Automated docking for novel drug discovery

Article *in* Expert Opinion on Drug Discovery · May 2013

Impact Factor: 3.54 · DOI: 10.1517/17460441.2013.794780 · Source: PubMed

3 authors:

Martiniano Bello
Instituto Politécnico Nacional
27 PUBLICATIONS **162** CITATIONS
SEE PROFILE

Marlet Martínez-Archundia
Instituto Politécnico Nacional
11 PUBLICATIONS **43** CITATIONS
SEE PROFILE

José Correa-Basurto
Instituto Politécnico Nacional
136 PUBLICATIONS **613** CITATIONS
SEE PROFILE

EXPERT OPINION

1. Introduction
2. Algorithms for docking studies
3. Expert opinion

Automated docking for novel drug discovery

Martiniano Bello, Marlet Martínez-Archundia & José Correa-Basurto†
Laboratorio de Modelado Molecular y Bioinformática de la Escuela Superior de Medicina, Instituto Politécnico Nacional, México, CP, USA

Introduction: The volume of three-dimensional structural information of macromolecules and the number of computational tools to predict binding modes and affinities of molecular complexes are increasing daily. Molecular docking is a rational structural approach employed to predict thermodynamic parameters based on molecular recognition between two or more molecules. In addition, docking studies have become very important for therapeutic applications in modern structure-based drug design because this computational tool uses few economic resources. However, they omit many biological conditions that critically influence small and macromolecular structural motions. To mimic physiological conditions, it is necessary to consider other environmental factors, such as the presence of water molecules and the flexibility of ligands and side chain residues of proteins. Furthermore, molecular dynamics simulations have been coupled with docking procedures to expand the boundaries and obtain more reliable information.

Areas covered: In this article, we review current advances in protein-small molecule docking and possible future directions.

Expert opinion: Docking studies include many conformations to predict binding free energies (scoring functions) and to search (scoring sampling) for the most representative binding conformations. Therefore, several biological properties, from side chain residues to complete protein motions, have been included in docking studies to improve theoretical predictions.

Keywords: docking, molecular dynamics simulations, protein target, small ligands

Expert Opin. Drug Discov. [Early Online]

1. Introduction

Since the isolation of morphine from opium in 1805 by the German pharmacist Friedrich Wilhelm Sertürner, pharmaceutical companies have been aware of the importance of finding other naturally occurring substances with potential medicinal uses [1].

In the nineteenth century, the physiologists John Newport Langley and Paul Ehrlich introduced the concept of a receptor [2]. Paul Ehrlich conceptualized the idea that different chemoreceptors can be found in parasites, microorganisms or cancer cells based on his observations of their response to certain dyes [3]. In 1905, Langley and his collaborators proposed that binding a substrate to a receptor could activate or inhibit the receptor [4]. These theories served as the basis for the study of drug action on cells in the 1930s by Joseph Clark [5].

Experimental techniques, such as X-ray crystallography and nuclear magnetic resonance (NMR), are currently the most utilized methods to obtain information concerning the three-dimensional (3D) structure of water-soluble proteins and membrane proteins. However, it is very difficult to study membrane proteins because of their intrinsic flexibility and limited water solubility [6]. For this reason, other crystallization conditions have been explored for membrane proteins, such as co-crystallization of membrane proteins with their respective ligands [7] or the

10.1517/17460441.2013.794780 © 2013 Informa UK, Ltd. ISSN 1746-0441, e-ISSN 1746-045X

Article highlights.

- Docking studies are dependent on the search and scoring functions.
- Docking studies include side chain motions.
- Docking studies depict pharmacophore moieties of the tested ligands.
- Docking studies are more efficient at identifying a potential drug than traditional procedures.
- MD simulations coupled with docking could yield more reliable data than traditional procedures.
- Incorporating protein and ligand flexibility yield better results than rigid systems.

This box summarizes key points contained in the article.

design of chimeric proteins, such as the human beta-2 adrenergic receptor [8]. In the field of drug discovery, it is essential to study co-crystallized complexes of target proteins and their ligands [9]. This experimental data provides detailed information at the atomic level about receptor structure and particularly binding sites for use in drug design. In addition, these experimental data help theoretical researchers design algorithms that can be validated with respect to large numbers of ligand–protein complexes and subsequently used to perform molecular docking studies [10].

In response to current demands for the discovery and design of new molecular structures with the potential for use as drugs, computational methods (*in silico*) are being applied to discover new receptor molecules by employing algorithms for *de novo* ligand design and ligand docking as well as for scoring protein–ligand binding arrangements and energies [11]. *In silico* tools are very useful for the identification of binding sites on proteins [12] and the prediction of ligand–receptor interactions that determine ligand–protein affinity [13].

Because docking studies are very useful for the prediction of protein–ligand interactions, they provide an opportunity to explore recognition properties and identify potential pharmacophores [14]. Currently, most of the computational docking programs, such as Autodock 3.0.5 and Autodock Vina [15], consider the protein as a rigid body and the ligand as a flexible molecule, which reduces the computational cost. However, these calculations omit conformational changes that occur in protein molecules due to ligand binding.

Due to the importance of molecular motion and the flexibility of targets and/or ligands, other types of docking programs have been designed. For example, some docking algorithms consider the flexibility of certain residues in the protein side chains and ligands to make accurate predictions of protein–ligand associations [16,17]. However, these docking programs do not consider backbone protein motion, which are very important in the process of ligand recognition [18]. To incorporate protein flexibility in ligand–protein coupling studies under a docking procedure, many research groups employ molecular dynamics (MD) simulations [19] or introduced protocols for fits [20,21], which can provide more reliable predictions than traditional protocols. However, when several ligands are examined with virtual screening protocols, the predictions may require longer times and be computationally more expensive.

2. Algorithms for docking studies

2.1 Search algorithms

Docking studies can be regarded as combinations of two components, a search strategy and a scoring function. The search algorithm is defined by a set of rules and parameters used to predict ligand-receptor conformations. A search algorithm should generate an optimum number of putative ligand binding orientations/conformations at the binding site of a protein, which include the experimentally determined binding mode [22]. After these configurations are generated, they are evaluated with scoring functions to distinguish the experimental binding mode from all of the binding modes identified by the search algorithm in the docking program. A rigorous search algorithm would explore all of the possible binding modes between the two molecules, but this would be impractical due to the size of the search space [22].

Docking applications can be classified into at least two groups based on ligand flexibility and/or receptor docking algorithms: rigid-body and flexible docking. Rigid-body docking methods consider neither ligand nor receptor flexibility, which limits the specificity and accuracy of their results. Despite the lack of molecular flexibility, it is possible for rigid-body docking methods to predict the correct ligand binding sites for several different proteins relative to their corresponding co-crystallized complexes [23-28]. Therefore, rigid-body docking methods are used as the fastest method for performing an initial screen of a small molecule database [29]. In contrast, flexible docking methods allow the choice of several conformations of the ligand or the receptor or for both molecules simultaneously at greater computational expense.

In general, docking procedures use four types of search algorithms: shape matching (SM) [30], systematic search, stochastic algorithm and MD simulations. SM methods are among the simplest algorithms used in the early stages of the docking processes. SM methods place the ligand into the protein binding site using a criterion that the molecular surfaces must complement each other. The degrees of freedom of the ligand enable the ligand to adopt different conformer orientations in the macromolecule binding site. Therefore, the main goal of SM algorithms is to place the ligand inside the binding site with good shape complementarities. In fact, several docking applications, such as ZDOCK [31], FRED [32], FLOG [33], Surflex [34], LigandFit [35] and EUDOC [36], adopt the basic concept of SM. For example, EUDOC performs a systematic search of rigid body rotations and translations of a rigid ligand within a rigid active site [36]. This application is based on the earlier SYSDOC program [37] that uses a fast affine transformation to perform the systematic search [36] and utilizes the AMBER force field [38] to calculate the

intermolecular energy with a distance-dependent dielectric. This method was applied to virtual screening of farnesyltransferase inhibitors. Based on this study, 21 hits were obtained, and 4 of these hits were selected as good inhibitors [22]. However, when using the SM method with flexible-ligand docking, it is necessary to dock an ensemble of pre-generated ligand conformations into the protein [39] and then re-rank the docked arrangements with respect to their energy scores [40]. Recently, a 'progressive distributed docking method' has been reported that iteratively combined a shape-matching method with a multiple receptor conformation docking method [41]. A validation of this method with a selection of PPARγ- and PI3Kγ-oriented compounds revealed that the combination of these two methods provided a better selection of the target-oriented virtual hits.

Systematic search algorithms are commonly used in flexible-ligand docking procedures. With this type of algorithm, it is possible to generate every possible ligand binding conformation by exploring all degrees of freedom for the ligand. Systematic searches can be divided into three types: exhaustive, conformational ensemble and fragment-based. Exhaustive searches are the most straightforward method, and in these searches, flexible-ligand docking is performed by rotating all of the possible bonds of the ligand at a given interval. However, for some cases, the large number of rotatable bonds in a ligand would result in a huge number of conformations, and geometrical or chemical constraints are used to filter the ligand conformations. Glide [42,43] and FRED [32] are two examples of exhaustive search methods. FRED utilizes a Gaussian function [32] to generate a smooth and clearly searchable energy landscape. In addition, this method allows a broad freedom for errors in the protein atom positions. Therefore, although the use of a pre-generated database of multiple ligand conformers is an appropriate technique when a soft and error-tolerant approach is used, the method has limited precision in accurately ranking ligands resulting in further scoring [44]. In fact, comparative research between the FRED and Glide scoring schemes using the binding sites of different proteins [45] indicated that hard functions, such as those used in Glide, performed better than the softer scoring function in FRED. Nevertheless, when hydrophobic effects outweighed electrostatic interactions, FRED was found to produce better results for hydrophobic binding sites indicating that these methods react differently to specific features in the binding sites.

Conformational ensemble methods [39] simulate ligand flexibility by using an ensemble of ligand conformations previously generated through a rigid-docking program. Afterward, the ligand binding conformations obtained from different docking runs are collected and ranked with respect to their binding free energies. FLOG [33], DOCK [46], PhDOCK [47], MS-DOCK [48], MDock [49,50] and Q-Dock [51] are examples of conformational ensemble methods for docking. For example, FLOG [33] generates database conformations based on distance geometries. Once adequate conformations have been performed, the algorithm explores them in a manner similar to DOCK [46].

In fragment-based methods, the ligand is divided into separate portions or fragments. These fragments are individually docked into the protein-binding site, and then linked covalently. Some well-known docking programs that use fragment-based methods are DOCK [46] and FlexX [52]. In FlexX, the first step is the selection phase of the base fragment for the ligand from which possible conformations are made based on the MIMUMBA torsion angle database [53]. Although this selection was manually performed in early implementations, this process has now been automated [54]. All of the fragments identified for a particular ligand serve as starting points for the docking. After placement of a base fragment, the complete ligand is constructed by adding the remaining components. At each step, the interactions are calculated and the best solution is selected according to the docking score. The docking score uses the model of molecular interactions developed by Klebe [55] and Böhm [56].

Originally, this algorithm was validated on 19 complexes, from which 10 of the docked complexes with the best score reproduced the experimental binding mode with maximum RMSD values of 1.04 Å, although these were not necessarily the lowest free energy values. More recent extensions to FlexX include the placement of explicit waters into the binding site through the docking procedure [57]. This process was validated using 200 complexes, and based on this study, some improvements were found for some targets, such as HIV protease.

The docking program DOCK [46,58] basically performs a search for geometrically allowed ligand-binding poses by modeling the ligand and receptor cavity as spheres, matching the sphere groups, placing the ligand and scoring each position using the AMBER force field [38]. Some extensions have been applied to this protocol in which the bipartite graphs consisting of the protein and ligand interactions are combined into a single graph, giving rise to a pair of molecular interactions from which group detection is performed using the methodology developed by Bron and Kerbosch [59]. In addition, this technique was evaluated [60,61] and classified as one of the most efficient methods to locate a cluster that has the maximal pairs of interactions between matching sites. DOCK 4.0 incorporates ligand flexibility using a modified scoring function [62]. In this version, the ligand fragment is chosen and placed on the receptor. Then, a rigid body minimization is performed, and the conformations of the remaining parts of the ligand are searched by a limited backtrack method and minimized. This protocol was evaluated using 10 structures from which 7 docked complexes reproduced the crystallographic complex with a maximum RMSD value of 1.03 Å with the remaining 3 complexes within 1.88 Å [22]. Some of the new features added in a recent version (i.e., DOCK 6.0) include minimization that uses a conjugate gradient method, MD simulation capabilities, ligand and receptor desolvation, ligand conformational

entropy corrections, receptor flexibility and an efficient MPI-based parallel implementation [63].

Although DOCK and FlexX are fragment-based methods, they produce quite different results. Whereas DOCK performs well with apolar binding sites, FlexX shows the totally opposite behavior. In fact, it has a bit lower hit rate than DOCK, but provides better estimates of RMSD for compounds with correctly predicted binding modes. Furthermore, there is a version of FlexX named FlexE with flexible receptors that has produced better results with significantly lower computational times [64].

In stochastic methods, the ligand binding conformations are sampled by making random changes to the ligand in both the conformational space and degrees of freedom [40] and deciding whether to accept or reject the change according to a probabilistic criterion. Stochastic methods can be divided into four types: Monte Carlo (MC) simulations, genetic algorithms (GA), Tabu search methods (TS) and particle swarm optimization (PSO).

MC simulations were first used as minimization procedures in some MD simulations, such as GROMACS [65], and then they were adopted for use by certain flexible docking algorithms, such as MCDOCK [66], ICM [17] and AutoDock [67,68]. MC simulations involve the generation of random moves in the system and then accept or reject these moves based on a Boltzmann probability function [22]. Because of the combination of atomistic potential energy models with stochastic search techniques, MC simulations are among the most powerful methods available for analyzing different thermodynamic conditions for both structure optimization and prediction. For flexible docking, the MC procedure places the ligand inside the receptor binding site by exploring many random positions and rotations, which decreases the likelihood of being trapped in local minima [22]. Next, each random structure is minimized using a force field. MC simulations are not appropriate for assessing time-dependent processes. At the present time, DockVision [69], ICM [70], QXP [71], Prodock [72] and MCDOCK are examples of programs that use MC methods [66].

Prodock [72] uses a MC minimization approach to dock flexible ligands into a flexible binding site using two force fields (AMBER [38] and ECEPP/3 [73]) and a solvation model based on solvent exposed volume. However, this method differs from a standard MC procedure in that after each random motion a local gradient-based minimization is performed. Then, the resulting structure would be acceptable based on the Metropolis acceptance criteria. In this program, the sampling is facilitated during the docking process by scaling the magnitudes of the potential energy terms, thus, allowing the reduction of barriers that restrict sampling. In its first version, MCDOCK [66] used a MC minimization approach with simulated annealing using a scoring function based on the CHARMM force field [74]. MCDOCK applied a multiple step strategy to dock a flexible ligand into a rigid receptor. Next, the overlaps between the ligand and protein atoms are reduced by applying random motions. Then, a MC simulation that incorporates an adjustable temperature is performed. This method have been tested using a set of 19 complexes taken from a set obtained with the FlexX program, and the RMSD between the theoretical and experimental binding modes oscillated from 0.25 Å to 1.84 Å [75]. In contrast to Prodock and MCDOCK, DockVision applies a rigid ligand and receptor. First, this docking algorithm generates a random ligand orientation, and an MC method is applied to the system, except for the energy function, which is replaced by a geometric score for atomic overlap. Next, MC-simulated annealing is performed using a simple potential energy function, and the same procedure is repeated for a large number of random ligand orientations. Finally, the generated ligand orientations are clustered based on an RMSD score. This method has been tested using two inhibitor–protein complexes, and the binding geometry was properly predicted in both cases [69].

GA belongs to the class of evolutionary programming algorithms that solve docking problems by trying to find the exact conformation of the global energy minimum, or similar conformations, based on approaches that adapt the principles of biological competition and population dynamics. Basically, the concept of GA is the evolution of a population of possible solutions through genetic operators (mutations, crossovers and migrations) to attain a final population [22]. The process of applying GA starts with encoding the variables. In this case, the degrees of freedom are encoded into genes or binary strings [22]. A random initial population of solutions is created, and a genetic operator is applied to generate a new population that is scored and ranked using the survival of the fittest. Therefore, the probability that a conformation advances to the next iteration round depends on its score. However, in contrast to MC and MD, which require only a single initial structure, GA requires the generation of an initial population of structures. Docking programs that use GA include GOLD [76], AutoDock [77], DIVALI [78], DARWIN [79], MolDock [80], PSI-DOCK [81], FLIPDock [82], Lead finder [83] and EADock [84].

GOLD was originally developed by Jones *et al.* [76]. The program has been commercially released by the Cambridge Crystallographic Data Center. GOLD 3.2 uses a GA to explore the rotational flexibility for selected receptor hydrogen bonds along with total ligand flexibility [85]. The ligand–receptor hydrogen bonds are subsequently matched with a least squares fitting protocol to maximize the number of intermolecular hydrogen bonds between the accepting and donating hydrogen bonding groups. This feature is unique to GOLD. In addition, the GA optimizes GOLD for flexible dihedrals, ligand ring geometries and for dihedrals of protein OH^- and NH^{3+} groups [85]. Given the 3D structures of both protein and ligand, an initial population of ligand-binding poses is randomly generated. Then, fitness is assigned to each individual of the population based on its predicted binding affinity. Three scoring functions, referred to as GoldScore, ChemScore and ASP, are implemented in GOLD

3.2 to perform this task [86]. Kellenberger *et al.* [87] tested GOLD on 100 protein–ligand complexes and found that it successfully docked 80% of the ligands with an RMSD value of 2.0 Å. Perola *et al.* [85] performed a virtual screening of HIV-1 protease where GOLD predicts the binding of 60% of the ligands when investigating the top 10% of the ranked complexes.

AutoDock 4.0 uses a MC method and simulated annealing in combination with GA for constructing the possible conformations. GA is used for global optimization. The search method in combination with free energy values obtained using AMBER force field [38] is utilized to evaluate the binding positions with several scoring functions based on the free energy. Autodock has two software programs (AutoDock and AutoGrid). AutoDock performs the docking of the ligand to a set of grids depicting the target protein, while AutoGrid pre-calculates these grids. AutoDock calculates the atomic affinity using grids whereas its graphical user interface, named 'AutoDockTools (ADT)', depicts the ligand–protein interactions, which supports the analyses of the docking results. In contrast with some popular commercial software packages for conducting molecular docking simulations, such as GOLD, FlexX and ICM, AutoDock 4.0 has the advantage of being an open-source package. This program has been successfully utilized in a number of virtual screenings and in the development of the HIV integrase inhibitor raltegravir [88-90]. Recently, a parallelized version of 'AutoDock 4.2' using MPI and OpenMP to create mpAD4 has been released [91].

TS methods are iterative procedures designed to solve optimization problems [29]. In TS methods, new states are randomly produced from an initial random ligand conformation. These novel solutions are graded and rated in ascending order. The best solution is selected as the new current solution, and the procedure is repeated. The probability of approval will depend on the previously visited areas in the conformational space of the ligand [40] because a random configurational change will be rejected if the RMSD between the current ligand binding conformation and the previous one is less than a specific cut-off, otherwise, the configurational change will be accepted [40]. Some examples of docking programs that employ this algorithm are PRO_LEADS [92] and PSI-DOCK [81].

PRO_LEADS uses Chemscore [93] as a scoring function. PRO_LEADS has been tested for the flexible ligand docking of 50 ligand–protein complexes, from which a success rate of 86% of the solutions within an RMSD of 1.5 Å of the co-crystallized complexes was achieved [92]. This program has also been applied to 70 ligand–receptor complexes for which the experimental binding affinity and binding geometry are known, and 79% of the solutions were within 2.0 Å [94].

PSI-DOCK uses an improved score function capable of reproducing the absolute binding free energy of a training set of 200 protein complexes with a correlation coefficient of 0.788 and a standard error of 8.13 kJ/mol [81]. In addition, this program has been shown to be highly efficient at identifying the experimental binding pose. For a test dataset of 194 complexes, PSI-DOCK achieved a 67% success rate (RMSD < 2.0 Å) with one run. However, PSI-DOCK achieved a 74% success rate in 10 runs [81]. Although PSI-DOCK is similar to programs such as GOLD that have a high success rate for binding pose predictions, it also precisely estimated the experimental binding free energies and is extremely easy to use compared to other docking programs [81].

PSO methods share many similarities with evolutionary algorithms, such as GA. In PSO methods, the system starts with a population of random solutions and searches for an optimal solution in a search space. However, unlike GAs in which evolutionary operators, such as crossover and mutations, are used, the PSO methods use swarm intelligence [29]. In PSO methods, the motion of each ligand through the search space is influenced by its best known local position and guided toward its best value in the conformational space by information about the best positions of its neighbors. Some examples of docking programs that use PSO include SODOCK [95], PSO@Autodock [96] and Tribe-PSO [97].

SODOCK has been described as being more simple and efficient than GA-based methods [95]. The docking performance of SODOCK was evaluated using 37 complexes whose ligands had a number of torsions ranging from 0 to 19, and the smallest RMSD was obtained for 19 of 37. The average RMSD value (2.29 Å) of SODOCK was better than those of the other docking programs, such as FlexX [52], DOCK [46], GOLD [76] and LGA of AutoDock 3.05 [68], which were all above 3.0 Å. PSO@Autodock, which is a program constructed on Autodock 3.0 [68,77], has been reported as a highly efficient tool for performing flexible peptide–protein docking and virtual screening studies [96]. The performance of PSO@Autodock has been compared with the docking programs GOLD [76], DOCK [46], FlexX [52], Autodock 3.05 [68] and SODOCK [95] using a set of 37 complex structures with highly flexible ligands. For example, there was a ligand with 23 rotatable bonds that was successfully docked within as few as 100,000 computing steps (RMSD = 0.87 A), corresponding to only 10% of the computing time required by Autodock [96]. In addition, PSO@Autodock clearly outperforms the other docking programs, providing the smallest RMSD values for 12 in 37 protein–ligand complexes and an average RMSD value of 1.4 Å. This value was significantly lower than those obtained with the other docking programs, which were all above 2.0 A [96]. Therefore, PSO@autodock is a highly efficient docking program in terms of speed and quality for flexible peptide–protein docking and virtual screening studies.

Currently, MD simulations are one of the most versatile and popular computational approaches for studying biological ligand-target complexes. MD simulations involve the calculation of solutions to Newton's equations of motion [65], and there are many programs that can be used to perform

MD simulations, such as GROMACS [65], AMBER [38] and NAMD [98], among others. MD simulations are useful not only for exploring conformational changes coupled to complex stabilization [99] and intrinsic protein mobility coupled to metal coordination [100] but also for refining structures that are experimentally determined by X-ray and NMR studies [50,101].

Ideally, MD simulations could be used to simulate a ligand docked to a receptor. These methods have the advantage of taking into account all of the degrees of freedom of a protein, thus approaching physiological conditions while explicitly considering the solvent. In addition, acceptable binding free energies can be obtained using non-covalent terms of the association process along with the free energy perturbation (FEP) method [102] or the linear interaction energy (LIE) method [103]. Unfortunately, MD simulations are computationally expensive, and the computational power needed to simulate a complete ligand diffusion process without approximation is out of reach currently and in the near future. However, there are reports that these types of studies have been performed [104].

In contrast, MD simulations have the drawback of being often unable to cross high-energy barriers within the allowed simulation time and, therefore, might only accommodate ligands in local minima of the rugged hypersurface [105]. Thus, in some cases, the results from MD simulations depend on the starting conformation. Therefore, other approaches have been used with MD simulations to cross these barriers. In fact, MD simulations at various temperatures have been performed [106], whereas other authors have proposed methodologies to place the ligand in different binding poses [44]. In addition, another approach is to model full protein mobility, perform a MD simulation of the receptor without the ligand, and use an ensemble of conformations to represent the conformational diversity of the protein in water. Then, these conformations can be docked using a set of ligands employing traditional rigid-protein/flexible-ligand methods, as we have used in our research group [107]. It is important to mention that a MD simulation procedure not only allows for the sampling of many protein snapshots under a docking procedure, but it is also an important computational tool that can be used to refine new 3D structures constructed with homology modeling methods [108]. MD simulations allow a protein to achieve its optimal conformation energy value, which would represent its natural behavior. However, after refining a model, its structure needs to be validated by verifying several properties, such as Ramachandran maps, RMSD values versus template, as well as validating against its natural or synthetic ligands.

2.2 Scoring functions

Scoring functions are fast mathematical methods used to estimate the binding affinity between a ligand and its receptor for a complex predicted by a search algorithm [22]. Scoring functions are also considered the primary tool for lead optimization of virtual screening results because they make it possible to determine the highest-affinity ligand for a target. Then, another procedure is performed to discriminate between correct and incorrect arrangements. Therefore, the design of a trusted scoring function is critical. Free energy simulations have been used to predict the binding free energy [109,110]. However, this type of calculation has a high computational cost because it analyzes a large number of protein–ligand complexes. Therefore, scoring functions implemented in docking programs make various simplifications in the evaluation of the binding free energy. Nevertheless, scoring functions do not take into account a number of thermodynamic parameters that determine molecular recognition, for example, the entropic component.

Scoring functions can be classified into three different groups: force field based, [111], knowledge based [108,112] and empirically based [113].

Force field-based scoring functions have been developed based on physicochemical atomic interactions that include van der Waals (VDW) interactions, electrostatic interactions, bond stretching energies, bond angle bending energies, bond torsion energies and hydrogen bond energies. All of these parameters are commonly obtained from experimental and ab initio quantum mechanical calculations. Despite its simple physical meaning, a major challenge for force field scoring functions is how to treat the solvent in ligand binding [114]. Several force field scoring functions are based on different force field parameters. For example, DOCK and Autodock are based on the AMBER force field [38]. Recently, a torsional entropy for ligands in G-Score and the inclusion of explicit protein–ligand hydrogen bond terms, which might increase the specificity in molecular recognition, has been included in the force field-based scoring functions of some docking programs, such as Autodock [68] and Gold [44,76]. In the AMBER force field, the scoring function is composed of two energy components including a Lennard–Jones VDW term and an electrostatic term.

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

where *rij* represents the distance between protein atom *i* and ligand atom *j*; A_{ij} and B_{ij} denote the VDW repulsion and attraction parameters, respectively; and *qi* and *qj* are the partial charges corresponding to atoms *i* and *j*, respectively. In this equation, the solvent effect is implicitly considered by introducing a straightforward distance dependent dielectric constant $Dr(r_{ij})$ in the Coulombic term. The force field scoring function has a high computational efficiency. The distant-dependent dielectric constant cannot describe the desolvation effect [114]. There are other available methods that take into account the solvent by using explicit water molecules, such as the FEP method [105] and LIE method [103]. However, these methods are too computationally expensive to be used in virtual database screening. Therefore, accelerated force field models have been developed to perform molecular docking incorporating a more appropriate method for handling the

solvent effects by treating water as a continuum dielectric medium. Two popular examples that have been successfully used for performing virtual screening test are the Poisson–Boltzmann/surface area (PB/SA) model [115-117] and the generalized-Born/surface area (GB/SA) model [118-120].

Another force field-based scoring function is MedusaScore [121], in which the VDW attraction and repulsion terms, *vdw_attr* and *vdw_rep*, respectively; solvation (*sol*); and hydrogen bonding energies are included. The hydrogen bonding energies are divided into three groups: bonds formed between backbone atoms (*bb_hbond*), between side chains (*sc_hbond*) and between backbone and side chains (*bb_sc_hbond*) are taken into account. However, the hydrogen bond interactions between the protein and ligand are not present in the force field:

$$E = W_{vdw\_attr}E_{vdw\_attr} + W_{vdw\_rep}E_{vdw\_rep} + W_{sol}E_{sol} + W_{bb\_hbond} + W_{sc\_hbond}E_{sc\_hbond} + W_{bb\_sc\_hbond} \quad (2)$$

where the VDW energy parameters were taken from the CHARMM 19 united atom force field of [74], the hydrogen bond energy model was taken from Kortemme *et al.* [122], and the solvation energy model is an approximation of the EEF1 model of Lazaridis *et al.* [123]. These coefficients were trained to match the native structure of 38 high-resolution crystal structures determined by X-ray crystallography [124]. It is worth mentioning that the electrostatic energy term that is present in the AMBER force field is not present in MedusaScore even though the electrostatic interactions involved in hydrogen bond formation are implicitly considered, their mathematic form is not Columbic [125].

Knowledge-based scoring functions are collected from experimentally determined 3D data regarding the ligand-target complexes [126,127]. These scoring functions work according to the principle of the potential of mean force (PMF), where the energy of the complex is the sum of all of the interaction terms of the protein–ligand atom pairs [40]. Among the most important features of knowledge-based scoring functions is the introduction of an appropriate reference state for an optimal description of the residue–residue or atom–atom pairs in the non-interaction state. Since it is known that the predictive power of a knowledge-based scoring functions will depend on the definition of a suitable reference state, several attempts have been performed to establish an appropriate reference state value [128,129]. Derived from these early studies, it was stated that a reference sphere radius of at least 7 – 8 Å is needed to capture solvation effects [129]. However, despite the introduction of this useful approximation of the reference state, there were still some limitations as the fact of not including explicitly the contributions from solvation and entropy contributions [108]. More recently, it has been developed a new computational model that explicitly includes the contributions from solvation and entropy in the knowledge-based scoring functions named as ITScore/SE [108], which was evaluated using three important benchmarks of diverse protein–ligand complexes. Two popular scoring functions that include this implementation are DrugScore [108] and SMoG [130] where DrugScore includes solvent-accessibility corrections to the pair-wise potentials and SMoG utilizes pair-wise atom potentials to evaluate protein–ligand interactions. The major attraction of many knowledge-based scoring functions is that compared to the force field and empirical scoring functions, the knowledge-based scoring functions offer a good balance between precision and speed that allows efficient screening of a large database of compounds. A disadvantage is that these functions were primarily designed to reproduce experimental structures instead of evaluate binding energies.

In empirical scoring functions, the binding energy of a complex is scored by summing several intermolecular interaction terms:

$$\Delta G = \sum_i Wi.\Delta G_i \quad (3)$$

where ΔG_i represents several energy terms, such as VDW energy, electrostatic energy, hydrogen bonding energy, desolvation and entropy terms. The coefficient W_i is determined through a linear regression procedure in which the theoretical values are fitted to the experimental data. Therefore, empirical scoring functions are more computationally efficient than force field scoring functions.

One of the first empirical scoring functions (SCORE1) was developed by Bohm *et al.*, and consisted of four energy terms describing hydrogen bonds, ionic interactions, the lipophilic protein–ligand contact surface and the number of rotatable bonds in the ligand [131]. This scoring function was calibrated using 45 protein–ligand complexes. Then, this empirical scoring function was upgraded by increasing the dataset to 82 known 3D structures of protein–ligand complexes with known binding constants and including a more extended list of energy terms [132]. ChemScore [93] is an empirical scoring function developed by Eldridge *et al.*, in which some energetic terms and functions are similar to those adopted by Bohm *et al.* Some of these terms describe hydrogen bonds, metal atoms, the lipophilic effects of atoms and the effective number of rotatable bonds in the ligand [93]. This scoring function was calibrated with 82 protein–ligand complexes with known binding affinities and tested using two other sets of ligand–receptor complexes. This scoring function is implemented in several docking programs including GOLD [76] and FRED [32].

X-Score is a recently developed empirical scoring function [133] consisting of four energy terms describing VDW interactions, hydrogen bonds, hydrophobic effects and effective rotatable bonds. X-Score is based on a larger set of 200 protein–ligand complexes. Two hundred and eight hundred protein–ligand complexes were used to calibrate two versions of X-Score. Empirical scoring functions have been widely utilized in some well-known protein–ligand docking programs, such as FlexX [52] and Surflex [34]. However, the applicability of an empirical scoring function will depend on its training set [40], which often yields different weighting factors for the various terms. Therefore, the terms of different scoring functions cannot simply be recombined into a new scoring function.

In empirical scoring functions, some of the terms describing non-bonded interactions can be implemented in different ways. For example, in the early LUDI function [134], the hydrogen-bonding term was separated into neutral and ionic hydrogen bonds, while the hydrophobic contributions were based on a representation of the molecular surface area. In contrast, ChemScore [93] does not distinguish between different types of hydrogen bonds. ChemScore evaluates the contacts between hydrophobic atom pairs and F-score adds an extra term to account for aromatic interactions [52]. Non-enthalpic contributions, which is the so-called rotor term, can also be included in empirical scoring functions. These contributions approximate entropy penalties on binding from a weighted sum of the number of rotatable bonds in the ligand. In other cases, a more complicated form describes the molecular environment surrounding each rotatable bond, as implemented in the ligand rotational entropy of ChemScore [93]. Solvation and desolvation effects have also been addressed through complex functions, as reported in the Fresno scoring functions [135] that explicitly account for ligand desolvation and desolvation energies using a continuum electrostatic model [136,137]. During docking procedures, scoring functions are used to optimize ligand placement. After a docking procedure is complete, the scoring functions are used to rank each ligand from the docking procedure. This ranking process will predict which ligand has the best affinity.

Among the methods for estimating the binding constant between a receptor and its ligand, scoring functions are able to determine the binding free energy or binding constant (*Ka*). The binding free energy is obtained using the integrated form of the Gibbs–Helmholtz equation at constant temperature:

$$\Delta G = \Delta H - T\Delta S \quad (4)$$

where ΔG is the change in the free energy of binding, ΔH is the change in enthalpy, T is the temperature in degrees Kelvin and ΔS is the change in entropy. The binding constant for a complex, *Ka*, can be related to the standard free energy change ΔG^0 by Equations 5 and 6.

$$\Delta G^0 = \mathrm{RT} \ln K_\mathrm{a} \quad (5)$$

$$k_\mathrm{a} = \frac{1}{k_\mathrm{d}} + \frac{[PL]}{[P][L]} \quad (6)$$

where *Ka* is the binding constant (association), its inverse is termed the dissociation constant (*Kd*) in the pharmacology field, *L* refers to the ligand, *P* refers to the protein and *L–P* is a ligand–protein complex (**Figure 1**).

Figure 1 shows the predicted binding conformation for a complex between valproic acid and Cytochrome 2E1 (from molecular modeling) obtained using docking procedures under equilibrium conditions by our work group.

However, it is known that the binding free energy estimate greatly decreases in accuracy when side chains are misplaced in the binding site [138]. A new mean force potential for the knowledge-based scoring function ROTA [139] has been developed that is primarily used for scoring protein and small ligand complexes with modeled side-chain conformations.

3. Expert opinion

3.1 Side chain flexibility

It is known that physiologically, molecules (small and macro) are in constant motion inside water and lipid–water environments. Therefore, several programs utilized for docking studies have incorporated some flexibility properties in which backbones are fixed and side-chains are flexible [140]. This approach has many advantages, particularly when the principal residues undergo significant conformational changes, as in some GPCR proteins, such as Trp 6.48, which is better known as the toggle switch [141]. In addition, side chain flexibility can be used to explain the catalytic process of some enzymes, where the substrate fits into the active site and yields some products [142]. Therefore, more detailed studies are required that are focused on experimental data and aided by quantum chemistry, which has not been widely used [143]. Consequently, rigid proteins and flexible ligands are still used in virtual screening computations [144]. In addition, there are some disadvantages to modeling side-chain flexibility. It can be difficult to obtain reliable data because the process of ligand recognition in proteins involves more than side chain flexibility and the computational costs are high. For example, we used both flexible and non-flexible side chain residues of topoisomerase II in a study that showed that the free energy of side chain residues is higher in flexible systems than in rigid systems [143].

One approach utilized to incorporate protein flexibility during docking is the use of a side-chain prediction tool, such as IRECS [139]. The IRECS program allows for the creation of an ensemble of rotamers for the side chains involved in the complex stabilization. To simultaneously include more protein components, such as side chain residue properties, and be able to quickly attain docking results, some research groups have developed search algorithms named Iterative Multi-Greedy Docking (IMGDock) [145]. This algorithm places the ligand into the potential binding site by trying all of the possible conformations to determine the ones that have the lowest free energy values. Furthermore, during protein– ligand recognition, more than side chain residues are involved. Recently, a stochastic tunneling algorithm (STUN) has been used that takes into account the backbone loop motions [146].

3.2 MD and docking simulations

Because ligands and targets are flexible and that ligands can reach their targets in different conformational states, MD simulations can be used to sample many snapshot structures of a protein–ligand complex. This methodology has been recently explored by our group [107,147] because it has the advantage of exploring the natural behavior of proteins. This approach allows one to obtain information about ligand selection that is more reliable than information from a rigid-body approach, similar to the methodology we used to explain CYP450 metabolism [148],

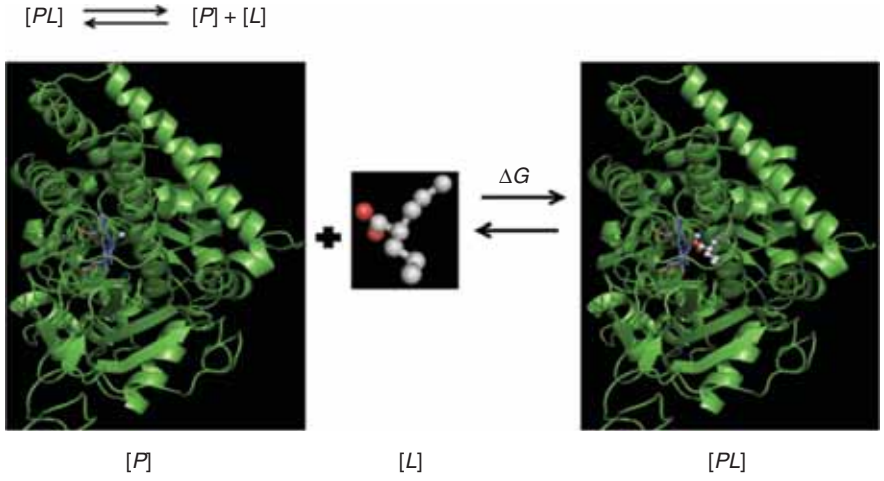

Figure 1. Docking simulation results, superposition of the highest-ranked conformation between valproic acid and Cytochrome 2E1, and the thermodynamic cycle describing the formation of the complex by our work group.

which did not depict the conformational changes of the ligand and protein during their coupling process. However, conformational changes during coupling in solution can be explored using MD simulations, as previously performed in our research group [147]. Unfortunately, MD simulations have high computational costs, but they can help to explain some interactions, including ligand accommodation and catalysis mechanisms.

3.3 Induced fit

Small ligands and targets are dynamic at all times, including during recognition processes [149]. MD simulations can be used to explore dynamics under free conditions in water or water/membranes. However, when a small ligand approaches a protein molecule, the protein can undergo structural changes due to non-covalent interactions with the ligand [149]. In fact, during this recognition process, many conformational changes occur until the ligand and protein reach the best free energy value of coupling. This physical process can be explored under a protocol named 'induced fit' even though only a few programs are designed to perform these theoretical assays [150].

3.4 Docking simulations for drug design

One method to validate a docking procedure is to reproduce the crystal complexes of small ligands and targets. After performing the protein evaluation and validation, the 3D model of the target is reliable and can be used for virtual screening in docking applications, which is reflected by the results for many ligand–protein complexes that reproduce experimental findings, as previously reported in the literature [151]. Thereby, many designers of docking programs have delivered their software after validation with Autodock or other docking programs [101].

It is possible that in the near future, many drugs could be designed with this strategy [152]. For example, HIV protease inhibitors have been developed through structure-based design and screening [153]. However, at this time, the number of drugs discovered with these theoretical strategies is still low because these methodologies require several steps before selecting the best compounds, which must then be synthesized and tested in preclinical assays, which is a process that may require several years. It is interesting that new methodologies have been developed to improve the performance of molecular docking, such as high-throughput screening and combinatorial chemistry [154].

In conclusion, we have reviewed the current aspects of the protein–small ligand problem, such as search algorithms, scoring functions and protein flexibility. In contrast with search algorithms and scoring functions, which have been extensively studied in the last two decades, protein flexibility has only recently been incorporated in docking studies. However, this method still has a high computational cost due to the enormous degrees of freedom of proteins (amino acid residues). Nevertheless, it has recently been shown through several computational studies that coupling docking with MD simulations can produce successful results [155]. Furthermore, modeling the different conformations present in a solvated system and then docking them with a ligand may enable researchers to better elucidate the biological aspect of molecular recognition.

Acknowledgements

M Bello and M Martínez-Archundia are supported by a fellowship from Consejo Nacional de Ciencia y Tecnología (CONACYT). The authors also acknowledge support through grants from the Instituto de Ciencia y Tecnología del Distrito Federal (ICyTDF), grant identifier PIRIVE09-9, PIFI-SIP-COFAA/IPN projects SIP (20130379) and CONACYT grant CB132353.

Declaration of interest

The authors state no conflict of interest and have received no payment in preparation of this manuscript.

Bibliography

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Hong-Fang J, Xue-Juan L, Hong-Yu Z. Natural products and drug discovery. Can thousands of years of ancient medical knowledge lead us to new and powerful drug combinations in the fight against cancer and dementia? EMBO Rep 2009;10(3):194-200
2. Cay-Rüdiger P. Part of a scientific master plan? Paul Ehrlich and the origins of his receptor concept. Med Hist 2003;47(3):332-56
3. Ehrlich P, Himmelweit F. Gesammelte Arbeiten. Springer-Verlag; Berlin: 1957
4. Langley JN. On the reaction of cells and of nerve-endings to certain poisons, chiefly as regards the reaction of striated muscle to nicotine and to curare. J Physiol 1905;33(4-5): 374-413
5. Maehle AH, Prüll CR, Halliwell RF. The emerge of the drug receptor theory. Nat Rev Drug Discov 2002;1(8):637-41

• **Refers to protein target.**

6. Carpenter EP, Konstantinos BD, Cameron A, Iwata S. Overcoming the challenges of membrane protein crystallography. Curr Opin Struct Biol 2008;18(5):581-6
7. Ozawa T. Crystallization of intrinsic membrane proteins. J Bioenerg Biomembr 1984;16(5-6):321-33

• **Obtaining 3D model of protein targets.**

8. Cherezov V, Rosenbaum DM, Hanson MA, et al. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. Science 2007;318(5854):1258-65

• **Obtaining 3D model of protein targets.**

9. Danley DE. Crystallization to obtain protein-ligand complexes for structure-aided drug design. Acta Cryst 2006;62(pt 6):569-75
10. Mandal S, Moudgil M, Mandal SK. Rational drug design. Eur J Pharmacol 2009;625(1-3):90-100

•• **Refer to designing compounds under *in silico*.**

11. Jackson RC. Update on computer-aided drug design. Curr Opin Biotechnol 1995;6(6):646-51
12. Leis S, Zacharias M. ReFlex in: a flexible receptor protein-ligand docking scheme evaluated on HIV-1 protease. PLoS One 2012;7(10):e48008

•• **Flexible docking.**

13. Dhanik A, Kavraki LE. Protein-ligand interactions: computational Docking. John Wiley & Sons; 2001. DOI: 10.1002/9780470015902.a0004105.pub2. Available at: http://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0004105.pub2/abstract
14. Arulalapperumal V, Sakkiah S, Thangapandian S. Ligand base pharmacophore identification and molecular docking studies for Grb2 inhibitors. Korean Chem Soc 2012;33(5):1707
15. 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 2010;31(2):455-61
16. Zavodszky MI, Kuhn LA. Side-chain flexibility in protein-ligand binding: the minimal rotation hypothesis. Protein Sci 2005;14(4):1104-14

•• **Flexible docking.**

17. Totrov M, Abagyan R. Flexible protein-ligand docking by global energy optimization in internal coordinates. Proteins 1997;1:215-20

•• **Flexible docking.**

18. Scott DR, Borbulevych OY, Piepenbrink KH, et al. Disparate degrees of hypervariable loop flexibility control T-cell receptor cross-reactivity, specificity, and binding mechanism. J Mol Biol 2011;414(3):385-400

•• **Flexible loop explain its biological properties.**

19. Cohen EM, Machado KS, Cohen M, de Souza ON. Effect of the explicit flexibility of the InhA enzyme from Mycobacterium tuberculosis in molecular dockingsimulations. BMC Genomics 2011;12(Suppl 4):S7
20. Nabuurs SB, Wagener M, Vlieg J. A flexible approach to induced fit docking. J Med Chem 2007;50(26):6507-18

• **Induced fit improve docking predictions.**

21. Koska J, Spassov VZ, Maynard AJ, et al. Fully automated molecular mechanics based induced fit protein-ligand docking method. J Chem Inf Model 2008;48(10):1965-73

• **Induced fit improve docking predictions.**

22. Taylor RD, Jewsbury PJ, Essex JW. A review of protein-small molecule docking methods. J Comput Aided Mol Des 2002;16(3):151-66

•• **Widely information about scoring function and sampling.**

23. Perozzo R, Kuo M, Sidhu AbS, et al. Structural elucidation of the specificity of the antibacterial agent triclosan for malarial enoyl acyl carrier protein reductase. J Biol Chem 2002;277(15):13106-14
24. Pereira JH, de Oliveira JS, Canduri F, et al. Structure of shikimate kinase from Mycobacterium tuberculosis reveals the binding of shikimic acid. Acta Crystallogr D Biol Crystallogr 2004;60(Pt 12 Pt 2):2310-19
25. Oliveira JS, Pereira JH, Canduri F, et al. Crystallographic and pre-steady-state kinetics studies on binding of NADH to wild-type and isoniazid-resistant enoyl-ACP(CoA) reductase enzymes from Mycobacterium tuberculosis. J Mol Biol 2006;359(3):646-66
26. Schulze-Gahmen U, De Bondt HL, Kim SH. High-resolution crystal structures of human cyclin-dependent kinase 2 with and without ATP: bound waters and natural ligand as guides for inhibitor design. J Med Chem 1996;39(23):4540-6
27. Chrencik JE, Staker BL, Burgin AB, et al. Mechanisms of camptothecin resistance by human topoisomerase I mutations. J Mol Biol 2004;339(4):773-84
28. Timmers LF, Caceres RA, Vivan AL, et al. Structural studies of human purine nucleoside phosphorylase: towards a new specific empirical scoring function. Arch Biochem Biophys 2008;479:28-38
29. Dias R, de Azevedo WF Jr. Molecular docking algorithms. Curr Drug Targets 2008;9(12):1040-7

•• **Widely information about scoring function and sampling.**

30. Kuntz ID, Blaney JM, Oatley SJ, et al. A geometric approach to

macromolecule-ligand interactions. J Mol Biol 1982;161(2):269-88

31. Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. Proteins 2003;52(1):80-7
32. McGann MR, Almond HR, Nicholls A, et al. Gaussian docking functions. Biopolymers 2003;68(1):76-90
33. 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):153-74
34. Jain AN. Surflex: fully automatic molecular docking using a molecular similarity-based search engine. J Med Chem 2003;46(4):499-511
35. Venkatachalam CM, Jiang X, Oldfield T, Waldan M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. J Mol Graph Model 2003;21(4):289-307
36. Pang YP, Perola E, Xu K, Prendergast FG. EUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. J Comput Chem 2001;22(15):1750-71
37. Pang Y.P, Kozikowski A.P. Prediction of the binding sites of huperzine A in acetylcholinesterase by docking studies. J Comput Aid Mol Des 1994;8:669-81
38. Kollman PA. AMBER 5.0. University of California; San Francisco: 1996
39. Lorber DM, Shoichet BK. Hierarchical docking of databases of multiple ligand conformations. Curr Top Med Chem 2005;5(8):739-49
40. Huang SY, Zou X. Advances and challenges in protein-ligand docking. Int J Mol Sci 2010;11(8):3016-34

•• **Update about docking simulations.**

41. Choi J, He N, Kim N, Yoon S. Enrichment of virtual hits by progressive shape-matching and docking. J Mol Graph Model 2012;32:82-8
42. Friesner RA, Banks JL, Murphy RB, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 2004;47:1739-49

• **Accurate of docking simulations.**

43. Halgren TA, Murphy RB, Friesner RA, et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem 2004;47:1750-9
44. 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):935-49

• **Advantages of docking studies.**

45. Schultz-Gasch T, Stahl M. Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. J Mol Model 2003;9:47-57

• **Structural studies taking into account of docking.**

46. Lorber DM, Shoichet BK. Flexible ligand docking using conformational ensembles. Protein Sci 1998;7:938-50
47. Joseph-McCarthy D, Thomas BEIV, Belmarsh M, et al. Pharmacophore-based molecular docking to account for ligand flexibility. Proteins 2003;51:172-88
48. Sauton N, Lagorce D, Villoutreix B, Miteva M. MS-DOCK: accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. BMC Bioinformatics 2008;9:184

•• **Considering many protein conformers.**

49. Huang SY, Zou X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. Proteins 2007;66:399-421
50. Huang SY, Zou X. Efficient molecular docking of NMR structures: application to HIV-1 protease. Protein Sci 2007;16:43-51
51. Brylinski M, Skolnick J. Q-Dock: low-resolution flexible ligand docking with pocket-specific threading restraints. J Comput Chem 2008;29:1574-88
52. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. J Mol Biol 1996;261:470-89
53. Klebe G, Mietzner T. A fast and efficient method to generate biologically relevant conformations. J Comput Aided Mol Des 1994;8(5):583-606
54. Rarey M, Kramer B, Lengauer T. Multiple automatic base selection: protein-ligand docking based on incremental construction without manual intervention. J Comput Aided Mol Des 1997;11(4):369-84
55. Klebe G. The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands. J Mol Biol 1994;237:221-35
56. Böhm 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):243-56
57. Rarey M, Kramer B, Lengauer T. The particle concept: placing discrete water molecules during protein-ligand docking predictions. Proteins 1999;34(1):17-28
58. Ewing TJA, Kuntz ID. Critical evaluation of search algorithms for automated molecular docking and database screening. J Comput Chem 1997;18:1175-89
59. Bron C, Kerbosch J. Finding all cliques of an undirected graph. J Commun ACM 1973;16:575-6
60. Brint AT, Willett P. Algorithms for the identification of three-dimensional maximal common substructures. J Chem Inform Comput Sci 1987;27:152-8
61. Grindley HM, Artymiuk PJ, Rice DW, Willett P. Identification of tertiary structure resemblance in proteins using a maximal common subgraph isomorphism algorithm. J Mol Biol 1993;229(3):707-21
62. Makino S, Kuntz ID. Automated flexible ligand docking method and its application for database search. J Comput Chem 1997;18:1812-25
63. Biesiada J, Porollo A, Velayutham P, et al. Survey of public domain software for docking simulations and virtual screening. Hum Genomics 2011;5(5):497-505
64. Claussen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. J Mol Biol 2001;308(2):377-95
65. Spoel DVD, Lindahl E, Hess B, et al. GROMACS: fast, flexible, and free. J Comput Chem 2005;26(16):1701-18

•• **Including molecular dynamics simulations to docking studies.**

66. Liu M, Wang S. MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. J Comput Aided Mol Des 1999;13(5):435-51
67. Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. Proteins 1990;8(3):195-202

68. Morris GM, Goodsell DS, Huey R, Olson AJ. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. J Comput Aided Mol Des 1996;10(4):293-304

69. Hart TN, Read RJ. A multiple-start Monte Carlo docking method. Proteins 1992;13:206-22

70. Abagyan R, Totrov M, Kuznetzov 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

71. McMartin C, Bohacek RS. QXP: powerful, rapid computer algorithms for structure-based drug design. J Comput Aided Mol Des 1997;11:333-44

72. Trosset JY, Scheraga HA. Prodock: software package for protein modeling and docking. J Comput Chem 1999;20:412-27

73. Némethy G, Gibson KD, Palmer KA, et al. Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. J Phys Chem 1992;96:6472-84

74. Brooks BR, Bruccoleri RE, Olafson BD, et al. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4:187-217

75. Argos P, Abagyan R. Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. J Mol Biol 1992;225:519-32

76. Jones G, Willett P, Glen RC, et al. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267:727-48

• **Scoring sampling developments.**

77. Morris GM, Goodsell DS, Halliday RS, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 1998;19:1639-62

•• **Update about docking simulations.**

78. Clark KP. Flexible ligand docking without parameter adjustment across four ligand-receptor complexes. J Comput Chem 1995;16:1210-26

79. Taylor JS, Burnett RM. DARWIN: a program for docking flexible molecules. Proteins 2000;41:173-91

80. Thomsen R, Christensen MH. MolDock: a new technique for highaccuracy molecular docking. J Med Chem 2006;49:3315-21

81. Pei J, Wang Q, Liu Z, et al. PSI-DOCK: towards highly efficient and accurate flexible ligand docking. Proteins 2006;62:934-46

82. Zhao Y, Sanner MF. FLIPDock: docking flexible ligands into flexible receptors. Proteins 2007;68:726-37

• **Making side chain residues of proteins.**

83. Stroganov OV, Novikov FN, Stroylov VS, et al. Lead finder: an approach to improve accuracy of protein-ligand docking, binding energy estimation, and virtual screening. J Chem Inf Model 2008;48:2371-85

84. Grosdidier A, Zoete V, Michielin O. EADock: docking of small molecules into protein active sites with a multiobjective evolutionary optimization. Proteins 2007;67:1010-25

85. Perola E, Walters WP, Charifson PS. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins 2004;56:235-49

• **Advances of scoring functions.**

86. Li X, Li Y, Cheng T, et al. Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes. J Comput Chem 2010;31(11):2109-25

• **Evaluating programs for docking studies.**

87. Kellenberger E, Rodrigo J, Muller P, Rognan D. Comparative evaluation of eight docking tools for docking and virtual screening. Proteins 2004;57:225-42

• **Evaluating programs for docking studies.**

88. Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: current status and future challenges. Proteins 2006;65:15-26

89. Schames JR, Henchman RH, Siegel JS, et al. Discovery of a novel binding trench in HIV integrase. J Med Chem 2004;47(8):1879-81

• **Docking studies to develop drugs.**

90. Cosconati S, Forli S, Perryman AL, et al. Virtual screening with AutoDock: theory and practice. Expert Opin Drug Discov 2010;5(6):597-607

91. Norgan AP, Coffman PK, Kocher JP, et al. Multilevel parallelization of AutoDock 4.2. J Cheminform 2011;3(1):12

• **Increasing results under parallelization of autodock.**

92. Baxter CA, Murray CW, Clark DE, et al. Flexible docking using Tabu search and an empirical estimate of binding affinity. Proteins 1998;33:367-82

93. Eldridge MD, Murray CW, Auton TR, et al. 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;11:425-45

•• **Scoring function to calculate free energy values.**

94. Baxter CA, Murray CW, Waszkowycz B, et al. New approach to molecular docking and its application to virtual screening of chemical databases. J Chem Inform Comput Sci 2000;40:254-62

95. Chen HM, Liu BF, Huang HL, et al. SODOCK: swarm optimization for highly flexible protein-ligand docking. J Comput Chem 2007;28:612-23

96. Namasivayam V, Gunther R. PSO@Autodock: a fast flexible molecular docking program based on swarm intelligence. Chem Biol Drug Des 2007;70:475-84

97. Chen K, Li T, Cao T. Tribe-PSO: a novel global optimization algorithm and its application in molecular docking. Chemom Intell Lab Syst 2006;82:248-59

98. Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. J Comput Chem 2005;16:1781-802

• **Molecular dynamics simulations can be mixed with docking studies.**

99. Bello M, Gutiérrez G, García-Hernández E. Structure and dynamics of beta-lactoglobulin in complex with dodecyl sulfate and laurate: a molecular dynamics study. Biophys Chem 2012;165-166:79-86

•• **Ligand-protein complexes can be submit to molecular dynamics simulations.**

100. Bello M, Valderrama B, Serrano-Posada H, Rudiño-Piñera E. Molecular Dynamics of a Thermostable Multicopper Oxidase from Thermus thermophilus HB27: structural Differences between the Apo and

Holo Forms. PLoS One 2012;7(7):e40700

101. Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. Med Res Rev 2006;26(5):531-68

•• **Ligand-protein complexes can be submit to molecular dynamics simulations.**

102. Lopez MA, Kollman PA. Application of molecular dynamics and free energy perturbation methods to metalloporphyrin-ligand systems II: CO and dioxygen binding to myoglobin. Protein Sci 1993;2(11):1975-86

•• **Free energy values can be obtained under molecular dynamics simulations.**

103. Aqvist J, Marelius J. The linear interaction energy method for predicting ligand binding free energies. Comb Chem High Throughput Screen 2001;4(8):613-26

•• **Free energy values can be obtained under molecular dynamics simulations.**

104. Tsfadia Y, Friedman R, Kadmon J, et al. Molecular dynamics simulations of palmitate entry into the hydrophobic pocket of the fatty acid binding protein. FEBS Lett 2007;581:1243-7

105. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. Annu Rev Biophys Biomol Struct 2004;32:335-73

106. Di Nola A, Berendsen HJC, Roccatano D. Molecular dynamics simulation of the docking of substrates to proteins. Proteins 1994;19:174-82

107. Martinez-Ramos F, Fonseca-Sabater Y, Soriano-Ursúa MA, et al. o-Alkylselenenylated benzoic acid accesses several sites in serum albumin according to fluorescence studies, Raman spectroscopy and theoretical simulations. Protein Pept Lett 2012;Epub ahead of print

• **Sampling many snapshots of protein with docking.**

108. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol 2002;295(2):337-56

109. Kollman PA. Free energy calculations: applications to chemical and biochemical phenomena. Chem Rev 1993;93:2395-417

110. Simonson T, Archontis G, Karplus M. Free energy simulations come of age: protein-ligand recognition. Acc Chem Res 2002;35:430-7

111. Verdonk ML, Cole JC, Hartshorn MJ, et al. Improved protein-ligand docking using GOLD. Proteins 2003;52(4):609-23

112. 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):2770-80

113. Krammer A, Kirchhoff PD, Jiang X, et al. LigScore: a novel scoring function for predicting binding affinities. J Mol Graph Model 2005;23(5):395-407

114. Huang SY, Grinter SZ, Zou X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. Phys Chem Chem Phys 2010;12(40):12899-908

115. Rocchia W, Sridharan S, Nicholls A, et al. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. J Comput Chem 2002;23:128-37

116. Grant JA, Pickup BT, Nicholls A. A smooth permittivity function for Poisson-Boltzmann solvation methods. J Comput Chem 2001;22:608-40

• **Physical properties to calculate free energy values.**

117. Baker NA, Sept D, Joseph S, et al. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA 2001;98:10037-41

118. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112(16):6127-9

119. Hawkins GD, Cramer CJ, Truhlar DG. Pairwise solute descreening of solute charges from a dielectric medium. Chem Phys Lett 1995;246:122-9

120. Qiu DI, Shenkin PS, Hollinger FP, Still WC. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. J Phys Chem A 1997;101:3005-14

• **Physical properties to calculate free energy values.**

121. Yin S, Biedermannova L, Vondrasek J, Dokholyan NV. MedusaScore: an accurate force field-based scoring function for virtual drug screening. J Chem Inform Model 2008;48:1656-62

• **Scoring function to calculate free energy values.**

122. Kortemme T, Morozov AV, Baker D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. J Mol Biol 2003;326:1239-59

123. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Proteins Struct Funct Genet 1999;35:133-52

124. Ding F, Dokholyan NV. Emergence of protein fold families through rational design. Plos Comput Biol 2006;2:725-33

125. Wang JC, Lin JH. Scoring functions for prediction of protein-ligand interactions. Curr Pharm Des 2012;Epub ahead of print

• **Scoring function to calculate free energy values.**

126. Sippl MJ. Calculation of conformational ensembles from potentials of mean force. J Mol Biol 1990;213:859-83

127. 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;8:677-91

128. 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

129. Muegge I. A knowledge-based scoring function for protein-ligand interactions: probing the reference state. Perspect Drug Discov Des 2000;20:99-114

130. 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;118:11733-44

131. Böhm 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):243-56

132. Böhm 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):309-23

133. Wang R, Lai L, Wang S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. J Comput Aided Mol Des 2002;16(1):11-26

134. Böhm HJ. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. J Comput Aided Mol Des 1992;6:593-606

135. Rognan D, Lauemoller SL, Holm A, et al. Predicting binding affinities of protein ligands from three-dimensional models: application to peptide binding to class I major histocompatibility proteins. J Med Chem 1999;42:4650-8

136. Sitkoff DF, Sharp KA, Honig B. Accurate calculation of hydration free energies using macroscopic continuum models. J Phys Chem 1998;98:1978-83

137. Huo S, Wang J, Cieplak P, et al. Molecular dynamics and free energy analyses of cathepsin D–inhibitor interactions: insight into structure based ligand design. J Med Chem 2002;45:1412-19

138. Thorsteinsdottir HB, Schwede T, Zoete V, Meuwly M. How inaccuracies in protein structure models affect estimates of protein-ligand interactions: computational analysis of HIV-I protease inhibitor binding. Proteins 2006;65:407-23

139. Hartmann C, Antes I, Lengauer T. IRECS: a new algorithm for the selection of most probable ensembles of side-chain conformations in protein models. Protein Sci 2007;16:1294-307

140. Jackson RM, Gabb HA, Sternberg MJ. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. J Mol Biol 1998;276(1):265-85

141. Bhattacharya S, Hall SE, Vaidehi N. Agonist-induced conformational changes in bovine rhodopsin: insight into activation of G-protein-coupled receptors. J Mol Biol 2008;382(2):539-55

142. Flachner B, Varga A, Szabó J, et al. Substrate-assisted movement of the catalytic Lys 215 during domain closure: site-directed mutagenesis studies of human 3-phosphoglycerate kinase. Biochemistry 2005;44(51):16853-65

143. Correa-Basurto J, Ramos-Morales FR, Matus MH, et al. Docking and DFT Studies to explore the Topoisomerase II ATP Pocket employing 3-Substituted 2,6-Piperazindiones for drug design. Mol Simul 2012;38(13):1072-84

144. Chang MW, Ayeni C, Breuer S, Torbett BE. Virtual screening for HIV protease inhibitors: a comparison of AutoDock 4 and Vina. PLoS One 2010;5(8):e11955

•• **Testing a new faster program than autodock.**

145. Schumann M, Armen RS. Systematic and efficient side chain optimization for molecular docking using a cheapest-path procedure. J Comput Chem 2013; [Epub ahead of print]

146. Flick J, Tristram F, Wenzel W. Modeling loop backbone flexibility in receptor-ligand docking simulations. J Comput Chem 2012;33(31):2504-15

• **Including loop motion in docking simulations.**

147. Durán LA, Hernández MC, Hernández-Rodríguez M, et al. Mapping myeloperoxidase to identify its promiscuity properties using docking and molecular dynamics simulations. Curr Pharm Des 2012

148. Rosales-Hernández MC, Mendieta-Wejebe JE, Trujillo-Ferrara JG, Correa-Basurto J. Homology modeling and molecular dynamics of CYP1A1 and CYP2B1 to explore the metabolism of aryl derivatives by docking and experimental assays. Eur J Med Chem 2010;45(11):4845-55

• **Explaining substrate recognition on cytochomes under docking and molecular dynamics simulations.**

149. Cook PD, Carney AE, Holden HM. Accommodation of GDP-linked sugars in the active site of GDP-perosamine synthase. Biochemistry 2008;47(40):10685-93

150. Kalid O, Toledo Warshaviak D, Shechter S, et al. Consensus Induced Fit Docking (cIFD): methodology, validation, and application to the discovery of novel Crm1 inhibitors. J Comput Aided Mol Des 2012;26(11):1217-28

•• **Induce fit could yield more reliable data than traditional docking procedure.**

151. Grosdidier A, Zoete V, Michielin O. Blind docking of 260 protein-ligand complexes with EADock 2.0. J Comput Chem 2009;30(13):2021-30

152. Ruvinsky AM. Role of binding entropy in the refinement of protein-ligand docking predictions: analysis based on the use of 11 scoring functions. J Comput Chem 2007;28(8):1364-72

153. Vacca JP, Condra JH. Clinically effective HIV-1 protease inhibitors. Drug Discov Today 1997;2:261-72

154. Okimoto N, Futatsugi N, Fuji H, et al. High-performance drug discovery: computational screening by combining docking and molecular dynamics simulations. PLoS Comput Biol 2009;5(10):e1000528

155. Lu SY, Jiang YJ, Lv J, et al. Molecular docking and molecular dynamics simulation studies of GPR40 receptor-agonist interactions. J Mol Graph Model 2010;28(8):766-74

Affiliation

Martiniano Bello, Marlet Martínez-Archundia & José Correa-Basurto†
†Author for correspondence
Laboratorio de Modelado Molecular y Bioinformática de la Escuela Superior de Medicina, Instituto Politécnico Nacional, México. Plan de San Luis Y Diaz Mirón S/N, Col. Casco de Santo Tomas, Mexico city, México, CP 11340, USA
Tel: +1 5255 57296000 Ext 62747 and 62767; Fax: +1 5255 57296000 Ext 62747 and 62767; E-mail: jcorreab@ipn.mx