORE2 [67], ChemScore [66], RankScore [68], LigScore [69], GlideScore [51], HINT [70], etc. The empirical scoring functions take into account many different energy terms and thus the problem of unknowingly double-counting of certain energy terms difficult issue to tackle.

Knowledge based scoring functions: these are derived from the structural information embedded in experimentally determined atomic structures. The functions use statistical analysis on crystal structures of complexes to obtain the interatomic contact frequencies between the protein and the ligand based on the presumption that stronger an interaction is, the greater the frequency of its occurrence will be. The overall score is calculated with the help of Eq. (3) by accounting for favorable contacts and repulsive interactions between each atom in the ligand and protein lying within a sphere with a specified cutoff [71–78].

$$w(r) = -k_B T \ln [g(r)], g(r) = (r)\rho(r)/\rho^*(r) \quad (3)$$

Here,  $k_B$  is the Boltzmann constant,  $T$  is the absolute temperature of the system,  $\rho(r)$  is the number density of the protein-ligand atom at distance  $r$ ,  $\rho^*(r)$  is the pair



density in the reference state where interatomic interactions are zero and  $g(r)$  is pair distribution function.

Popular knowledge based functions include DrugScore [79], PMF [72, 80], MScore [81], SMOG [71], BLEEP [74], ITScore/SE [75], etc. The computational simplicity of such functions is a major advantage especially when one has large databases at hand however, the accuracy of predicting the reference state and underrepresentation of interactions with halogens and metals are the major hurdles.

Each of the above classified have their inherent drawbacks, in this regard, combination of more than one scoring functions has given improved results. This approach has been widely regarded as “Consensus Scoring” [46].

Another set of scoring functions which have recently started to attract attention are based on machine learning. One of the programs based on functions incorporating machine learning was able to achieve an astounding hit rate of 88.6% [82]. The nexus of machine learning and scoring functions is promising but the development of such a tool is slow owing to its complexity.

In order to compare the variety of scoring functions that have been developed up until now, comparative assessment of scoring functions (CASF) is an incredible platform to begin with [83].

There is another set of classification proposed for the scoring functions namely physics-based methods, empirical scoring functions, knowledge-based potentials, and descriptor-based scoring functions but there is still no clear consensus on which classification of scoring functions would be appropriate [84].

### 3. Applications

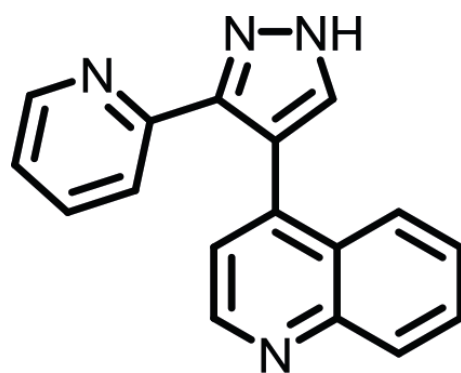
Molecular docking has been developed and improving for many years, but its ability to generate a viable drug is still generally questioned. In the section below, you will find examples where docking approach lead to recognition of active hits for a variety of different receptors/targets.

HIV 1 Integrase—a new binding site for drugs treating AIDS was discovered by Schames et al. using docking while considering the flexibility of the receptor through molecular dynamics. The group used AutoDock in conjunction with the relaxed-complex method to discover novel mode of inhibition of HIV integrase [85].

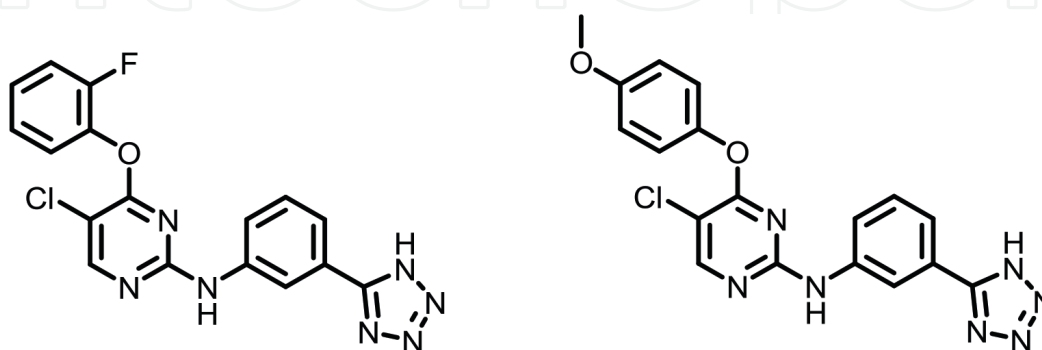
$\alpha$ 1A Adrenergic receptor—Evers et al. generated a model of the receptor using homology modeling based on the X-ray crystallographic structure of bovine rhodopsin. Hierarchical virtual screening method was performed by them on the Aventis in-house compound repository in a stepwise manner. 22,950 filtered compounds were then docked into the  $\alpha$ 1A receptor homology model with the program GOLD and scored with PMF. The top scoring compounds were finally clustered according to their unity fingerprint similarity, and a diverse set of 80 compounds was tested in a radio ligand displacement assay. Thirty-seven compounds displayed a  $K_i < 10 \mu\text{M}$  with the most active having  $K_i = 1.4 \text{ nM}$  [86].

Type I TGF-beta receptor kinase—A striking example and a proof of the benefit of in silico approach over classical high-throughput screening involves the discovery of novel Type I TGF-beta receptor kinase inhibitor. The same molecule (HTS-466284); **Figure 1**, a 27 nM inhibitor, was discovered independently using virtual screening [87] and also by traditional enzyme and cell-based high-throughput screening in the same year [88]. The compound discovered experimentally required in vitro screening of a large library of compounds in a TGF- $\beta$ -dependent cell-based assay which required more time, proved to be costlier and required usage of a variety of chemicals when compared to its computational counterpart.





**Figure 1.**  
Structure of HTS-46628, type I TGF-beta receptor kinase inhibitor.

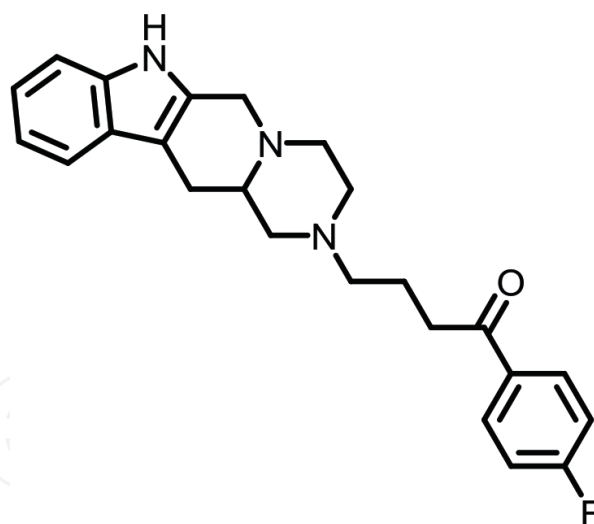


**Figure 2.**  
Structures for Aurora Kinase A inhibitor with  $IC_{50}$  12 and 43 pM respectively.

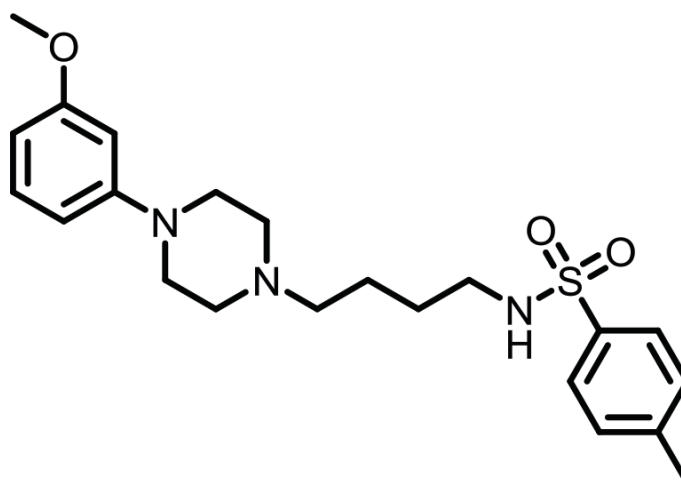
Aurora Kinase A—A major improvement was seen in the inhibitory activity of Aurora Kinase A inhibitors which were designed using *in silico* techniques by Park et al. [89]. This research group made use of a genetic algorithm to carry out the sampling while the scoring function involved the energy terms from the AutoDock program with a slight modification of the dehydration energy term. The design strategy and tools used to carry out the study proved to be immensely successful with some inhibitors revealing exceptionally high potency at low picomolar levels; **Figure 2** [89].

Dopamine D3 receptor—The 3D structure of the Dopamine 3 (D3) subtype receptor was modeled by Varady et al. from the X-ray crystallographic structure of rhodopsin and validated using experimental data. A D3 pharmacophore model was devised by them from 10 selective and potent known D3 receptor ligands. Using their model, 250,251 compound were screened from the National Cancer Institute (NCI) 3D database. The hit list of 2478 potential ligands was then filtered for known chemotypes. After removal of all compounds that were structurally similar to known D3 receptor ligands, 1314 candidates remained. At the end, 20 compounds supplied by NCI to the group were tested, out of which eight had  $K_i$  values below 500 nM, among which one of the compounds had  $K_i = 11$  nM; **Figure 3** [90].

Serotonin receptor (5HT1A)—Due to lack of structural information available for the receptor, Becker et al. made use of PREDICT, to develop a unique non-homology model for building a virtual 3D structure of the receptor. Using the model, 40,000 compounds from Predix's compound library were screened for molecular docking and 78 virtual hits were discovered and then purchased by them from respective vendors. The *in vitro* 5-HT1A binding assays elucidated that 16 of the 78 compounds tested by the group were found to be hits with  $K_i < 5$   $\mu$ M, reflecting a 21% hit rate, 9 of which had a  $K_i < 1$   $\mu$ M. The most potent molecule had  $K_i = 1$  nM (**Figure 4**) and was selected as a lead molecule for further optimization. One significant feature of the study which highlights the utility of docking was that



**Figure 3.**  
Structure of dopamine D<sub>3</sub> receptor inhibitor with  $K_i = 11$  nM.



**Figure 4.**  
Structure of serotonin receptor inhibitor with  $K_i = 1$  nM.

the complete discovery process, i.e., from in silico screening through lead optimization, preclinical, and into clinical studies, was very rapid, requiring less than a couple of years from program initiation to Phase I clinical trial [91].

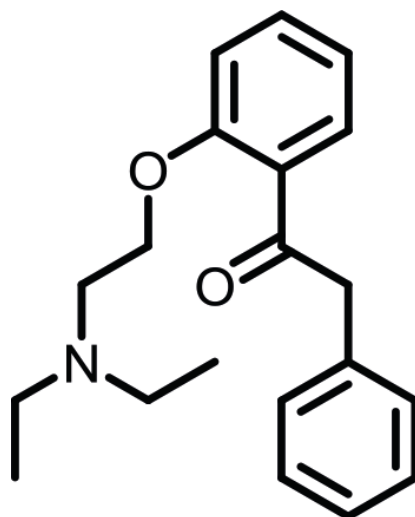
Crystal structure prediction challenge—The International Blind Test is a challenge organized by the Cambridge Crystallographic Data Center wherein a previously determined crystal structure is only revealed once all the participants submit their respective structures. In the Fifth International Blind Test, the challenge was toughened by including flexible molecules with 50–60 atoms. The successful prediction by two participants of the crystal structure of molecule XX in the blind test indicated that search methods and models for lattice energy are capable of providing worthwhile results, both in terms of the range of structures considered in the search and relative energies of the structures and thus can act as efficient ranking systems [92].

Muscarinic M<sub>3</sub> receptor—A pharmacophore model was constructed by Marriot et al. from the known molecules showing significant M<sub>3</sub> potency [93]. The research group utilized the program DISCO, which generated five models. Three models were rejected based on structural overlay. 3D screening was performed by Unity 3D of the Astra compound database. The first model developed by them gave 176 hits while the second model gave 173 hits; 172 compounds were common to the two sets and were tested for their M<sub>3</sub>-antagonistic potency. Several compounds with

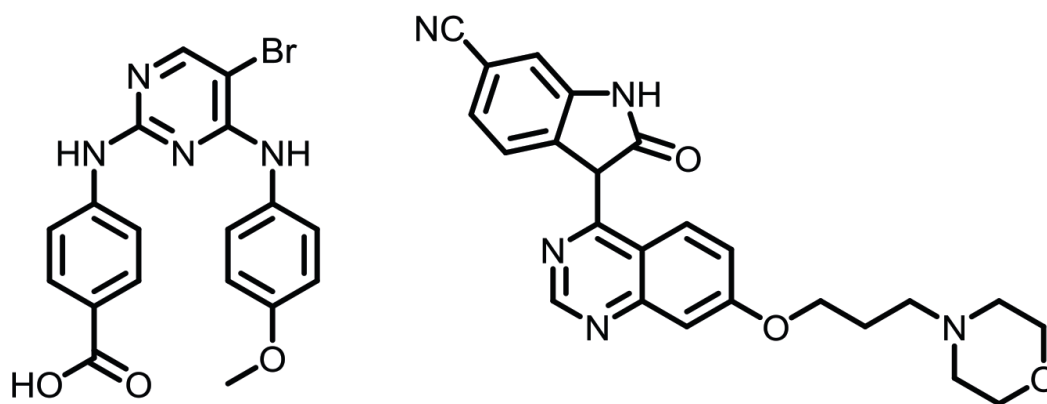
micromolar and even submicromolar activities resulted, for example, compound below had A50 M3 antagonism  $\approx 0.2 \mu\text{M}$ ;  $pA_2 = 6.67$ ; **Figure 5** [93].

Checkpoint Kinase 1—Lyne et al. utilized virtual screening to discover Checkpoint Kinase 1 (Chk-1) inhibitors [94]. Compounds with molecular weight  $> 600$  or with more than 10 rotatable bonds were excluded from the database. Then 3D structures of the ligands were generated using Corina and a maximum of 8 stereoisomers were generated for each molecule. A 3D pharmacophore search was performed with their in-house program Plurality to eliminate compounds that do not have the typical binding motif for the kinase region. The remainder of the compounds were docked into the ATP binding site of Chk-1, using the program FlexX-Pharm, which considers full flexibility of the ligand but treats the protein as a rigid structure. The research group then utilized consensus scoring to identify molecules which were consistently giving good score with different scoring functions. Finally, visual inspection by the group of the 250 highest scoring hits for unfavorable interactions with the binding site or compounds with unrealistic conformations resulted in a list of 103 compounds for biological testing. Thirty-six hits were identified with  $IC_{50}$  ranging from 110 nM to 68  $\mu\text{M}$ ; **Figure 6** [94].

Human Cathepsin K—Schröder et al. presented the implementation of a docking-based virtual screening workflow for the retrieval of covalent binders, human cathepsin K was utilized as a test case [95]. By using the filter of electrophilic war heads, a database with two million structurally diverse compounds with a



**Figure 5.**  
Structure of muscarinic M3 receptor antagonist.



**Figure 6.**  
Structures of checkpoint kinase 1 inhibitor with  $IC_{50}$  450 nM and 4  $\mu\text{M}$  respectively.

variety of functional groups was reduced to a data set of just 343 test compounds. Molecular docking was performed by them and the top scoring poses of the GoldScore ranking list were taken into account for the manual selection of the virtual hits based on visual inspection of the appropriate fit of the molecule in the active site. A data set of 44 compounds including the five low scoring compounds were finally selected for experimental evaluation. The activity of 21 out of the selected 39 *in silico* hits was experimentally confirmed and four out of the five structures predicted as inactive showed no activity on cathepsin K. This study demonstrated to a huge extent the ability of docking to generate positive outcomes (Figure 7) [95].

Human aldose reductase (ALR2)—ALR2 catalyzes a key reaction in the polyol pathway of glucose metabolism, a process implicated in the long-term complications of diabetes. Its inhibitors were designed by Wang et al. using molecular dynamic (MD) simulations and virtual screening [96]. A major challenge encountered by them in the *in silico* studies was that the binding site of the enzyme underwent large conformational changes and adopted distinct configurations upon binding different classes of ligands. To address this issue, the group sampled potentially accessible binding site conformations by MD simulations based on the available crystallographic structures of ALR2. After this procedure, three average conformations were selected for the docking. FlexX was utilized to carry out docking of 7200 compounds of which 128 compounds were selected by them for further screening. Out of these 72 molecules were selected which had RMSD < 3.00 Å for experimental assay, of which 15 novel ALR2 inhibitors hits were discovered. The most potent inhibitor had an  $IC_{50} = 0.24 \mu\text{M}$ ; Figure 8 [96].

Cyclooxygenase-2 (COX-2) and  $\beta$ -amyloid aggregation inhibitors—Dadashpour et al. made use of AutoDock4.2 to carry out docking studies of designed molecules

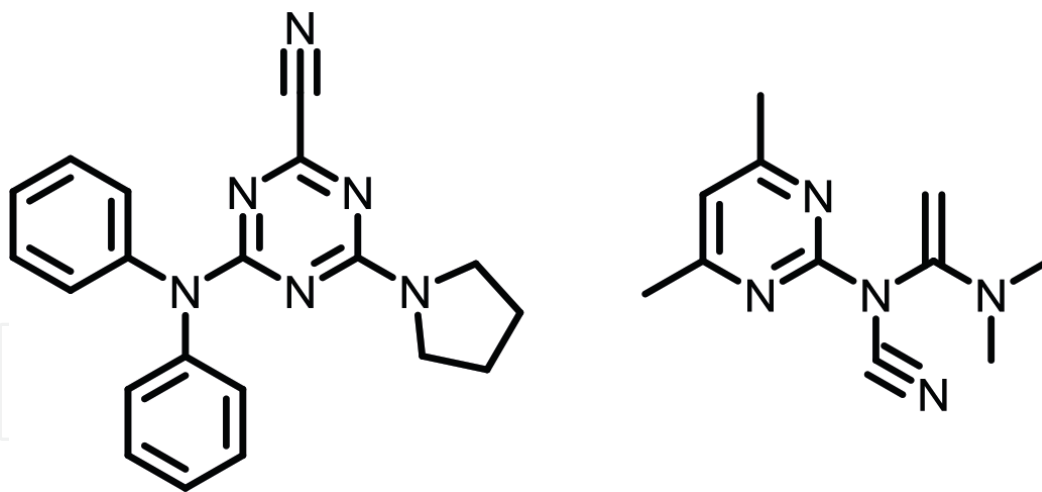


Figure 7.  
Respective structures for active and inactive covalent binders of human cathepsin K.

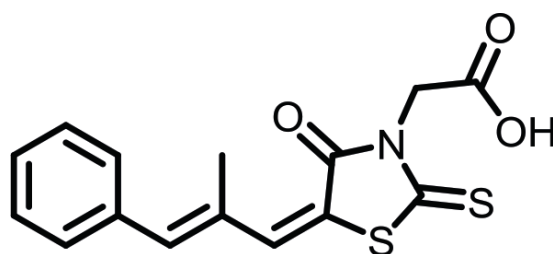
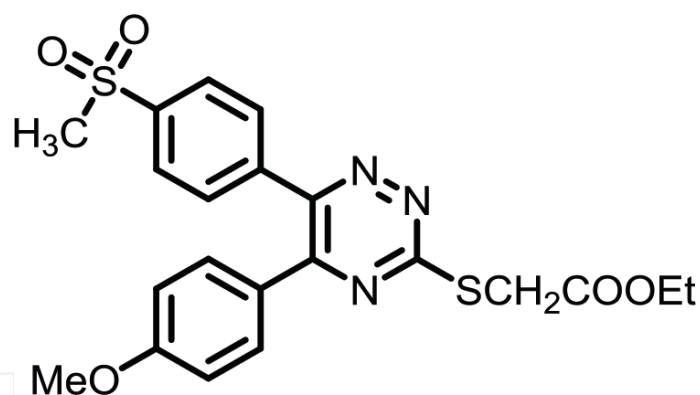


Figure 8.  
Structure of human aldose reductase inhibitor with  $IC_{50} = 0.24 \mu\text{M}$ .





**Figure 9.**  
Structure of cyclooxygenase-2 inhibitor with  $IC_{50} = 10.1 \mu M$ .

based on diaryltriazine as lead. To validate the enzyme-inhibitor complex, the key molecular interactions and calculated binding energy were considered by them. Among the designed molecules, one of the compounds (**Figure 9**) showed an  $IC_{50}$  of  $10.1 \mu M$  in experimental COX-2 assay. In addition, it showed potent anti-aggregation activity on  $\beta$  peptides [97].

#### 4. Limitations

The major limitation of molecular docking is due to the lack of confidence on the ability of scoring functions to give accurate binding energies. This stems from the fact that some intermolecular interaction terms are hardly predicted accurately, such as solvation effect and entropy change [98]. In addition, some intermolecular interactions are rarely considered in scoring functions which have been proven to be of significance. For instance, halogen bonding is verified to make a contribution to protein-ligand binding affinity [99] and so do guanidine-arginine interactions [100], but are not considered.

Transthyretin-thyroxine complex—One critical example wherein energy functions failed is that of transthyretin-thyroxine complex. The docking simulations with energy functions resulted in generation of two binding modes, one similar to the native binding mode of thyroxine and the other belonging to an alternate binding domain with a root mean square deviation (RMSD) of  $8.97 \text{ \AA}$  from native binding state. The energy simulation was carried out and the lower energy solution picked by the docking program was the one with higher RMSD. Thus, in this case molecular docking failed to make the correct prediction of binding mode. Thereby, it would be fair to conclude that we might get many false negatives during the process of VS. [101].

It is still an unsolved problem to accurately deal with the water molecules in binding pocket during docking process, which is tough task and needs a lot of attention in the near future due to two reasons. Firstly, the x-ray crystal structures lack the coordinate information of hydrogen, due to inefficient scattering by smaller atoms. Not knowing the exact position of hydrogen leads to inaccuracies in identifying water molecules which might be acting as a bridging molecule between the ligand and the receptor. Secondly, no reliable theoretical approach is available to accurately predict how water molecules are affected by ligands and how strong the effect is. On top of that, it impossible with our current knowledge to predict how many water molecules in the binding pocket would be replaced by potential ligands and how the hydrogen bonding network would be disturbed by ligand binding [102].

One of the major challenges faced in the field of docking is that of rigid receptor. A protein can adopt many different conformations depending upon the ligand to which it binds. As a result, docking performed using a rigid receptor will correspond to a single receptor conformation, which leads to false negatives in many cases where later the ligand was found to be active. This happens because a protein can exist in constant motion between different conformational states having similar energies, which is usually neglected in docking [58].

Finally, the spectrum of activity against off-target proteins is something rarely seen even in computational screens and is only dealt by animal and human trials.

## 5. Conclusion

Thus, it is quite evident from the case studies highlighted above and many more success stories that one can find in literature related to computer aided drug design, that *in silico* approaches in combination with biophysical data, experimental high throughput screening and biology/toxicology/clinical studies are an indispensable tool in the process of drug discovery. It assists in decision making, conceptualizing new ideas and exploring them in a rapid manner to test a hypothesis, bringing solutions to problems that cannot be assessed experimentally either because the experiments is too difficult to design or because it would cost too much.

Undoubtedly, many challenges still remain to be addressed such as role of water molecules, solvent effects, entropic effects, and receptor flexibility.


There is more than sufficient information now that proves the utility of computational tools in drug design and there is no scope for any debate regarding the effectiveness and advantage of computational tools in the process of drug discovery.

### Author details

Aaftaab Sethi, Khusbhoo Joshi, K. Sasikala and Mallika Alvala\*  
National Institute of Pharmaceutical Education and Research, Hyderabad,  
Telangana, India

\*Address all correspondence to: [mallikaalvala@yahoo.in](mailto:mallikaalvala@yahoo.in)

### IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] Tufts Center for the Study of Drug Development. Cost to develop and win marketing approval for a new drug is \$2.6 billion; 2014
- [2] Muntha P. Drug discovery & development. *Journal of Pharmacy and Pharmaceutical Sciences*. March, 2016;5(1):135-142
- [3] Available from: <https://www.fda.gov/forpatients/approvals/drugs/ucm405382.htm>
- [4] Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: The pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery*. 2010;9(3):203
- [5] Karamelic J, Ridic O, Ridic G, Jukic T, Coric J, Subasic D, et al. Financial aspects and the future of the pharmaceutical industry in the United States of America. *Materia Socio Medica*. 2013;25(4):286. DOI: 10.5455/msm.2013.25.286-290
- [6] Alvarez JC. High-throughput docking as a source of novel drug leads. *Current Opinion in Chemical Biology*. 2004;8(4):365-370. DOI: 10.1016/j.cbpa.2004.05.001
- [7] Available from: <https://www.rcsb.org/stats/growth/xray>
- [8] Diller DJ, Merz KM Jr. High throughput docking for library design and library prioritization. *Proteins: Structure, Function, and Bioinformatics*. 2001;43(2):113-124. DOI: 10.1002/1097-0134(20010501)43:2<113::AID-PROT1023>3.0.CO;2-T
- [9] Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*. 1982;161(2):269-288. DOI: 10.1016/0022-2836(82)90153-X
- [10] Wu SY, McNae I, Kontopidis G, McClue SJ, McInnes C, Stewart KJ, et al. Discovery of a novel family of CDK inhibitors with the program LIDAEUS: Structural basis for ligand-induced disordering of the activation loop. *Structure*. 2003;11(4):399-410. DOI: 10.1016/S0969-2126(03)00060-1
- [11] Joseph-McCarthy D, Thomas BE IV, Belmarsh M, Moustakas D, Alvarez JC. Pharmacophore-based molecular docking to account for ligand flexibility. *Proteins: Structure, Function, and Bioinformatics*. 2003;51(2):172-188. DOI: 10.1002/prot.10266
- [12] Goto J, Kataoka R, Hirayama N. Ph4Dock: Pharmacophore-based protein–ligand docking. *Journal of Medicinal Chemistry*. 2004;47(27):6804-6811. DOI: 10.1021/jm0493818
- [13] Jackson RM. Q-fit: A probabilistic method for docking molecular fragments by sampling low energy conformational space. *Journal of Computer-Aided Molecular Design*. 2002;16(1):43-57. DOI: 10.1023/A:1016307520660
- [14] Burkhard P, Taylor P, Walkinshaw M. An example of a protein ligand found by database mining: Description of the docking method and its verification by a 2.3 Å X-ray structure of a thrombin-ligand complex1. *Journal of Molecular Biology*. 1998;277(2):449-466. DOI: 10.1006/jmbi.1997.1608
- [15] Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*. 1996;261(3):470-489. DOI: 10.1006/jmbi.1996.0477
- [16] Ewing TJ, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: Search strategies for automated molecular docking of

- flexible molecule databases. *Journal of Computer-Aided Molecular Design*. 2001;**15**(5):411-428. DOI: 10.1023/A:1011115820450
- [17] Welch W, Ruppert J, Jain AN. Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites. *Chemistry & Biology*. 1996;**3**(6):449-462. DOI: 10.1016/S1074-5521(96)90093-9
- [18] Schnecke V, Kuhn LA. Virtual screening with solvation and ligand-induced complementarity. In: *Virtual Screening: An Alternative or Complement to High Throughput Screening?* Netherlands: Springer; 2000. pp. 171-190. DOI: 10.1007/0-306-46883-2\_10
- [19] Zsoldos Z, Reid D, Simon A, Sadjad BS, Peter Johnson A. eHiTS: An innovative approach to the docking and scoring function problems. *Current Protein and Peptide Science*. 2006;**7**(5): 421-435. DOI: 10.2174/138920306778559412
- [20] Alberts IL, Todorov NP, Dean PM. Receptor flexibility in de novo ligand design and docking. *Journal of Medicinal Chemistry*. 2005;**48**(21): 6585-6596. DOI: 10.1021/jm050196j
- [21] Seifert MH. ProPose: Steered virtual screening by simultaneous protein-ligand docking and ligand-ligand alignment. *Journal of Chemical Information and Modeling*. 2005;**45**(2): 449-460. DOI: 10.1021/ci0496393
- [22] Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research*. 2005;**33**(suppl\_2):W363-W367. DOI: 10.1093/nar/gki481
- [23] Fradera X, Kaur J, Mestres J. Unsupervised guided docking of covalently bound ligands. *Journal of Computer-Aided Molecular Design*. 2004;**18**(10):635-650. DOI: 10.1007/s10822-004-5291-4
- [24] Miller MD, Kearsley SK, Underwood DJ, Sheridan RP. FLOG: A system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. *Journal of Computer-Aided Molecular Design*. 1994;**8**(2):153-174. DOI: 10.1007/BF00119865
- [25] Hart TN, Read RJ. A multiple-start Monte Carlo docking method. *Proteins: Structure, Function, and Bioinformatics*. 1992;**13**(3):206-222. DOI: 10.1002/prot.340130304
- [26] Baxter CA, Murray CW, Clark DE, Westhead DR, Eldridge MD. Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins: Structure, Function, and Bioinformatics*. 1998;**33**(3):367-382. DOI: 10.1002/(SICI)1097-0134(19981115)33:3<367::AID-PROT6>3.0.CO;2-W
- [27] Taylor RD, Jewsbury PJ, Essex JW. FDS: Flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. *Journal of Computational Chemistry*. 2003;**24**(13):1637-1656. DOI: 10.1002/jcc.10295
- [28] Tietze S, Apostolakis J. GlamDock: Development and validation of a new docking tool on several thousand protein-ligand complexes. *Journal of Chemical Information and Modeling*. 2007;**47**(4):1657-1672. DOI: 10.1021/ci7001236
- [29] Totrov M, Abagyan R. Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins: Structure, Function, and Bioinformatics*. 1997;**29**(S1):215-220. DOI: 10.1002/(SICI)1097-0134
- [30] Liu M, Wang S. MCDOCK: A Monte Carlo simulation approach to the



- molecular docking problem. *Journal of Computer-Aided Molecular Design*. 1999;**13**(5):435-451. DOI: 10.1023/A:1008005918983
- [31] Trosset JY, Scheraga HA. PRODOCK: Software package for protein modeling and docking. *Journal of Computational Chemistry*. 1999;**20**(4):412-427. DOI: 10.1002/(SICI)1096-987X(199903)20:4<412::AID-JCC3>3.0.CO;2-N
- [32] McMartin C, Bohacek RS. QXP: Powerful, rapid computer algorithms for structure-based drug design. *Journal of Computer-Aided Molecular Design*. 1997;**11**(4):333-344. DOI: 10.1023/A:1007907728892
- [33] Meiler J, Baker D. ROSETTALIGAND: Protein-small molecule docking with full side-chain flexibility. *Proteins: Structure, Function, and Bioinformatics*. 2006;**65**(3):538-548. DOI: 10.1002/prot.21086
- [34] Morley SD, Afshar M. Validation of an empirical RNA-ligand scoring function for fast flexible docking using RiboDock<sup>®</sup>. *Journal of Computer-Aided Molecular Design*. 2004;**18**(3):189-208. DOI: 10.1023/B:JCAM.0000035199.48747.1e
- [35] Choi V. Yucca: An efficient algorithm for small-molecule docking. *Chemistry & Biodiversity*. 2005;**2**(11):1517-1524. DOI: 10.1002/cbdv.200590123
- [36] Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. *Proteins: Structure, Function, and Bioinformatics*. 1990;**8**(3):195-202. DOI: 10.1002/prot.340080302
- [37] Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*. 1998;**19**(14):1639-1662. DOI: 10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B
- [38] Yang JM, Kao CY. Flexible ligand docking using a robust evolutionary algorithm. *Journal of Computational Chemistry*. 2000;**21**(11):988-998. DOI: 10.1002/1096-987X(200008)21:11<988::AID-JCC8>3.0.CO;2-H
- [39] Clark KP. Flexible ligand docking without parameter adjustment across four ligand-receptor complexes. *Journal of Computational Chemistry*. 1995;**16**(10):1210-1226. DOI: 10.1002/jcc.540161004
- [40] Oshiro CM, Kuntz ID, Dixon JS. Flexible ligand docking using a genetic algorithm. *Journal of Computer-Aided Molecular Design*. 1995;**9**(2):113-130. DOI: 10.1007/BF00124402
- [41] Jones G, Willett P, Glen RC. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of Molecular Biology*. 1995;**245**(1):43-53. DOI: 10.1016/S0022-2836(95)80037-9
- [42] Österberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins: Structure, Function, and Bioinformatics*. 2002;**46**(1):34-40. DOI: 10.1002/prot.10028
- [43] Taylor JS, Burnett RM. DARWIN: A program for docking flexible molecules. *Proteins: Structure, Function, and Bioinformatics*. 2000;**41**(2):173-191. DOI: 10.1002/1097-0134(20001101)41:2<173::AID-PROT30>3.0.CO;2-3
- [44] Corbeil CR, Englebienne P, Moitessier N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *Journal of Chemical Information*

and Modeling. 2007;**47**(2):435-449.  
DOI: 10.1021/ci6002637

[45] Zhao Y, Sanner MF. FLIPDock: Docking flexible ligands into flexible receptors. *Proteins: Structure, Function, and Bioinformatics*. 2007;**68**(3):726-737. DOI: 10.1002/prot.21423

[46] Charifson PS, Corkery JJ, Murcko MA, Walters WP. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *Journal of Medicinal Chemistry*. 1999; **42**(25):5100-5109. DOI: 10.1021/jm990352k

[47] Li H, Li C, Gui C, Luo X, Chen K, Shen J, et al. GAsDock: A new approach for rapid flexible docking based on an improved multi-population genetic algorithm. *Bioorganic & Medicinal Chemistry Letters*. 2004;**14**(18): 4671-4676. DOI: 10.1016/j.bmcl.2004.06.091

[48] Verdonk ML, Chessari G, Cole JC, Hartshorn MJ, Murray CW, Nissink JWM, et al. Modeling water molecules in protein-ligand docking using GOLD. *Journal of Medicinal Chemistry*. 2005; **48**(20):6504-6515. DOI: 10.1021/jm050543p

[49] Pei J, Wang Q, Liu Z, Li Q, Yang K, Lai L. PSI-DOCK: Towards highly efficient and accurate flexible ligand docking. *Proteins: Structure, Function, and Bioinformatics*. 2006;**62**(4): 934-946. DOI: 10.1002/prot.20790

[50] Grinter SZ, Zou X. Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design. *Molecules*. 2014; **19**(7):10150-10176. DOI: 10.3390/molecules190710150

[51] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method

and assessment of docking accuracy. *Journal of Medicinal Chemistry*. 2004; **47**(7):1739-1749. DOI: 10.1021/jm0306430

[52] Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, et al. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of Medicinal Chemistry*. 2004;**47**(7):1750-1759. DOI: 10.1021/jm030644s

[53] Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews Drug Discovery*. 2004;**3**(11): 935. DOI: 10.1038/nrd1549

[54] Liao C, Sitzmann M, Pugliese A, Nicklaus MC. Software and resources for computational medicinal chemistry. *Future Medicinal Chemistry*. 2011;**3**(8): 1057-1085. DOI: 10.4155/fmc.11.63

[55] Huang N, Kalyanaraman C, Irwin JJ, Jacobson MP. Physics-based scoring of protein-ligand complexes: Enrichment of known inhibitors in large-scale virtual screening. *Journal of Chemical Information and Modeling*. 2006;**46**(1): 243-253. DOI: 10.1021/ci0502855

[56] Meng EC, Shoichet BK, Kuntz ID. Automated docking with grid-based energy evaluation. *Journal of Computational Chemistry*. 1992;**13**(4): 505-524. DOI: 10.1002/jcc.540130412

[57] Available from: <https://www.cambridge.com/medchemconsulting.com/resource/s/solvation.html>

[58] Shoichet BK, Leach AR, Kuntz ID. Ligand solvation in molecular docking. *Proteins: Structure, Function, and Bioinformatics*. 1999;**34**(1):4-16. DOI: 10.1002/%28SICI%291097-0134%2819990101%2934%3A1%3C4%3A%3AAID-PROT2%3E3.0.CO%3B2-6

- [59] Nicholls A, Honig B. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *Journal of Computational Chemistry*. 1991;**12**(4): 435-445. DOI: 10.1002/jcc.540120405
- [60] Rashin AA. Hydration phenomena, classical electrostatics, and the boundary element method. *Journal of Physical Chemistry*. 1990;**94**(5): 1725-1733. DOI: 10.1021/j100368a005
- [61] Jones-Hertzog DK, Jorgensen WL. Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method. *Journal of Medicinal Chemistry*. 1997;**40**(10):1539-1549. DOI: 10.1021/jm960684e
- [62] van Dijk M, van Dijk AD, Hsu V, Boelens R, Bonvin AM. Information-driven protein-DNA docking using HADDOCK: It is a matter of flexibility. *Nucleic Acids Research*. 2006;**34**(11): 3317-3325. DOI: 10.1093/nar/gkl412
- [63] Böhm H-J. LUDI: Rule-based automatic design of new substituents for enzyme inhibitor leads. *Journal of Computer-Aided Molecular Design*. 1992;**6**(6):593-606. DOI: 10.1007/BF00126217
- [64] Böhm H-J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *Journal of Computer-Aided Molecular Design*. 1994;**8**(3):243-256
- [65] Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. *Annual Review of Biophysics and Biomolecular Structure*. 2003;**32**(1): 335-373. DOI: 10.1146/annurev.biophys.32.110601.142532
- [66] Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of Computer-Aided Molecular Design*. 1997;**11**(5):425-445
- [67] Böhm H-J. Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *Journal of Computer-Aided Molecular Design*. 1998;**12**(4):309-309
- [68] Moitessier N, Therrien E, Hanessian S. A method for induced-fit docking, scoring, and ranking of flexible ligands. Application to peptidic and pseudopeptidic  $\beta$ -secretase (BACE 1) inhibitors. *Journal of Medicinal Chemistry*. 2006;**49**(20):5885-5894. DOI: 10.1021/jm050138y
- [69] Krammer A, Kirchhoff PD, Jiang X, Venkatachalam C, Waldman M. LigScore: A novel scoring function for predicting binding affinities. *Journal of Molecular Graphics and Modelling*. 2005;**23**(5):395-407. DOI: 10.1016/j.jm gm.2004.11.007
- [70] Cozzini P, Fornabaio M, Marabotti A, Abraham DJ, Kellogg GE, Mozzarelli A. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 1. Models without explicit constrained water. *Journal of Medicinal Chemistry*. 2002;**45**(12):2469-2483. DOI: 10.1021/jm0200299
- [71] Ishchenko AV, Shakhnovich EI. Small molecule growth 2001 (SMoG2001): An improved knowledge-based scoring function for protein-ligand interactions. *Journal of Medicinal Chemistry*. 2002;**45**(13):2770-2780. DOI: 10.1021/jm0105833
- [72] Muegge I, Martin YC. A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *Journal of Medicinal*



- Chemistry. 1999;**42**(5):791-804. DOI: 10.1021/jm980536j
- [73] Feher M, Deretey E, Roy S. BHB: A simple knowledge-based scoring function to improve the efficiency of database screening. *Journal of Chemical Information and Computer Sciences*. 2003;**43**(4):1316-1327. DOI: 10.1021/ci030006i
- [74] Mitchell JB, Laskowski RA, Alex A, Thornton JM. BLEEP-potential of mean force describing protein-ligand interactions: I. Generating potential. *Journal of Computational Chemistry*. 1999;**20**(11):1165-1176. DOI: 10.1002/(SICI)1096-987X(199908)20:11<1165::AID-JCC7>3.0.CO;2-A
- [75] Huang S-Y, Zou X. Inclusion of solvation and entropy in the knowledge-based scoring function for protein-ligand interactions. *Journal of Chemical Information and Modeling*. 2010;**50**(2):262-273. DOI: 10.1021/ci9002987
- [76] Kumar A, Goyal R, Jain S. *Docking Methodologies and Recent Advances, Oncology: Breakthroughs in Research and Practice*. Hershey, PA: IGI Global; 2017. pp. 804-828. DOI: 10.4018/978-1-5225-0549-5.ch031
- [77] Brás N, Cerqueira N, Sousa S, Fernandes P, Ramos M. Protein ligand docking docking in drug discovery drug discovery. In: *Protein Modelling*. London: Springer; 2014. pp. 249-286. DOI: 10.1007/978-3-319-09976-7\_11
- [78] Náray-Szabó G. *Protein Modelling*. London: Springer; 2014. DOI: 10.1007/978-3-319-09976-7
- [79] Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. *Journal of Molecular Biology*. 2000; **295**(2):337-356. DOI: 10.1006/jmbi.1999.3371
- [80] Muegge I. PMF scoring revisited. *Journal of Medicinal Chemistry*. 2006; **49**(20):5895-5902. DOI: 10.1021/jm050038s
- [81] Yang C-Y, Wang R, Wang S. M-score: A knowledge-based potential scoring function accounting for protein atom mobility. *Journal of Medicinal Chemistry*. 2006;**49**(20):5903-5911. DOI: 10.1021/jm050043w
- [82] Wójcikowski M, Ballester PJ, Siedlecki P. Performance of machine-learning scoring functions in structure-based virtual screening. *Scientific Reports*. 2017;**7**:46710. DOI: 10.1038/srep46710
- [83] Su M, Yang Q, Du Y, Feng G, Liu Z, Li Y, et al. Comparative assessment of scoring functions: The CASF-2016 update. *Journal of Chemical Information and Modeling*. 25 Feb 2019; **59**(2):895-913. DOI: 10.1021/acs.jcim.8b00545. Epub 2018 Dec 11
- [84] Liu J, Wang R. Classification of current scoring functions. *Journal of Chemical Information and Modeling*. 2015;**55**(3):475-482. DOI: 10.1021/ci500731a
- [85] Schames JR, Henchman RH, Siegel JS, Sotriffer CA, Ni H, McCammon JA. Discovery of a novel binding trench in HIV integrase. *Journal of Medicinal Chemistry*. 2004;**47**(8):1879-1881. DOI: 10.1021/jm0341913
- [86] Evers A, Klabunde T. Structure-based drug discovery using GPCR homology modeling: Successful virtual screening for antagonists of the alpha1A adrenergic receptor. *Journal of Medicinal Chemistry*. 2005;**48**(4):1088-1097. DOI: 10.1021/jm0491804
- [87] Singh J, Chuaqui CE, Boriack-Sjodin PA, Lee W-C, Pontz T, Corbley MJ, et al. Successful shape-based virtual screening: The discovery of a potent inhibitor of the type I TGFβ receptor kinase (TβRI). *Bioorganic & Medicinal Chemistry Letters*. 2003;**13**(24):



4355-4359. DOI: 10.1016/j.bmcl.2003.09.028

[88] Sawyer JS, Anderson BD, Beight DW, Campbell RM, Jones ML, Herron DK, et al. Synthesis and activity of new aryl- and heteroaryl-substituted pyrazole inhibitors of the transforming growth factor- $\beta$  type I receptor kinase domain. *Journal of Medicinal Chemistry*. 2003; **46**(19):3953-3956. DOI: 10.1021/jm0205705

[89] Park H, Jung H-Y, Mah S, Hong S. Systematic computational design and identification of low Picomolar inhibitors of Aurora kinase a. *Journal of Chemical Information and Modeling*. 2018; **58**(3):700-709. DOI: 10.1021/acs.jcim.7b00671

[90] Varady J, Wu X, Fang X, Min J, Hu Z, Levant B, et al. Molecular modeling of the three-dimensional structure of dopamine 3 (D3) subtype receptor: Discovery of novel and potent D3 ligands through a hybrid pharmacophore- and structure-based database searching approach. *Journal of Medicinal Chemistry*. 2003; **46**(21): 4377-4392. DOI: 10.1021/jm030085p

[91] Becker OM, Dhanoa DS, Marantz Y, Chen D, Shacham S, Cheruku S, et al. An integrated in silico 3D model-driven discovery of a novel, potent, and selective amidosulfonamide 5-HT<sub>1A</sub> agonist (PRX-00023) for the treatment of anxiety and depression. *Journal of Medicinal Chemistry*. 2006; **49**(11): 3116-3135. DOI: 10.1021/jm0508641

[92] Kazantsev AV, Karamertzanis PG, Adjiman CS, Pantelides CC, Price SL, Galek PT, et al. Successful prediction of a model pharmaceutical in the fifth blind test of crystal structure prediction. *International Journal of Pharmaceutics*. 2011; **418**(2):168-178. DOI: 10.1016/j.ijpharm.2011.03.058

[93] Marriott DP, Dougall IG, Meghani P, Liu Y-J, Flower DR. Lead generation

using pharmacophore mapping and three-dimensional database searching: Application to muscarinic M<sub>3</sub> receptor antagonists. *Journal of Medicinal Chemistry*. 1999; **42**(17):3210-3216. DOI: 10.1021/jm980409n

[94] Lyne PD, Kenny PW, Cosgrove DA, Deng C, Zabludoff S, Wendoloski JJ, et al. Identification of compounds with nanomolar binding affinity for checkpoint kinase-1 using knowledge-based virtual screening. *Journal of Medicinal Chemistry*. 2004; **47**(8): 1962-1968. DOI: 10.1021/jm030504i

[95] Schröder JR, Klinger A, Oellien F, Marhöfer RJ, Duszenko M, Selzer PM. Docking-based virtual screening of covalently binding ligands: An orthogonal lead discovery approach. *Journal of Medicinal Chemistry*. 2013; **56**(4):1478-1490. DOI: 10.1021/jm3013932

[96] Wang L, Gu Q, Zheng X, Ye J, Liu Z, Li J, et al. Discovery of new selective human aldose reductase inhibitors through virtual screening multiple binding pocket conformations. *Journal of Chemical Information and Modeling*. 2013; **53**(9):2409-2422. DOI: 10.1021/ci400322j

[97] Dadashpour S, Tuylu Kucukkilinc T, Unsal Tan O, Ozadali K, Irannejad H, Emami S. Design, synthesis and in vitro study of 5,6-diaryl-1,2,4-triazine-3-ylthioacetate derivatives as COX-2 and  $\beta$ -amyloid aggregation inhibitors. *Archiv der Pharmazie*. 2015; **348**(3): 179-187. DOI: 10.1002/ardp.201400400

[98] Yuriev E, Agostino M, Ramsland PA. Challenges and advances in computational docking: 2009 in review. *Journal of Molecular Recognition*. 2011; **24**(2):149-164. DOI: 10.1002/jmr.1077

[99] Ren J, He Y, Chen W, Chen T, Wang G, Wang Z, et al. Thermodynamic and structural characterization of halogen bonding in

protein-ligand interactions: A case study of PDE5 and its inhibitors. *Journal of Medicinal Chemistry*. 2014;**57**(8): 3588-3593. DOI: 10.1021/jm5002315

[100] Yang Y, Xu Z, Zhang Z, Yang Z, Liu Y, Wang J, et al. Like-charge guanidinium pairing between ligand and receptor: An unusual interaction for drug discovery and design? *The Journal of Physical Chemistry B*. 2015;**119**(36): 11988-11997. DOI: 10.1021/acs.jpcc.5b04130

[101] Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Arthurs S, Colson AB, et al. Deciphering common failures in molecular docking of ligand-protein complexes. *Journal of Computer-Aided Molecular Design*. 2000;**14**(8):731-751. DOI: 10.1023/A:1008158231558

[102] Spyrakis F, Cavasotto CN. Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration and active site water molecules description. *Archives of Biochemistry and Biophysics*. 2015; **583**:105-119. DOI: 10.1016/j.abb.2015.08.002