Published in final edited form as:
Curr Comput Aided Drug Des. 2011 June 1; 7(2): 146–157.

Molecular Docking: A powerful approach for structure-based drug discovery

Xuan-Yu Meng[1,2], **Hong-Xing Zhang**[1,*], **Mihaly Mezei**[3], and **Meng Cui**[2,*]

[1]State Key Laboratory of Theoretical and Computational Chemistry, Institute of Theoretical Chemistry, Jilin University, Changchun, 130023, China

[2]Department of Physiology and Biophysics, Virginia Commonwealth University, 1101 East Marshall Street, PO Box 980551, Richmond, VA 23298, USA

[3]Department of Structural and Chemical Biology, Mount Sinai School of Medicine, Box 1677, New York, NY 10029, USA

Abstract

Molecular docking has become an increasingly important tool for drug discovery. In this review, we present a brief introduction of the available molecular docking methods, and their development and applications in drug discovery. The relevant basic theories, including sampling algorithms and scoring functions, are summarized. The differences in and performance of available docking software are also discussed. Flexible receptor molecular docking approaches, especially those including backbone flexibility in receptors, are a challenge for available docking methods. A recently developed Local Move Monte Carlo (LMMC) based approach is introduced as a potential solution to flexible receptor docking problems. Three application examples of molecular docking approaches for drug discovery are provided.

Introduction

The completion of the human genome project has resulted in an increasing number of new therapeutic targets for drug discovery. At the same time, high-throughput protein purification, crystallography and nuclear magnetic resonance spectroscopy techniques have been developed and contributed to many structural details of proteins and protein–ligand complexes. These advances allow the computational strategies to permeate all aspects of drug discovery today [1-5], such as the virtual screening (VS) techniques [6] for hit identification and methods for lead optimization. Compared with traditional experimental high-throughput screening (HTS), VS is a more direct and rational drug discovery approach and has the advantage of low cost and effective screening [7-9]. VS can be classified into ligand-based and structure-based methods. When a set of active ligand molecules is known and little or no structural information is available for targets, the ligand-based methods, such as pharmacophore modeling and quantitative structure activity relationship (QSAR) methods can be employed. As to structure-based drug design, molecular docking is the most common method which has been widely used ever since the early 1980s [10]. Programs based on different algorithms were developed to perform molecular docking studies, which have made docking an increasingly important tool in pharmaceutical research. Various excellent reviews on docking have been published in the past [5, 11-14], and many comparison studies were conducted to evaluate the relative performance of the programs [15-18].

[*]To whom correspondence should be addressed. zhanghx@mail.jlu.edu.cn (H.Z.); mcui@vcu.edu (M.C.) .

The molecular docking approach can be used to model the interaction between a small molecule and a protein at the atomic level, which allow us to characterize the behavior of small molecules in the binding site of target proteins as well as to elucidate fundamental biochemical processes [19]. The docking process involves two basic steps: prediction of the ligand conformation as well as its position and orientation within these sites (usually referred to as *pose*) and assessment of the binding affinity. These two steps are related to sampling methods and scoring schemes, respectively, which will be discussed in the theory section.

Knowing the location of the binding site before docking processes significantly increases the docking efficiency. In many cases, the binding site is indeed known before docking ligands into it. Also, one can obtain information about the sites by comparison of the target protein with a family of proteins sharing a similar function or with proteins co-crystallized with other ligands. In the absence of knowledge about the binding sites, cavity detection programs or online servers, e.g. GRID[20, 21], POCKET [22], SurfNet [23, 24], PASS [25] and MMC [26] can be utilized to identify putative active sites within proteins. Docking without any assumption about the binding site is called blind docking.

The early elucidation for the ligand-receptor binding mechanism is the lock-and-key theory proposed by Fischer [27], in which the ligand fits into the receptor like lock and key. The earliest reported docking methods [10] were based on this theory and both the ligand and receptor were treated as rigid bodies accordingly. Then the “induced-fit” theory [28, 29] created by Koshland takes the lock-and-key theory a step further, stating that the active site of the protein is continually reshaped by interactions with the ligands as the ligands interact with the protein. This theory suggests that the ligand and receptor should be treated as flexible during docking. Consequently, it could describe the binding events more accurately than the rigid treatment.

Considering the limitation of computer resources, docking has been performed with a flexible ligand and a rigid receptor for a long time, and remains the most popular method in use [7, 30-35]. Recently many efforts have been made to deal with the flexibility of the receptor [36-42], however, flexible receptor docking, especially backbone flexibility in receptors, still presents a major challenge for available docking methods. In our study, we propose a Local Move Monte Carlo (LMMC) approach as a potential solution to flexible receptor docking problems.

Theory of docking

Essentially, the aim of molecular docking is to give a prediction of the ligand-receptor complex structure using computation methods. Docking can be achieved through two interrelated steps: first by sampling conformations of the ligand in the active site of the protein; then ranking these conformations via a scoring function. Ideally, sampling algorithms should be able to reproduce the experimental binding mode and the scoring function should also rank it highest among all generated conformations. From these two perspectives, we give a brief overview of basic docking theory.

Sampling algorithms

With six degrees of translational and rotational freedom as well as the conformational degrees of freedom of both the ligand and protein, there are a huge number of possible binding modes between two molecules. Unfortunately, it would be too expensive to computationally generate all the possible conformations. Various sampling algorithms have been developed and widely used in molecular docking software (Table 1).

Matching algorithms (MA) [43-45] based on molecular shape map a ligand into an active site of a protein in terms of shape features and chemical information. The protein and the ligand are represented as pharmacophores. Each distance of the pharmacophore within the protein and ligand is calculated for a match; new ligand conformations are governed by the distance matrix between the pharmacophore and the corresponding ligand atoms. Chemical properties, like hydrogen-bond donors and acceptors, can be taken into account during the match. Matching algorithms have the advantage of speed; thus they may be used for the enrichment of active compounds from large libraries [7]. Matching algorithms for ligand docking are available in DOCK [10], FLOG [46], LibDock [47] and SANDOCK [48] programs.

Incremental construction (IC) [30, 49, 50] methods put the ligand into an active site in a fragmental and incremental fashion. The ligand is divided into several fragments by breaking its rotatable bonds and then one of these fragments is selected to dock into the active site first. This anchor is usually the largest fragment or the piece which may have significant functional role or interaction with protein. The remaining fragments can be added incrementally. Different orientations are generated to fit in the active site, which realizes the flexibility of the ligand. The incremental construction method has been used in DOCK 4.0 [51], FlexX [30], Hammerhead [52], SLIDE [53] and eHiTS [54].

In addition to IC, Multiple Copy Simultaneous Search (MCSS) [55, 56] and LUDI [57] are fragment-based methods for the *de novo* design of ligands and modifications of known ligands that may enhance their binding to the target protein. MCSS makes 1,000 to 5,000 copies of a functional group, which are randomly placed in the binding site of interest and subjected to simultaneous energy minimization and/or quenched molecular dynamics in the forcefield of the protein. Copies only interact with the proteins and any interactions among the copies are omitted. Consequently a set of energetically favorable binding sites and orientations for the functional group is identified based on the interaction energies. The binding site is mapped by using different functional groups. New molecules which perfectly match the binding site can be designed through the linkage of those different functional groups.

LUDI focuses on the hydrogen bonds and hydrophobic contacts which could be formed between the ligand and protein. Its central concept are interaction sites, which are discrete positions in space suitable for forming hydrogen bonds or for filling a hydrophobic pocket [57]. A set of interaction sites is generated either by searching the database or using the rules. The fragment is then fitted onto the interaction sites and evaluated by distance criteria. The final step is the connection of some or all of the fitted fragments to a single molecule.

Stochastic methods search the conformational space by randomly modifying a ligand conformation or a population of ligands. Monte Carlo (MC) and genetic algorithms are two typical algorithms that belong to the class of stochastic methods.

Monte Carlo (MC) [58, 59] methods generate poses of the ligand through bond rotation, rigid-body translation or rotation. The conformation obtained by this transformation is tested with an energy- based selection criterion. If it passes the criterion, it will be saved and further modified to generate next conformation. The iterations will proceed until the pre-defined quantity of conformations is collected. The main advantage of MC is that the change can be quite large allowing the ligand to cross the energy barriers on the potential energy surface, a point that isn’t achieved easily by molecular dynamics based simulation methods. Examples of applying the Monte Carlo methods include an earlier version of AutoDock [60], ICM [61], QXP [62] and Affinity [63].

Genetic algorithms (GA) [31, 32, 64] form another class of well-known stochastic methods. The idea of the GA stems from Darwin's theory of evolution. Degrees of freedom of the ligand are encoded as binary strings called genes. These genes make up the 'chromosome' which actually represents the pose of the ligand. Mutation and crossover are two kinds of genetic operators in GA. Mutation makes random changes to the genes; crossover exchanges genes between two chromosomes. When the genetic operators affect the genes, the result is a new ligand structure. New structures will be assessed by scoring function, and the ones that survived (i.e., exceeded a threshold) can be used for the next generation. Genetic algorithms have been used in AutoDock [31], GOLD [65], DIVALI [66] and DARWIN [67].

Molecular dynamics (MD) [68-70] is widely used as a powerful simulation method in many fields of molecular modeling. In the context of docking, by moving each atom separately in the field of the rest atoms, MD simulation represents the flexibility of both the ligand and protein more effectively than other algorithms. However, the disadvantage of MD simulations is that they progress in very small steps and thus have difficulties in stepping over high energy conformational barriers, which may lead to inadequate sampling. On the other hand, MD simulations are often efficient at local optimization. Thus a current strategy is to use random search in order to identify the conformation of the ligand, followed by the further subtle MD simulations.

Scoring functions

The purpose of the scoring function is to delineate the correct poses from incorrect poses, or binders from inactive compounds in a reasonable computation time. However, scoring functions involve estimating, rather than calculating the binding affinity between the protein and ligand and through these functions, adopting various assumptions and simplifications. Scoring functions can be divided in force-field-based, empirical and knowledge-based scoring functions [5]. Table 2 shows some examples of scoring function formulae belonging to those three classes of scoring functions respectively.

Classical force-field-based scoring functions [71-73] assess the binding energy by calculating the sum of the non-bonded (electrostatics and van der Waals) interactions. The electrostatic terms are calculated by a Coulombic formulation. Since such point charge calculations have problems in modeling the protein's real environment a distance-dependent dielectric function is generally used to modulate the contribution of charge–charge interactions. The van der Waals terms are described by a Lennard-Jones potential function. Adopting different parameter sets for the Lennard-Jones potential can vary the "hardness" of the potential which controls how close a contact between protein and ligand atoms can be acceptable. Force-field-based scoring functions also have the problem of slow computational speed. Thus cut-off distance is used to handle the non-bonded interactions. This also results in decreasing the accuracy of long-range effects involved in binding.

Extensions of force-field-based scoring functions consider the hydrogen bonds, solvations and entropy contributions. Software programs, such as DOCK [10, 50, 51, 74], GOLD [65] and AutoDock [31], offer users such functions. They have some differences in the treatment of hydrogen bonds, the form of the energy function etc.. Furthermore, the results of docking with force-field-based functions can be further refined with other techniques, such as linear interaction energy [75] and free-energy perturbation methods (FEP) [71, 76] to improve the accuracy in predicting binding energies.

In empirical scoring functions [77-81], binding energy decomposes into several energy components, such as hydrogen bond, ionic interaction, hydrophobic effect and binding entropy. Each component is multiplied by a coefficient and then summed up to give a final score. Coefficients are obtained from regression analysis fitted to a test set of ligand-protein complexes with known binding affinities.

Empirical scoring functions have relatively simple energy terms to evaluate. However, it is unclear as to how well they are suited for ligand-protein complexes beyond the training set. Additionally, each term in empirical scoring functions may be treated in a different manner by different software, and the numbers of the terms included are also different. LUDI [57], PLP [78, 79, 82], ChemScore [83] are examples derived from empirical scoring functions

Knowledge-based scoring functions [84-89] use statistical analysis of ligand-protein complexes crystal structures to obtain the interatomic contact frequencies and/or distances between the ligand and protein. They are based on the assumption that the more favorable an interaction is, the greater the frequency of occurrence will be. These frequency distributions are further converted into pairwise atom-type potentials. The score is calculated by favoring preferred contacts and penalizing repulsive interactions between each atom in the ligand and protein within a given cutoff.

The appeal of knowledge-based functions is computational simplicity, which can be exploited to screen large compound databases. They can also model some uncommon interactions like sulphur-aromatic or cation-π, which are often poorly handled in empirical approaches. However, they are still faced with the problem that some interactions are underrepresented in the limited training sets of crystal structures as well as by the bias inherent in the selection of proteins for successful structure determination thus the obtained parameters may not be suitable for widespread use, especially with interactions involving metals or halogens. PMF [84], DrugScore [90], SMoG [91] and Bleep [85] are examples of knowledge-based functions which differ mainly in the size of training sets, the form of the energy function, the definition of atom types, distance cutoff or other parameters.

Consensus scoring [92] is a recent strategy that combines several different scores to assess the docking conformation. A pose of ligand or a potential binder could be accepted when it scores well under a number of different scoring schemes. Consensus scoring usually substantially improves enrichments (i.e., the percentage of strong binder among the high-scoring ligands) in virtual screening, and improves the prediction of bound conformations and poses [93]. However, the prediction of binding energies might still be inaccurate. Also, the usefulness of consensus scoring diminishes when terms in different scoring functions are significantly correlated [5, 93]. CScore [94] is an example of which combines DOCK, ChemScore, PMF, GOLD, and FlexX scoring functions.

Typical scoring functions face the problem of affinity prediction partly because of the limited treatment of solvation effect. One of the ways to solve this problem is physics-based scoring, e.g. MM-PB/SA and MM-GB/SA (MM stands for molecular mechanics, PB and GB for Poisson-Boltzmann and Generalized Born, respectively, SA for solvent-accessible surface area), which is involved in rescoring or lead optimization to improve the accuracy of binding affinity prediction. Promising results were obtained using MM-PB/SA [95, 96] or MM-GB/SA [97] in some studies. However, recently Guimarães and Mathiowetz reported that the GB/SA model poorly estimated protein desolvation on certain systems, while incorporating WaterMap into the MM-GB/SA method instead of GB/SA protein desolvation gave the best ranking result [98]. Singh and Warshel compared several methods for evaluating the affinity of protein-ligand complexes and suggested that PDLD/S-LRA/β (protein dipoles Langevin dipoles linear response approximation) appears to offer an appealing option for the final stages of massive VS and in contrast, PB/SA appears to provide erroneous estimates of the absolute binding energies because of its incorrect estimation of entropies and the problematic treatment of electrostatic energies [99].

Docking methodologies

Rigid ligand and rigid receptor docking

When the ligand and receptor are both treated as rigid bodies, the search space is very limited, considering only three translational and three rotational degrees of freedom. In this case, ligand flexibility could be addressed by using a pre-computed a set of ligand conformations, or by allowing for a degree of atom–atom overlap between the protein and ligand. The early versions of DOCK [10, 50, 51, 74], FLOG [46] and some protein-protein docking programs, such as FTDOCK [100], adopted such a method that kept the ligand and receptor rigid during the process of the docking.

DOCK is the first automated procedure for docking a molecule into a receptor site and is being continuously developed. It characterizes the ligand and receptor as sets of spheres which could be overlaid by means of a clique detection procedure [101]. Geometrical and chemical matching algorithms are used, and the ligand-receptor complexes can be scored by accounting for steric fit, chemical complementation or pharmacophore similarity. Within its improved versions, incremental construction method and exhaustive search are added to consider the ligand flexibility. The exhaustive search randomly generates a user-defined number of conformers as a multiple of the number of rotatable bonds in the ligand. With respect to scoring, the latest version DOCK 6.4 has included both an AMBER-derived force-field scoring with implicit solvent [102] and GB/SA, PB/SA solvation scoring [97, 103].

FLOG generates ligand conformations on the basis of distance geometry and uses a clique-finding algorithm to calculate the sets of distances. Up to 25 explicit conformations of the ligand could be used to dock for some flexibility. FLOG allows users to define essential points which must be paired with a ligand atom. This approach is useful if an important interaction is already known before docking. Conformations are scored with a function considering van der Waals, electrostatics, hydrogen bonding and hydrophobic interactions.

Flexible ligand and rigid receptor docking

For systems whose behavior follows the induced fit paradigm [28, 29], it is of vital importance to consider the flexibilities of both the ligand and receptor since in that case both the ligand and receptor change their conformations to form a minimum energy perfect-fit complex. However, the cost is very high when the receptor is also flexible. Thus the common approach, also a trade-off between accuracy and computational time, is treating the ligand as flexible while the receptor is kept rigid during docking. Almost all the docking programs have adopted this methodology, such as AutoDock [31], FlexX [30].

AutoDock 3.0 incorporates Monte Carlo simulated annealing, evolutionary, genetic and Lamarckian genetic algorithm methods to model the ligand flexibility while keeping the receptor rigid. The scoring function is based on the AMBER force field, including van der Waals, hydrogen bonding, electrostatic interactions, conformational entropy and desolvation terms. Each term is weighted using an empirical scaling factor obtained from experimental data. AutoDock 4.0 is able to model receptor flexibility by allowing side-chains to move. Additionally, interaction of protein-protein docking could be evaluated in this version of AutoDock. AutoDock Vina was recently released as the latest version for molecular docking and virtual screening [104]. By redocking the 190 receptor-ligand complexes that had been used as a training set for the AutoDock 4, AutoDock Vina simultaneously showed approximately a two orders exponential improvement of magnitude in speed and a significantly better accuracy of the binding mode prediction.

FlexX uses an incremental construction algorithm to sample ligand conformations. The base fragment is first docked into the active site by matching hydrogen bond pairs and metal and aromatic ring interactions between the ligand and protein. Then the remaining components are incrementally built-up in accordance with a set of predefined rotatable torsion angles to account for ligand flexibility. The FlexX scoring function is based on Böhm's work [105]. Its current version includes terms of electrostatic interactions, directional hydrogen bonds, rotational entropy, and aromatic and lipophilic interactions. The interactions between functional groups are also taken into account through assigning the type and geometry for groups.

Flexible ligand and flexible receptor docking

The intrinsic mobility of proteins has been proved to be closely related to ligand binding behavior and it has been reviewed by Teague [106]. Incorporating the receptor flexibility is significant challenge in the field of docking. Ideally, using MD simulations could model all the degrees of freedom in the ligand-receptor complex. But MD has the problem of inadequate sampling that we mentioned earlier. Another hurdle is its high computational expense, which prevents this method from being used in the screening of large chemical database.

In addition to the historic induced fit several theoretical models, conformer selection and conformational induction, have been proposed to illustrate the flexible ligand-protein binding process. According to the definition given by Teague [106], conformer selection refers to a process when a ligand selectively binds to a favorable conformation from a number of protein conformations; conformational induction describes a process in which the ligand converts the protein into a conformation that it would not spontaneously adopt in its unbound state. In some cases, this conformational conversion can be likened to a partial refolding of the protein.

Various methods are currently available to implement the receptor flexibility (Table 3). The simplest one is so-called "soft-docking" [37, 107, 108], decreases the van der Waals repulsion energy term in the scoring function to allow for a degree of atom-atom overlap between the receptor and ligand. For example, the LJ 8-4 potential in GOLD and smooth potential in AutoDock 3.0 belong to this class. This method may not include adequate flexibility. Nevertheless, it has the advantage of computational efficiency as the receptor coordinates are fixed, simply by adjusting van der Waals parameters.

Utilizing rotamer libraries [109, 110] is another approach to modeling receptor flexibility. Rotamer libraries include a set of side-chain conformations which are usually determined from statistical analysis of structural experimental data. The advantage of using rotamers is the relative speed in sampling, and the avoiding of minimization barriers. ICM (Internal Coordinates Mechanics) [61] is a program using rotamer libraries with the biased probability methodology [111], coupled with Monte Carlo search of the ligand conformation.

AutoDock 4 [112] adopts a simultaneous sample method to deal with side chain flexibility. Several side chains of the receptor can be selected by users and simultaneously sampled with a ligand using the same methods. Other portions of the receptor are treated rigidly with a grid energy map during sampling. Grid energy map introduced by Goodford [20] is used to store energy information of the receptor and simplify interaction energy calculation between ligand and receptor.

Still another way to deal with the protein flexibility is to use an ensemble of protein conformations, which corresponds to the theory of conformer selection [113, 114]. A ligand is separately docked into a set of rigid protein conformations rather than a single one, and the results are merged depending on the method of choice [115]. This method was originally implemented in DOCK, which generates an average potential energy grid of the ensemble

[113] and is extended in many programs in different ways. For example, FlexE [38] collects multiple crystal structures of a certain protein, merging the similar parts while marking the dissimilar areas as different alternatives. During the incremental construction of a ligand discrete protein conformations are sampled in a combinatorial fashion. The highest scoring protein structure is selected based on a comparison between the ligand and each alternative.

Hybrid method is another practical strategy to model receptor flexibility. One example is Glide [33], a very popular program in the field of docking. Glide designs a series of hierarchical filters to search the possible poses and orientations of the ligand within the binding site of the receptor. Ligand flexibility is handled by an exhaustive search of the ligand torsion angle space. Initial ligand conformations are selected based on torsion energies and docked into receptor binding sites with soft potentials. Then a rotamer exploration is used to further model receptor flexibility [36]. IFREDA [115] utilizes a hybrid method that combines soft potential and multiple receptor conformations, accounting for receptor flexibility. Other programs, like QXP [62] and Affinity [63], perform a Monte Carlo search of ligand conformations followed by a minimization step. During minimization, the user-defined parts of the protein are allowed to move in order to avoid atom clashes between the ligand and receptor. SLIDE [53] is designed to incorporate flexibility with the ability to remove clashes by directed, single bond rotation of either the ligand or the side chains of the protein. An optimization approach based on the mean-field theory is applied to model induced-fit complementarities between the ligand and protein.

Methods mentioned above either include only side chain flexibility or full flexibility of the receptor. We have known that loops forming active sites play an important role in ligand binding. In some cases the loop may undergo dramatic conformational change whereas in other portions of the receptor there is little change upon ligand binding. For this situation, side chain flexibility methods fail to sample the correct protein conformation and full flexibility seems to be a computational waste. Figure 1 shows superimposed crystal structures of triosephosphate isomerase as an example. The active site of triosephosphate isomerase has an 11-residue loop which moves 7Å upon ligand binding [116]. However, the rest of the enzyme has no movement in comparison to their apo and holo structures. Several enzyme families also involve loop rearrangement within the active site responsible for ligand binding, such as Bromodomain, an extensive family related to acetyl-lysine binding, or Dihydrofolate reductase, responsible for the maintenance of the cellular pools of tetrahydrofolate, as well as other kinds of kinases [117, 118]. In the next section, we present the Local Move Monte Carlo (LMMC) loop sampling method, a new approach which focuses on sampling ligand conformation within loop-containing active sites.

Local Move Monte Carlo sampling for flexible receptor docking

Local move (also referred to as ‘window move’) starts with changing one torsion angle (called the driver torsion) followed by the adjustment of the six subsequent torsions to allow the rest of the chain to remain in its original position while preserving all bond lengths and bond angles (Figure 2). The pioneering work on local move was done by Go and Scheraga [119], who developed a solution for the system of equations defining the values of the six torsion angles that preserve the backbone bond lengths and angles. Hoffmann and Knapp first applied the local move method in a MC simulation of polyalanine folding that included a suitable Jacobian [120], required for maintaining detailed balance. They demonstrated that this method samples the conformational space more efficiently than single move [121]. The method has been further tested on proline-containing peptides [122], proteins and nucleic acids [123]. Mezei introduced the ‘reverse proximity criterion’ for filtering all possible loop closure solutions to select the most structurally conservative one and tested it on a solvated lipid bilayer [124].

We have developed an improved local move Monte Carlo (LMMC) loop sampling approach for loop predictions. The method generates loop conformations based on simple moves of the torsion angles of side chains and local moves of the backbones of loops. To reduce the computational costs for energy evaluations, we developed a grid-based force field to represent the protein environment and solvation effect. Simulated annealing has been used to enhance the efficiency of the LMMC loop sampling and identify low-energy loop conformations. The prediction quality was evaluated on a set of protein loops with a known crystal structure that has been previously used by others to test different loop prediction methods. The results show that this approach can reproduce the experimental results with root mean square deviation (RMSD) within 1.8 Å for the all the test cases [125]. Figure 3 shows the loop structures of 2act (198-205) sampled by the LMMC method. This LMMC loop prediction approach could be useful for flexible receptor docking. In our future studies, we will develop our LMMC based molecular docking approach, which samples not only the side chains but also the backbone loops in the binding site of proteins and flexible ligands as well. A flowchart of the LMMC based molecular docking approach is given in Figure 4.

Application examples of molecular docking for drug discovery

Molecular docking has been the most widely employed technique. Though the main application lies in structure-based virtual screening for identification of new active compounds towards a particular target protein, in which it has produced a number of success stories [126], it is actually not a stand-alone technique but is normally embedded in a workflow of different *in silico* as well as experimental techniques [127]. Several research groups focus on evaluating of the performance of various docking programs or on making improvements to the scoring functions when experimental testing has already been done. Such efforts could give meaningful guidance to choose the methodology for a particular target system. Docking, combined with other computational techniques and experimental data, also could be involved in analyzing drug metabolism to obtain some useful information from the cytochrome P450 system [128-130], for example. In the following, three examples of successful applications of docking are presented.

DNA gyrase is a bacterial enzyme that introduces negative supercoils into bacterial DNA and unwinds of DNA, thus being studied as antibacterial target. HTS failed to find novel inhibitors of DNA gyrase. Boehm et. al. used *de novo* design for this enzyme and successfully obtained several new inhibitors [131]. Firstly, 3D complex structures of DNA gyrase with known inhibitors, ciprofloxacin and novobiocin, were carefully analyzed to get a common binding pattern, in which both inhibitors donate one hydrogen bond to Asp73 and accept one hydrogen bond from a conserved water molecule. In addition, some lipophilic fragments should be included in the molecule to have lipophilic interaction with the receptor. Based on this information, LUDI and CATALYST were employed to search the Available Chemicals Directory (ACD) and a part of the Roche compound inventory (RIC), respectively, and collected about 600 compounds. Close analogs of these compounds were also considered, thus in total 3000 compounds were further tested using biased screening. Consequently 150 hits were selected and clustered into 14 classes of which 7 classes were proven to be the true and novel inhibitors. Subsequent hit optimization relied strongly on the knowledge of 3D structures of the binding site and eventually generated a series of highly potent DNA gyrase inhibitors.

Another example is focused on the validation of docking and scoring applied in cytochromes P450 and other heme-containing proteins [132]. Docking against heme-containing complexes appears to be difficult because certain ligands coordinate directly to the heme iron atom and the precise energetics of this contact for different chelating groups needs to be properly balanced with other energetic terms, and in the case of the P450s, the environment above the heme group is very hydrophobic compared to other enzymes and some scoring functions and docking methods perform poorly on interactions driven entirely by lipophilic contacts. In this study, 45 complexes from the PDB database comprising heme-containing proteins and ligands were selected. The native ligands were removed and then docked into the defined active cavities using the GOLD [65] software which employs genetic algorithms to generate ligand conformations. The scoring functions used to rank the docking poses were Goldscore [32] and Chemscore [65]. The results show that the success rates are 64% and 57% for Chemscore and Goldscore respectively, which is significantly lower than the value of 79% observed with both scoring functions for the full GOLD validation set. Additionally, it is apparent from the data that the search algorithm was very unlikely to be responsible for the failure in docking. Further research indicated that re-parameterization of metal-acceptor interactions and lipophilicity of planar nitrogen atoms in the scoring functions resulted in a significant increase in the percentage of successful docking poses against the heme binding proteins (Chemscore 73%, Goldscore 65%), which might be useful in docking applications on P450 enzymes and other heme-binding proteins.

Concerning VS and HTS, comparative research has been done by Doman et al. [133]. Both VS and HTS were applied to screen the inhibitors of the protein tyrosine phosphatase-1B (PTP-1B). For the HTS a library of approximately 400,000 compounds from a corporate collection were screened. Some 85 compounds were found with IC_{50} values less than 100 μM, corresponding to a hit rate of 0.021%. And the most active had an IC_{50} value of 4.2 μM. For VS, 235,000 commercially available molecules were docked into the crystal structure of PTP-1B (PDB code 1pty) using the Northwestern University version [134-137] of DOCK3.5 [102, 138]. After docking, the top-scoring 1000 molecules (500 for the ACD and 500 for the combined BioSpecs and Maybridge databases) were considered for further evaluation. A total of 889 molecules were actually available, and after visual inspection 365 compounds were chosen for testing. Of these, 127 molecules were found to be active with IC_{50} <100 μM, corresponding to a hit rate of 34.8%. Structure-based docking therefore enriched the hit rate by 1700-fold over random screening. Another point that should be noted is that the hits from VS and HTS are very different from each other, which implies combination of VS and HTS may be more helpful for lead discovery.

Concluding Remarks

Receptor flexibility, especially backbone flexibility and movement of several key secondary elements of the receptor involving ligand binding and the catalyst, is still a major hurdle in docking studies. Some methods to deal with side chain flexibility have been proven effective and adequate in certain cases. With respect to global flexibility, an ensemble of proteins is a popular solution which accords with the viewpoint of conformer selection. It requires an efficient way to obtain and select reliable protein structures used for docking, which means structures that the ligand can fit in should be included in the ensembles. Besides, computational cost is another limitation for this method. LMMC could be an appropriate method for sampling a ligand within loop-containing active sites since loop tends to be more flexible and hard to model using existing approaches especially due to their possibly dramatic movements. Another advantage is the adjustment of the extent of flexibility. Either the side chain or full movement of the loop can be directly controlled by users.

Scoring function is a fundamental component worth being further improved upon in docking. Successful application examples show that computational approaches have the power to screen hits from a huge database and design novel small molecules. However, the realistic interactions between small molecules and receptors are still relied on experimental technology. Accurate as well as low computational cost scoring functions may bring docking application to a new stage.

Acknowledgments

We gratefully acknowledge financial support from National Institute Health (DC008996 and S10RR027411 to M.C.). The computations were partially supported by grants of the National Center for Supercomputing Applications (MCB070095T and MCB090019 to M.C.).

References

[1]. Jorgensen WL. The many roles of computation in drug discovery. Science. 2004; 303(5665): 1813–1818. [PubMed: 15031495]

[2]. Bajorath J. Integration of virtual and high-throughput screening. Nat Rev Drug Discov. 2002; 1(11):882–894. [PubMed: 12415248]

[3]. Walters WP, Stahl MT, Murcko MA. Virtual screening - an overview. Drug Discov. Today. 1998; 3:160–178.

[4]. Langer T, Hoffmann RD. Virtual screening: an effective tool for lead structure discovery? Curr Pharm Des. 2001; 7(7):509–527. [PubMed: 11375766]

[5]. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov. 2004; 3(11):935–949. [PubMed: 15520816]

[6]. Gohlke H, Klebe G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew Chem Int Ed Engl. 2002; 41(15):2644–2676. [PubMed: 12203463]

[7]. Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. Br J Pharmacol. 2008; 153(Suppl 1):S7–26. [PubMed: 18037925]

[8]. Shoichet, BK.; McGovern, SL.; Wei, B.; Irwin, JJ. Hits, leads and artifacts from virtual and high throughput screening. 2002. Molecular Informatics: Confronting Complexity.

[9]. Bailey D, Brown D. High-throughput chemistry and structure-based design: survival of the smartest. Drug Discov Today. 2001; 6(2):57–59. [PubMed: 11166243]

[10]. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. J Mol Biol. 1982; 161(2):269–288. [PubMed: 7154081]

[11]. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: An overview of search algorithms and a guide to scoring functions. Proteins. 2002; 47(4):409–443. [PubMed: 12001221]

[12]. Coupez B, Lewis RA. Docking and scoring--theoretically easy, practically impossible? Curr Med Chem. 2006; 13(25):2995–3003. [PubMed: 17073642]

[13]. Kontoyianni M, Madhav P, Suchanek E, Seibel W. Theoretical and practical considerations in virtual screening: a beaten field? Curr Med Chem. 2008; 15(2):107–116. [PubMed: 18220766]

[14]. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. Annu Rev Biophys Biomol Struct. 2003; 32:335–373. [PubMed: 12574069]

[15]. ten Brink T, Exner TE. Influence of protonation, tautomeric, and stereoisomeric states on protein-ligand docking results. J Chem Inf Model. 2009; 49(6):1535–1546. [PubMed: 19453150]

[16]. Cross JB, Thompson DC, Rai BK, Baber JC, Fan KY, Hu Y, Humblet C. Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. J Chem Inf Model. 2009; 49(6):1455–1474. [PubMed: 19476350]

[17]. Li X, Li Y, Cheng T, Liu Z, Wang R. Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes. J Comput Chem. 2010; 31(11):2109–2125. [PubMed: 20127741]

[18]. Plewczynski D, Lazniewski M, Augustyniak R, Ginalski K. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. J Comput Chem. 2010 doi: 10.1002/jcc.21643.

[19]. McConkey BJ, Sobolev V, Edelman M. The performance of current methods in ligand-protein docking. Current Science. 2002; 83:845–855.

[20]. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem. 1985; 28(7):849–857. [PubMed: 3892003]

[21]. Kastenholz MA, Pastor M, Cruciani G, Haaksma EE, Fox T. GRID/CPCA: a new computational tool to design selective ligands. J Med Chem. 2000; 43(16):3033–3044. [PubMed: 10956211]

[22]. Levitt DG, Banaszak LJ. POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. J Mol Graph. 1992; 10(4):229–234. [PubMed: 1476996]

[23]. Laskowski RA. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. J Mol Graph. 1995; 13(5):323–330. 307–328. [PubMed: 8603061]

[24]. Glaser F, Morris RJ, Najmanovich RJ, Laskowski RA, Thornton JM. A method for localizing ligand binding pockets in protein structures. Proteins. 2006; 62(2):479–488. [PubMed: 16304646]

[25]. Brady GP Jr. Stouten PF. Fast prediction and visualization of protein binding pockets with PASS. J Comput Aided Mol Des. 2000; 14(4):383–401. [PubMed: 10815774]

[26]. Mezei M. A new method for mapping macromolecular topography. J Mol Graph Model. 2003; 21(5):463–472. [PubMed: 12543141]

[27]. Fischer E. Einfluss der configuration auf die wirkung derenzyme. Ber. Dt. Chem. Ges. 1894; 27:2985–2993.

[28]. Koshland DE Jr. Correlation of Structure and Function in Enzyme Action. Science. 1963; 142:1533–1541. [PubMed: 14075684]

[29]. Hammes GG. Multiple conformational changes in enzyme catalysis. Biochemistry. 2002; 41(26): 8221–8228. [PubMed: 12081470]

[30]. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. J Mol Biol. 1996; 261(3):470–489. [PubMed: 8780787]

[31]. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of Computational Chemistry. 1998; 19(14):1639–1662.

[32]. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol. 1997; 267(3):727–748. [PubMed: 9126849]

[33]. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. Glide: a new approach for rapid, accurate docking and scoring. 1 Method and assessment of docking accuracy. J Med Chem. 2004; 47(7): 1739–1749. [PubMed: 15027865]

[34]. McGann MR, Almond HR, Nicholls A, Grant JA, Brown FK. Gaussian docking functions. Biopolymers. 2003; 68(1):76–90. [PubMed: 12579581]

[35]. Perola E, Walters WP, Charifson PS. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins. 2004; 56(2):235–249. [PubMed: 15211508]

[36]. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. Novel procedure for modeling ligand/receptor induced fit effects. J Med Chem. 2006; 49(2):534–553. [PubMed: 16420040]

[37]. Jiang F, Kim SH. “Soft docking”: matching of molecular surface cubes. J Mol Biol. 1991; 219(1):79–102. [PubMed: 2023263]

[38]. Claussen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. J Mol Biol. 2001; 308(2):377–395. [PubMed: 11327774]

[39]. Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. Med Res Rev. 2006; 26(5):531–568. [PubMed: 16758486]

[40]. Sander T, Liljefors T, Balle T. Prediction of the receptor conformation for iGluR2 agonist binding: QM/MM docking to an extensive conformational ensemble generated using normal mode analysis. J Mol Graph Model. 2008; 26(8):1259–1268. [PubMed: 18203639]

[41]. Subramanian J, Sharma S, C BR. A novel computational analysis of ligand-induced conformational changes in the ATP binding sites of cyclin dependent kinases. J Med Chem. 2006; 49(18):5434–5441. [PubMed: 16942017]

[42]. Subramanian J, Sharma S, C BR. Modeling and selection of flexible proteins for structure-based drug design: backbone and side chain movements in p38 MAPK. ChemMedChem. 2008; 3(2): 336–344. [PubMed: 18081134]

[43]. Brint AT, Willett P. Algorithms for the Identification of Three-Dimensional Maximal Common Substructures. J. Chem. Inf. Comput. Sci. 1987; 27:152–158.

[44]. Fischer D, Norel R, Wolfson H, Nussinov R. Surface motifs by a computer vision technique: searches, detection, and implications for protein-ligand recognition. Proteins. 1993; 16(3):278–292. [PubMed: 8394000]

[45]. Norel R, Fischer D, Wolfson HJ, Nussinov R. Molecular surface recognition by a computer vision-based technique. Protein Eng. 1994; 7(1):39–46. [PubMed: 8140093]

[46]. 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. J Comput Aided Mol Des. 1994; 8(2):153–174. [PubMed: 8064332]

[47]. Diller DJ, Merz KM Jr. High throughput docking for library design and library prioritization. Proteins. 2001; 43(2):113–124. [PubMed: 11276081]

[48]. Burkhard P, Taylor P, Walkinshaw MD. An example of a protein ligand found by database mining: description of the docking method and its verification by a 2.3 A X-ray structure of a thrombin-ligand complex. J Mol Biol. 1998; 277(2):449–466. [PubMed: 9514757]

[49]. DesJarlais RL, Sheridan RP, Dixon JS, Kuntz ID, Venkataraghavan R. Docking flexible ligands to macromolecular receptors by molecular shape. J Med Chem. 1986; 29(11):2149–2153. [PubMed: 3783576]

[50]. Kuntz ID, Leach AR. Conformational analysis of flexible ligands in macromolecular receptor sites. J. Comput. Chem. 1992; 13:730–748.

[51]. Ewing TJ, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des. 2001; 15(5):411–428. [PubMed: 11394736]

[52]. Welch W, Ruppert J, Jain AN. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. Chem Biol. 1996; 3(6):449–462. [PubMed: 8807875]

[53]. Schnecke V, Kuhn LA. Virtual Screening with Solvation and Ligand-Induced Complementarity. Perspectives in Drug Discovery and Design. 2000; 20:171–190.

[54]. Zsoldos Z, Reid D, Simon A, Sadjad BS, Johnson AP. eHiTS: an innovative approach to the docking and scoring function problems. Curr Protein Pept Sci. 2006; 7(5):421–435. [PubMed: 17073694]

[55]. Miranker A, Karplus M. Functionality maps of binding sites: a multiple copy simultaneous search method. Proteins. 1991; 11(1):29–34. [PubMed: 1961699]

[56]. Eisen MB, Wiley DC, Karplus M, Hubbard RE. HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. Proteins. 1994; 19(3):199–221. [PubMed: 7937734]

[57]. Bohm HJ. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. J Comput Aided Mol Des. 1992; 6(6):593–606. [PubMed: 1291628]

[58]. Goodsell DS, Lauble H, Stout CD, Olson AJ. Automated docking in crystallography: analysis of the substrates of aconitase. Proteins. 1993; 17(1):1–10. [PubMed: 8234239]

[59]. Hart TN, Read RJ. A multiple-start Monte Carlo docking method. Proteins. 1992; 13(3):206–222. [PubMed: 1603810]

[60]. Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. Proteins. 1990; 8(3):195–202. [PubMed: 2281083]

[61]. Abagyan R, Totrov M, Kuznetsov D. ICM-A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. J. Comput. Chem. 1994; 15:488–506.

[62]. McMartin C, Bohacek RS. QXP: powerful, rapid computer algorithms for structure-based drug design. J Comput Aided Mol Des. 1997; 11(4):333–344. [PubMed: 9334900]

[63]. Accelrys Inc., San Diego, CA, USA.

[64]. Oshiro CM, Kuntz ID, Dixon JS. Flexible ligand docking using a genetic algorithm. J Comput Aided Mol Des. 1995; 9(2):113–130. [PubMed: 7608743]

[65]. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. Proteins. 2003; 52(4):609–623. [PubMed: 12910460]

[66]. Clark KP. Ajay, Flexible ligand docking without parameter adjustment across four ligand-receptor complexes. J Comput Chem. 1995; 16:1210–1226.

[67]. Taylor JS, Burnett RM. DARWIN: a program for docking flexible molecules. Proteins. 2000; 41(2):173–191. [PubMed: 10966571]

[68]. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J. Am. Chem. Soc. 1995; 117:5179–5197.

[69]. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S Jr. Weiner P. New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. J. Am. Chem. Soc. 1984; 106:765–784.

[70]. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 1983; 4:187–217.

[71]. Kollman PA. Free energy calculations: Applications to chemical and biochemical phenomena. Chem. Rev. 1993; 93:2395–2417.

[72]. Aqvist J, Luzhkov VB, Brandsdal BO. Ligand binding affinities from MD simulations. Acc Chem Res. 2002; 35(6):358–365. [PubMed: 12069620]

[73]. Carlson HA, Jorgensen WL. An extended linear response method for determining free energies of hydration. J Phys Chem. 1995; 99:10667–10673.

[74]. Shoichet BK, Stroud RM, Santi DV, Kuntz ID, Perry KM. Structure-based discovery of inhibitors of thymidylate synthase. Science. 1993; 259(5100):1445–1450. [PubMed: 8451640]

[75]. Michel J, Verdonk ML, Essex JW. Protein-ligand binding affinity predictions by implicit solvent simulations: a tool for lead optimization? J Med Chem. 2006; 49(25):7427–7439. [PubMed: 17149872]

[76]. Briggs JM, Marrone TJ, McCammon JA. Computational Science New Horizons and Relevance to Pharmaceutical Design. Trends Cardiovasc. Med. 1996; 6:198–206. [PubMed: 21232297]

[77]. Bohm HJ. 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. J Comput Aided Mol Des. 1998; 12(4):309–323. [PubMed: 9777490]

[78]. Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, Fogel LJ, Freer ST. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. Chem Biol. 1995; 2(5):317–324. [PubMed: 9383433]

[79]. Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Arthurs S, Colson AB, Freer ST, Larson V, Luty BA, Marrone T, Rose PW. Deciphering common failures in molecular docking of ligand-protein complexes. J Comput Aided Mol Des. 2000; 14(8):731–751. [PubMed: 11131967]

[80]. Jain AN. Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities. J Comput Aided Mol Des. 1996; 10(5):427–440. [PubMed: 8951652]

[81]. Head RD, Smythe ML, Oprea TI, Waller CL, Green SM, Marshall GR. VALIDATE: A New Method for the Receptor-Based Prediction of Binding Affinities of Novel Ligands. J. Am. Chem. Soc. 1996; 118:3959–3969.

[82]. Gehlhaar DK, Moerder KE, Zichi D, Sherman CJ, Ogden RC, Freer ST. De novo design of enzyme inhibitors by Monte Carlo ligand generation. J Med Chem. 1995; 38(3):466–472. [PubMed: 7853340]

[83]. 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. J Comput Aided Mol Des. 1997; 11(5):425–445. [PubMed: 9385547]

[84]. Muegge I, Martin YC. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. J Med Chem. 1999; 42(5):791–804. [PubMed: 10072678]

[85]. Mitchell JBO, Laskowski RA, Alex A, Thornton JM. Bleep-potential of mean force describing protein-ligand interactions: I. generating potential. J. Comput. Chem. 1999; 20(11):1165–1176.

[86]. Ishchenko AV, Shakhnovich EI. SMall Molecule Growth 2001 (SMoG2001): an improved knowledge-based scoring function for protein-ligand interactions. J Med Chem. 2002; 45(13): 2770–2780. [PubMed: 12061879]

[87]. Feher M, Deretey E, Roy S. BHB: a simple knowledge-based scoring function to improve the efficiency of database screening. J Chem Inf Comput Sci. 2003; 43(4):1316–1327. [PubMed: 12870925]

[88]. Verkhivker G, Appelt K, Freer ST, Villafranca JE. Empirical free energy calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. Protein Eng. 1995; 8(7):677–691. [PubMed: 8577696]

[89]. Wallqvist A, Jernigan RL, Covell DG. A preference-based free-energy parameterization of enzyme-inhibitor binding. Applications to HIV-1-protease inhibitor design. Protein Sci. 1995; 4(9):1881–1903. [PubMed: 8528086]

[90]. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol. 2000; 295(2):337–356. [PubMed: 10623530]

[91]. DeWitte RS, Shakhnovich EI. SMoG: de Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1 Methodology and Supporting Evidence. J. Am. Chem. Soc. 1996; 118:11733–11744.

[92]. 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. J Med Chem. 1999; 42(25):5100–5109. [PubMed: 10602695]

[93]. Feher M. Consensus scoring for protein-ligand interactions. Drug Discov Today. 2006; 11(9-10): 421–428. [PubMed: 16635804]

[94]. Clark RD, Strizhev A, Leonard JM, Blake JF, Matthew JB. Consensus scoring for ligand/protein interactions. J Mol Graph Model. 2002; 20(4):281–295. [PubMed: 11858637]

[95]. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidateâ^'DNA Helices. Journal of the American Chemical Society. 1998; 120(37):9401–9409.

[96]. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE 3rd. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res. 2000; 33(12):889–897. [PubMed: 11123888]

[97]. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. J. Am. Chem. Soc. 1990; 112(16):6127–6129.

[98]. Guimaraes CR, Mathiowetz AM. Addressing limitations with the MM-GB/SA scoring procedure using the WaterMap method and free energy perturbation calculations. J Chem Inf Model. 50(4): 547–559. [PubMed: 20235592]

[99]. Singh N, Warshel A. Absolute binding free energy calculations: on the accuracy of computational scoring of protein-ligand interactions. Proteins. 2010; 78(7):1705–1723. [PubMed: 20186976]

[100]. Gabb HA, Jackson RM, Sternberg MJ. Modelling protein docking using shape complementarity, electrostatics and biochemical information. J Mol Biol. 1997; 272(1):106–120. [PubMed: 9299341]

[101]. Bron C, Kerbosch J. Algorithm 457: Finding All Cliques of an Undirected Graph. Communications of the ACM. 1973; 16(9):575–576.

[102]. Meng EC, Shoichet BK, Kuntz ID. Automated docking with grid-based energy evaluation. J. Comput. Chem. 1992; 13:505–524.

[103]. Zou XQ, Sun Y, Kuntz ID. Inclusion of Solvation in Ligand Binding Free Energy Calculations Using the Generalized-Born Model. J. Am. Chem. Soc. 1999; 121:8033–8043.

[104]. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2009

[105]. Bohm HJ. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J Comput Aided Mol Des. 1994; 8(3):243–256. [PubMed: 7964925]

[106]. Teague SJ. Implications of protein flexibility for drug discovery. Nat Rev Drug Discov. 2003; 2(7):527–541. [PubMed: 12838268]

[107]. Gschwend DA, Good AC, Kuntz ID. Molecular docking towards drug discovery. J Mol Recognit. 1996; 9(2):175–186. [PubMed: 8877811]

[108]. Totrov, M.; Abagyan, R. Protein-ligand docking as an energy optimization problem. In: Raffa, RB., editor. Drug-receptor thermodynamics: Introduction and experimental applications. John Wiley & Sons; New York: 2001. p. 603-624.

[109]. Leach AR. Ligand docking to proteins with discrete side-chain flexibility. J Mol Biol. 1994; 235(1):345–356. [PubMed: 8289255]

[110]. Desmet J, De Maeyer M, Hazes B, Lasters I. The dead-end elimination theorem and its use in protein sidechain positioning. Nature. 1992; 356:539–542. [PubMed: 21488406]

[111]. Abagyan R, Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J Mol Biol. 1994; 235(3):983–1002. [PubMed: 8289329]

[112]. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009; 30(16):2785–2791. [PubMed: 19399780]

[113]. Knegtel RM, Kuntz ID, Oshiro CM. Molecular docking to ensembles of protein structures. J Mol Biol. 1997; 266(2):424–440. [PubMed: 9047373]

[114]. Carlson HA, Masukawa KM, Rubins K, Bushman FD, Jorgensen WL, Lins RD, Briggs JM, McCammon JA. Developing a dynamic pharmacophore model for HIV-1 integrase. J Med Chem. 2000; 43(11):2100–2114. [PubMed: 10841789]

[115]. Cavasotto CN, Abagyan RA. Protein flexibility in ligand docking and virtual screening to protein kinases. J Mol Biol. 2004; 337(1):209–225. [PubMed: 15001363]

[116]. Derreumaux P, Schlick T. The loop opening/closing motion of the enzyme triosephosphate isomerase. Biophys J. 1998; 74(1):72–81. [PubMed: 9449311]

[117]. Zeng L, Zhou MM. Bromodomain: an acetyl-lysine binding domain. FEBS Lett. 2002; 513(1): 124–128. [PubMed: 11911891]

[118]. Venkitakrishnan RP, Zaborowski E, McElheny D, Benkovic SJ, Dyson HJ, Wright PE. Conformational changes in the active site loops of dihydrofolate reductase during the catalytic cycle. Biochemistry. 2004; 43(51):16046–16055. [PubMed: 15609999]

[119]. Go N, Scheraga HA. Ring Closure and Local Conformational Deformations of Chain Molecules. Macromolecules. 1970; 3(2):178–187.

[120]. Dodd LR, Boone TD, Theodorou DN. A concerted rotation algorithm for atomistic Monte Carlo simulation of polymer melts and glasses. Mol. Phys. 1993; 78:961–996.

[121]. Hoffmann D, Knapp W. Polypeptide folding with off-lattice Monte Carlo dynamics: the method. Eur. Biophys. J. 1996; 111:387–404.

[122]. Wu MG, Deem MW. Analytical rebridging Monte Carlo: application to cis/trans isomerization in proline-containing, cyclic peptides. J. Chem. Phys. 1999; 14:6625–6632.

[123]. Dinner AR. Local deformations of polymers with nonplannar rigid main chain internal coordinates. J. Comp. Chem. 2000; 21:1132–1144.

[124]. Mezei M. Efficient Monte Carlo sampling for long molecular chains using local moves, tested on a solvated lipid bilayer. J. Chem. Phys. 2003; 118:3874–3879.

[125]. Cui M, Mezei M, Osman R. Prediction of protein loop structures using a local move Monte Carlo approach and a grid-based force field. Protein Eng Des Sel. 2008; 21(12):729–735. [PubMed: 18957407]

[126]. Kubinyi, H. Computer Applications in Pharmaceutical Research and Development. John Wiley; New York: 2006.

[127]. Kroemer RT. Structure-Based Drug Design: Docking and Scoring. Current Protein and Peptide Science. 2007; 8:312–328. [PubMed: 17696866]

[128]. Venhorst J, ter Laak AM, Commandeur JN, Funae Y, Hiroi T, Vermeulen NP. Homology modeling of rat and human cytochrome P450 2D (CYP2D) isoforms and computational rationalization of experimental ligand-binding specificities. J Med Chem. 2003; 46(1):74–86. [PubMed: 12502361]

[129]. Williams PA, Cosme J, Ward A, Angove HC, Matak Vinkovic D, Jhoti H. Crystal structure of human cytochrome P450 2C9 with bound warfarin. Nature. 2003; 424(6947):464–468. [PubMed: 12861225]

[130]. Meng XY, Zheng QC, Zhang HX. A comparative analysis of binding sites between mouse CYP2C38 and CYP2C39 based on homology modeling, molecular dynamics simulation and docking studies. Biochim Biophys Acta. 2009; 1794(7):1066–1072. [PubMed: 19358898]

[131]. Boehm HJ, Boehringer M, Bur D, Gmuender H, Huber W, Klaus W, Kostrewa D, Kuehne H, Luebbers T, Meunier-Keller N, Mueller F. Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. J Med Chem. 2000; 43(14):2664–2674. [PubMed: 10893304]

[132]. Kirton SB, Murray CW, Verdonk ML, Taylor RD. Prediction of binding modes for ligands in the cytochromes P450 and other heme-containing proteins. Proteins. 2005; 58(4):836–844. [PubMed: 15651036]

[133]. Doman TN, McGovern SL, Witherbee BJ, Kasten TP, Kurumbail R, Stallings WC, Connolly DT, Shoichet BK. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. J Med Chem. 2002; 45(11):2213–2221. [PubMed: 12014959]

[134]. Shoichet BK, Leach AR, Kuntz ID. Ligand solvation in molecular docking. Proteins. 1999; 34(1):4–16. [PubMed: 10336382]

[135]. Lorber DM, Shoichet BK. Flexible ligand docking using conformational ensembles. Protein Sci. 1998; 7(4):938–950. [PubMed: 9568900]

[136]. Freymann DM, Wenck MA, Engel JC, Feng J, Focia PJ, Eakin AE, Craig SP. Efficient identification of inhibitors targeting the closed active site conformation of the HPRT from Trypanosoma cruzi. Chem Biol. 2000; 7(12):957–968. [PubMed: 11137818]

[137]. Su AI, Lorber DM, Weston GS, Baase WA, Matthews BW, Shoichet BK. Docking molecules by families to increase the diversity of hits in database screens: computational strategy and experimental evaluation. Proteins. 2001; 42(2):279–293. [PubMed: 11119652]

[138]. Gschwend DA, Kuntz ID. Orientational sampling and rigid-body minimization in molecular docking revisited: on-the-fly optimization and degeneracy removal. J Comput Aided Mol Des. 1996; 10(2):123–132. [PubMed: 8741016]

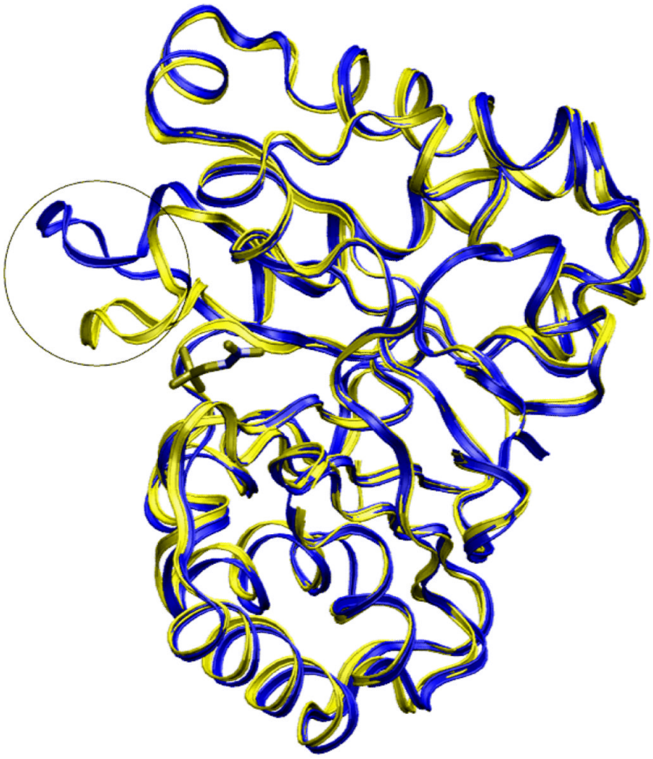

Figure 1.

Superimposed apo- (in yellow) and holo- (in blue) crystal structures of triosephosphate isomerase. PDB code 1YPI and 2YPI, respectively [116]. The 11 residue-loop composed of binding site is the only region that has large motion upon ligand binding (in circle).

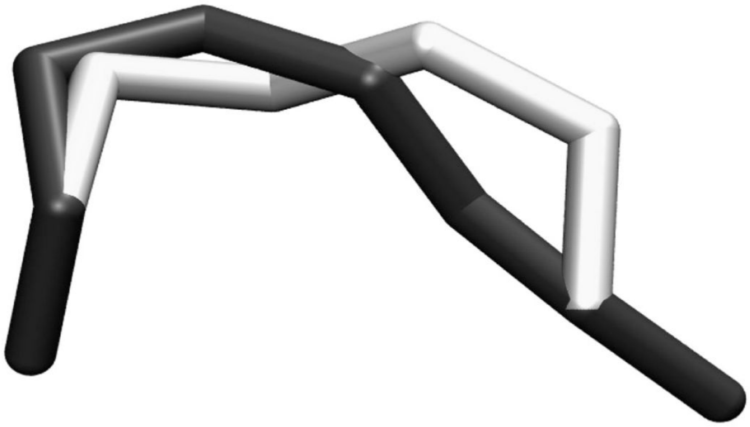

Figure 2.

Local move of a lipid tail. Six subsequent torsions change while keeping the rest of the chain to remain in its original position.

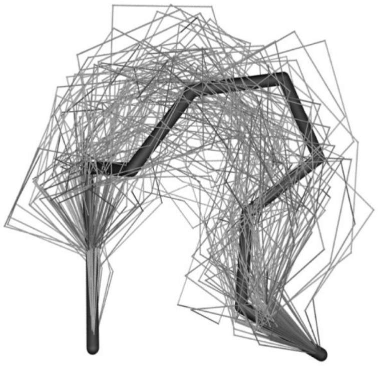

Figure 3.

Loop Structure of 2act (198-205) produced by the local move MC method at 5000K and followed by clustering to generate 100 representative conformations. Black stick represents the crystal loop structure, and gray wires represent the 100 representative loop conformations.

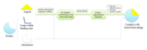
Ligand
Loops within binding sites
Protein
Complex with lowest total energy
cluster analysis
Cluster representative structures

Figure 4.

Flowchart of local move Monte Carlo (LMMC) loop sampling approach for protein-ligand docking. Abbreviate: MC, Monte Carlo. ESC, Exponential Coolong Scheme. LCS, Linear Cooling Scheme.

Table 1

Some sampling algorithms discussed in this paper.

Algorithms	Characteristic	Reference
Matching algorithms	Geometry-based, suitable to VS and database enrichment for its high speed	[43-45]
Incremental construction	Fragment-based and docking incrementally	[30, 49, 50]
MCSS	fragment-based methods for the *de novo* design	[55, 56]
LUDI	fragment-based methods for the *de novo* design	[57]
Monte Carlo	Stochastic search	[58, 59]
Genetic algorithms	Stochastic search	[31, 32, 64]
Molecular dynamics	For further refinement after docking	[68-70]

Table 2

Examples of scoring function formulae

Scoring function formulae
$V = W_{vdw}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + W_{hbond}\sum_{i,j}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + W_{elec}\sum_{i,j}\frac{q_i q_j}{\epsilon(r_{ij})r_{ij}} + W_{sol}\sum_{i,j}(S_i V_j + S_j V_i)e^{\left(-r_{ij}^2/2\sigma^2\right)}$ Extended force-field-based scoring function from AutoDock. For two atoms i, j, the pair-wise atomic energy is evaluated by the sum of van der Waals, hydrogen bond, coulomb energy and desolvation. *W* are weighted factors for calibrate the empirical free energy.
$\Delta G = \Delta G_0 + \Delta G_{rot} \times N_{rot} + \Delta G_{hb}\sum_{neutral\ H-bond} f(\Delta R,\ \Delta a) + \Delta G_{io}\sum_{ion\ init.} f(\Delta R,\ \Delta a) + \Delta G_{aro}\sum_{aro\ int.} f(\Delta R,\ \Delta a) + \Delta G_{lipo}{}_{lipo}$ Empirical scoring function from FlexX. ΔG is the estimated free energy of binding; ΔG_0 is the regression constant; ΔG_{rot} , ΔG_{hb} , ΔG_{io} , ΔG_{aro} and ΔG_{lipo} are regression coefficients for each corresponding free energy term; $f(\Delta R,\Delta \alpha)$ is scaling function penalizing deviations from the ideal geometry; N_{rot} is the number of free rotate bonds that are immobilized in the complex.
$PM\_score = \sum_{\substack{kI \\ r<r_{cut-off}^{ij}}} A_{ij}(r) \qquad A_{ij}(r) = -k_B T \ln\left[f_{Vol\_corr}^{j}(r)\frac{\rho_{seg}^{ij}(r)}{\rho_{bulk}^{ij}}\right]$ Knowledge-based scoring functions PMF. k_B is the Boltzmann constant; T is the absolute temperature; r is the atom pair distance. $f_{Vol\_corr}^{j}(r)$ is the ligand volume correction factor; $\frac{\rho_{seg}^{ij}(r)}{\rho_{bulk}^{ij}}$ designates the radial distribution function of a protein atom of type i and a ligand atom of type j.

Table 3

Some basic methods for including receptor flexibility.

Method	Description	Advantage	Disadvantage	Program
Soft potential	Change vdW to allow for overlap between receptor and ligand atoms	Computational efficiency. Easy to implement and use combined with other methods.	Inadequate flexibility. Describe flexibility in an implicit, rude and non- quantitative way.	GOLD [65] AutoDock [31]
Rotamer library	Search side chain library to obtain possible conformations	Relative computational efficiency. Avoid minimization barriers.	Strong dependence on the database used. No backbone flexibility.	ICM [61]
Receptor side chain flexibility	Sample both side chain and ligand conformations simultaneously using GA	Relative computational efficiency. Model the effect that ligand make on binding site residues.	Only selected side chains are involved. No backbone flexibility.	AutoDock 4 [112]
Ensemble of protein conformations	Docking ligand to a series of receptor structures which represent different conformational states.	Include full and explicit flexibility.	Expensive computational cost. Limited by protein conformations used in sampling.	DOCK [113] FlexE [38]