



# Molecular Docking an Important Tool for Drug Designing

**Rahul Haque, Mukti Mohammad and Md Maidul Islam\***

*Department of Chemistry, Aliah University, India*

**\*Corresponding author:** Md Maidul Islam, Department of Chemistry, Aliah University, Kolkata-700156, India

**Submission:** August 10, 2017; **Published:** January 08, 2018

## Abstract

Molecular Docking plays a significant role in drug discovery and the knowledge on molecular docking will help us to design appropriate drugs in future. Several types of docking methodologies were used depending on the criteria of drug designing. Scoring functions play a significant role to short out the proper data obtained from molecular docking. In this review we have discussed several method of docking and several methods of choosing scoring functions.

## Introduction

Molecular docking is a growing field in the discovery new biologically active molecules and medicinal research. It predicts the biological activity, and several physical parameters of a molecule. Molecular docking research focuses on computationally simulating the molecular recognition process. It aims to achieve an optimized conformation for both the protein/DNA/RNA and other biological receptors and ligand and relative orientation between receptors and ligand such that the free energy of the overall system is minimized.

Molecular docking is one of the most frequently used methods in structure-based drug design, due to its ability to predict the binding-conformation of small molecule ligands to the appropriate target binding site. Characterisation of the binding behaviour plays an important role in rational design of drugs as well as to elucidate fundamental biochemical processes [1].

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 [2]. Molecular docking attempts to arrange molecules in favorable configurations by matching complementary features [3]. This is a difficult task because there are many ways in which complex molecules can be associated. The problem is further complicated by an exponential dependence on molecule size, so that the number of possible configurations explodes when docking involves biological macromolecules such as proteins or nucleic acid polymers. Current docking methodologies thus invoke either geometric- or energy-based schemes to guide configurational sampling [4], the former relying upon the matching of topographical features and the latter upon optimization along a potential energy surface of some kind. As alluded to earlier, however, configurational sampling is only

half of the problem. The ranking of each configuration by some measurement of complementarity constitutes the other major hurdle [5]. The structure-based design methods used to optimize these leads into drugs are now often applied much earlier in the drug discovery process. Protein structure is used in target identification and selection (the assessment of the 'drug ability' or tractability of a target), in the identification of hits by virtual screening and in the screening of fragments. Additionally, the key role of structural biology during lead optimization to engineer increased affinity and selectivity into leads remains as important as ever. Each of these topics will be outlined, using the field of kinase drug discovery as an example of the role of structure in lead optimization.

Just like in protein folding, solving the docking problem also involves two components: an efficient search procedure and a good scoring function. The two critical elements in a search procedure are speed and effectiveness in covering the relevant conformational space. On the other hand, the scoring function should be fast enough to allow its application to a large number of potential solutions and, in principle, effectively discriminate between native and non-native docked conformations. The scoring function should include and appropriately weigh all the energetic ingredients. Hence, as in folding, the performance of a particular docking program should not be viewed as representing one complete piece. To solve the docking problem, ideally, the best matching algorithms and scoring schemes should be combined. Similar considerations and division have recently been discussed [6-8]. The three aspects of the docking are mutually inter related. The choice of the system (surface) representation decides the types of conformational search algorithms, and the ways to rank potential solutions. Below, we review the principles of the representation, available search algorithms, and scoring schemes. Based on these, we highlight some potential promising approaches.

Many reviews of docking algorithms and scoring functions have been published in recent years [9-15]. It is commonly reported in these reviews that docking programs are often able to reproduce the correct pose of protein-ligand complexes (alongside many incorrect poses), and that the problem lies in the accurate estimation of the relative binding affinities of ligand poses, i.e. the scoring function.

### Different Type of Scoring Function

The scoring function is one of the most important components in structure-based drug design. Despite considerable success, accurate and rapid prediction of protein-ligand interactions is

still a challenge in molecular docking. In this perspective, we have reviewed three basic types of scoring functions (force-field, empirical, and knowledge-based) and the consensus scoring technique that are used for protein-ligand docking. The different scoring functions have been developed that exhibit different accuracies and computational efficiencies. In this section, we will briefly review the scoring functions in literature developed for protein-ligand interactions in molecular docking [16]. Some of the commonly-used scoring functions are summarized in (Figure 1 and Tables 1-4).

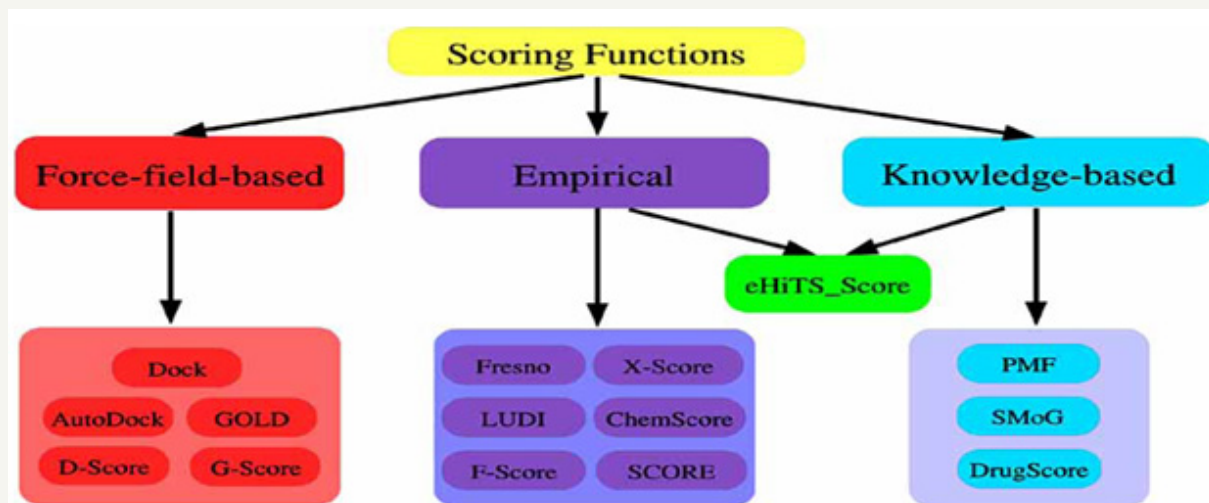


Figure 1

Table 1: Types of scoring functions

Type	Scoring Function
Force field-based	DOCK [17], DOCK 3.5(PB/SA) [18,19], DOCK/GBSA(SDOCK)[20-22], Auto Dock [23,24], GOLD [25], SYBYL/D-Score [17], SYBYL/G-Score [25]
Empirical	FlexX [26], Glide [27], ICM [28], LUDI [29,30], PLP [31,32], ChemScore [33], SCORE [34], X-Score [35], Surflex [36], SYBYL/F-Score [26], LigScore [37], MedusaScore [38], AIScore [39], SFCscore [40]
Knowledge-based	ITScore [41-43], PMF [44,45], DrugScore [46,47], DFIRE [48], SMOG [49,50], BLEEP [51,52], MScore [53], GOLD/ASP [54], KScore [55]

### Force-field-based scoring function

Affinities are estimated by summing the strength of intermolecular vander waals and electrostatic interactions between all atoms of the two molecules in the complex using a force field. The intra molecular energies (also referred to as strain energy) of the two binding partners are also frequently included. Molecular mechanics force fields usually quantify the sum of two energies, the receptor-ligand interaction energy and internal ligand energy (such as steric strain induced by binding).

Table 2: Success rates of 16 scoring functions for Wang et al.'s test set of 100 diverse protein-ligand complexes, using the criterion of rmsd  $\leq 2.0$  Å [43].

Scoring Function	Type Of Scoringa	Success Rate (%)
ITScore/SE [43]	K	91
Drug Score CSD [47]	K	87
ITScore [41]	K	82
Cerius2/PLP [31,32]	E	76
SYBYL/F-Score [26]	E	74
Cerius2/LigScore [37]	E	74
Drug ScorePDB [46]	K	72
Cerius2/LUDI [29,30]	E	67
X-Score [35]	E	66
Auto Dock [23]	F	62
DFIRE [48]	K	58
DOCK/FF [17]	F	58
Cerius2/PMF [44]	K	52
SYBYL/G-Score [25]	F	42
SYBYL/Chem Score [33]	E	35
SYBYL/D-Score [17]	F	26

a“K” stands for knowledge-based scoring functions, “E” for empirical scoring functions, and “F” for force field scoring functions, respectively.

Force-field-based scoring is further complicated by the fact that it generally requires the introduction of cut-off distances for the treatment of non-bonded interactions, which are more or less arbitrarily chosen and complicate the accurate treatment of long-range effects involved in binding. One typical force field scoring function in molecular docking is the scoring function of DOCK whose energy parameters are taken from the Amber force fields [17-19]. The scoring function is composed of two energy components of Lennard-Jones VDW and an electrostatic term

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

**Table 3:** Correlation coefficients between the experimentally determined binding energies and the calculated binding scores of 17 scoring functions for Wang et al.'s test [56] set of 100 complexes, [43].

Scoring Function	Function Type	Correlation (R)
ITScore/SE	K	0.65
ITScore	K	0.65
X-Score	E	0.64
DFIRE	K	0.63
Drug Score CSD	K	0.62
Drug Score PDB	K	0.6
Cerius2/PLP	E	0.56
SYBYL/G-Score	F	0.56
KScore	K	0.49
SYBYL/D-Score	F	0.48
SYBYL/Chem Score	E	0.47
Cerius2/PMF	K	0.4
DOCK/FF	F	0.4
Cerius2/LUDI	E	0.36
Cerius2/Lig Score	E	0.35
SYBYL/F-Score	E	0.3
Auto Dock	F	0.05

**Table 4:** Enrichments of nine scoring functions at the top 5% of the ranked data bases on four targets of ERα, MMP3, fXa, and AChE [42].

Scoring function	Function type	Enrichment at the top 5% (%)			
		ERα	MMP3	fXa	AChE
ITScore [41]	Iterative/knowledge-based	19.2	68.3	34.9	37
DOCK/FF [17]	Force-field-based	2.7	56.7	14	7.4
ICM-Score [28]	Empirical	38	36.7	29.5	0
ICM-PMF [28]	knowledge-based	9.6	20	19.4	1.9
SYBYL/F-Score [26]	Empirical	23.3	31.7	26.4	1.9
SYBYL/G-Score [25]	Force-field-based	0.7	31.7	31.8	11.1

SYBYL/ChemScore [33]	Empirical	0	73.3	23.3	9.3
SYBYL/PMF-Score [44]	Knowledge-based	0	5	21.7	0
SYBYL/D-Score [17]	Force-field-based	0	0	16.3	0
Maximum enrichments		39.2	88.3	43.7	97.5

<sup>a</sup>For each protein target, the constructed database includes known inhibitors (146 for ERα, 60 for MMP3, 129 for fXa, and 54 for AChE) and 999 random, diverse drug-like molecules served as a set of inactive compounds. <sup>b</sup> The last row lists the maximum theoretically possible enrichments at the top 5% of the ranked database, given the compositions of the databases including the active and inactive compounds.

Where  $r_{ij}$  stands for the distance between protein atom  $i$  and ligand atom  $j$ ,  $A_{ij}$  and  $B_{ij}$  are the VDW parameters, and  $q_i$  and  $q_j$  are the atomic charges. Here, the effect of solvent is implicitly considered by introducing a simple distance dependent dielectric constant  $\epsilon(r_{ij})$  in the Coulombic term. Despite the computational efficiency of the force field scoring function of DOCK, the distance-dependent dielectric factor cannot account for the desolvation effect, an important solvent effect that charged groups favor aqueous environments whereas non-polar groups tend to stay in non-aqueous environments. The desolvation energy is a many-body interaction term and depends on specific geometric and chemical surrounding environments of the considered solute atoms. If the desolvation effect is ignored, a scoring function would be biased on coulombic electrostatic interactions and therefore would tend to select highly charged ligands [16] (Table 5).

### Empirical scoring functions

The based on counting the number of various types of interactions between the two binding partners. Counting may be based on the number of ligand and receptor atoms in contact with each other or by calculating the change in solvent accessible surface area ( $\Delta SASA$ ) in the complex compared to the uncomplexed ligand and protein. The coefficients of the scoring function are usually fit using multiple linear regression methods. These interactions terms of the function may include for example: hydrophobic - hydrophobic contacts (favorable), Hydrophobic - hydrophilic contacts (unfavorable) (Accounts for unmet hydrogen bonds, which are an important enthalpic contribution to binding. One lost hydrogen bond can account for 1-2 orders of magnitude in binding affinity, Number of hydrogen bonds (favorable contribution to affinity, especially if shielded from solvent, if solvent exposed no contribution), Number of rotatable bonds immobilized in complex formation (unfavorable conformational entropy contribution).

These scoring functions are fit to reproduce experimental data, such as binding energies and/or conformations, as a sum of several parameterized functions, as first proposed by Böhm [20]. The design of empirical scoring functions is based on the idea that binding energies can be approximated by a sum of individual uncorrelated terms.

**Table 5:** Selected force field-based scoring functions.

	Protein-Ligand	Internal Ligand
SYBYLTM G-Score	$E_{vdW} + E_{H-bond} = \sum_{prot} \sum_{lig} \left[ \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + (E_{da} + E_{ww}) - (E_{dw} + E_{aw}) \right]$	$E_{vdW} + E_{torsion} = \sum_{lig} \left[ \left( \frac{C_{ij}}{d_{ij}^{12}} - \frac{D_{ij}}{d_{ij}^6} \right) + \sum_{lig} \frac{1}{2} V_l \left[ 1 + \frac{n}{ n } \right] \cos( n  \omega) - (E_{dw} + E_a) \right]$
SYBYLTM	$E_{vdW} + E_{electrostatic} = \sum_{prot} \sum_{lig} \left[ \left( \frac{A_{ij}}{d_{ij}^4} - \frac{B_{ij}}{d_{ij}^6} \right) + 332.0 \frac{q_i q_j}{\epsilon(d_{ij}) d_{ij}} \right]$	
Gold Score	$E_{vdW} + E_{electrostatic} = \sum_{prot} \sum_{lig} \left[ \left( \frac{A_{ij}}{d_{ij}^a} - \frac{B_{ij}}{d_{ij}^b} \right) + 332.0 \frac{q_i q_j}{\epsilon(d_{ij}) d_{ij}} \right]$	$E_{vdW} + E_{electrostatic} = \sum_{lig} \left[ \left( \frac{A_{ij}}{d_{ij}^a} - \frac{B_{ij}}{d_{ij}^b} \right) + 332.0 \frac{q_i q_j}{\epsilon(d_{ij}) d_{ij}} \right]$
Auto Dock v3.05	$E_{vdW} + E_{H-bond} + E_{electrostatic} = \sum_{prot} \sum_{lig} \left[ \left( \frac{A_{ij}}{d_{ij}^{12}} + \frac{B_{ij}}{d_{ij}^6} \right) + E(t) * \left( \frac{C_{ij}}{d_{ij}^{12}} + \frac{D_{ij}}{d_{ij}^{10}} \right) + 332.0 \frac{q_i q_j}{\epsilon(d_{ij}) d_{ij}} \right]$	$E_{vdW} + E_{H-bond} + E_{electrostatic} = \sum_{lig} \left[ \left( \frac{A_{ij}}{d_{ij}^{12}} + \frac{B_{ij}}{d_{ij}^6} \right) + E(t) * \left( \frac{C_{ij}}{d_{ij}^{12}} + \frac{D_{ij}}{d_{ij}^{10}} \right) + 332.0 \frac{q_i q_j}{\epsilon(d_{ij}) d_{ij}} \right]$
	where E(t)=angular weight factor	where E(t)=angular weight factor

The total  $\Delta G_{binding}$  is given by the sum of the terms corresponding to the interaction energy between the receptor and the ligand and the internal energy of the ligand (when available). Depending on the scoring function, these terms can be obtained by adding the following contributions: (a) electrostatic (i.e.  $E_{electrostatic}$ ); (b) vander Waals (i.e.  $E_{vw}$ ); (c) hydrogen bonding (i.e.  $E_{H-bond}$ ); and (d) torsional (i.e.  $E_{torsion}$ ). For two atoms i and j: (a)  $A_{ij}$  and  $B_{ij}$  correspond to van der Waals parameters for the given atom types; (b)  $d_{ij}$  corresponds to their interatomic distance; (c)  $q_i$  and  $q_j$  are the atomic partial charges; and (d)  $\epsilon(d_{ij})$  is a distance-dependent dielectric function. In G-Score, the hydrogen bonding term is a sum of the individual energies (i.e.  $E_{da}$ ,  $E_{ww}$ ,  $E_{dw}$ ,  $E_{aw}$ ) from all the donor acceptor pairs in the complex [59].

Second kind of scoring functions are empirical scoring functions, which estimate the binding affinity of a complex on the basis of a set of weighted energy terms

$$\Delta G = \sum_i W_i \cdot [\Delta G]_i$$

Where  $\Delta G_i$  represents different energy terms such as VDW energy, electrostatics, hydrogen bond, desolvation, entropy,

hydrophobicity, etc. The corresponding coefficients  $W_i$  are determined by fitting the binding affinity data of a training set of protein-ligand complexes with known three-dimensional structures [21-33]. Compared to the force field scoring functions, the empirical scoring functions are much faster in binding score calculations due to their simple energy terms [16]. (Table 6).

**Table 6:** Selected empirical scoring functions.

	Functional Form
LUDI	$\Delta G_{bind} = \Delta G_{H-bond} \sum_{H-bond} \phi(\Delta R, \Delta \alpha) + \Delta G_{ionic} \sum_{ionic} \phi(\Delta R, \Delta \alpha) + \Delta G_{hydrophobic} \sum_{hydrophobic}  A_{hydrophobic}  + \Delta G_{rotor} N_{rotor} + \Delta G_0$
SYBYLTM F-Score	$\Delta G_{bind} = \Delta G_{H-bond} \sum_{H-bond} \phi(\Delta R, \Delta \alpha) + \Delta G_{ionic} \sum_{ionic} \phi(\Delta R, \Delta \alpha) + \Delta G_{aromatic} \sum_{aromatic} \phi(\Delta A, \Delta \alpha) + \Delta G_{Contact} \sum_{contact} \phi(\Delta R) + \Delta G_{rotor} N_{rotor} + \Delta G_0$
SYBYLTM ChemScore	$\Delta G_{bind} = \Delta G_{H-bond} \sum_{H-bond} \phi(\Delta R, \Delta \alpha) + \Delta G_{metal} \sum_{metal} \phi(\Delta R, \Delta \alpha) + \Delta G_{lipo} \sum_{lipo} \phi(\Delta A, \Delta \alpha) + \Delta G_{rotor} \sum_{rotor} \phi(P_{nl}, P'_{nl}) + \Delta G_0$

The free energy of binding (i.e.  $\Delta G_{bind}$ ) is obtained by adding the contribution to the free energy of some terms that correspond to: (a) hydrogen bonding (in LUDI and SYBYLTM/F-Score the first two terms account for neutral and ionic hydrogen bonds, respectively); (b) hydrophobic or lipophilic (that accounts for the hydrophobic effect); (c) ligand rotational entropy (a term that counts all the rotatable single bonds in the ligand, which is supposed to be related with the torsional entropy loss of the ligand upon protein-ligand complexation.); (d) contact (that accounts for a general distance-dependent potential for protein-ligand atom contacts); and (e) metal (that accounts for metal ions residing inside the protein binding pocket). Thus, the different scoring functions differ in: (a) the number and the typology of the terms that contributes to  $\Delta G_{bind}$ ; and (b) the mathematical function (i.e.  $\phi$ ) used to calculate one specific contribution [where this function can depend on an angular ( $\Delta \alpha$ ) and/or a distance ( $\Delta R$ ) parameter/s that penalize/s the deviations from an ideal geometry].  $\Delta G_{H-bond}$ ,  $\Delta G_{ionic}$ ,  $\Delta G_{hydrophobic}$ ,  $\Delta G_{rotor}$ ,  $\Delta G_{aromatic}$ ,  $G_{contact}$ ,  $\Delta G_{metal}$ ,  $\Delta G_{lipo}$  are regression coefficients for each corresponding free energy term.  $\Delta G_0$  is a regression constant.  $A_{hydrophobic}$  corresponds to the molecular surface area [49].

### Knowledge-based scoring functions

The based on statistical observations of intermolecular close contacts in large 3D databases (such as the Cambridge Structural

Database or Protein Data Bank) which are used to derive “potentials of mean force”. This method is founded on the assumption that close intermolecular interactions between certain types of atoms



or functional groups that occur more frequently than one would expect by a random distribution are likely to be energetically favorable and therefore contribute favorably to binding affinity [34]. Knowledge-based scoring functions are based on knowing contact preferences and rely on the classical statistical physics idea that observed distributions of geometries can be used to deduce the potential that gives rise to the observed distribution. They were first proposed for studying protein folding [35] but they have also been used in protein ligand docking and to predict protein structures

[36] and protein-protein complexes [19,37]. In the protein-ligand docking field, they are used to score ligand binding poses by means of relatively simple atomic interaction-pair potentials that are built from statistical analyses of experimentally determined protein-ligand structures. These functions, then, are a direct consequence of the exponential number of protein-ligand complexes that have been deposited in the PDB database in recent years (20901 of the 27224 complexes that are now in the PDB were deposited after the first of January 2000) [38,39]. (Table 7).

**Table 7:** Selected Knowledge-Based Scoring Functions.

	Functional Form
PMF	<p>Parametrized pairwise potential PMF score:</p> $PMF = \sum_{prot} \sum_{lig} A_{ij}(d_{ij}) A_{ij}(d_{ij}) = -K_B T \ln \left[ f_{vol\_corr}^j(r) \frac{\rho_{seg}^{ij}(r)}{\rho_{bulk}^{ij}} \right]$ <p>where: (a) <math>k_B</math> is the Boltzmann constant; (b) <math>f_{vol\_corr}^j(r)</math> is a ligand volume correction factor; (c) and <math>(\rho_{seg}^{ij}(r))/(\rho_{bulk}^{ij})</math> indicates a radial distribution function for a protein atom <math>i</math> and a ligand atom <math>j</math></p>
Drug Score v1.2	$\Delta W = \sum_{prot} \sum_{lig} \Delta W(r) + (1 - \gamma) \left[ \sum_{lig} \Delta W_i(SAS, SAS_0) + \sum_{prot} \Delta W_i(SAS, SAS_0) \right]$ <p>where: (a) SAS correspond to the surface accessible area terms; (b) <math>W_{ij}</math> is a distance dependent pairwise potential; and (c) <math>\gamma</math> is an adjustable weight factor, normally set to 0.5</p>
where: (a) SAS correspond to the surface accessible area terms; (b) $W_{ij}$ is a distance dependent pairwise potential; and (c) $\gamma$ is an adjustable weight factor, normally set to 0.5	$G = \sum_{ij} -kT \log \left[ \frac{P_{ij}}{P} \right] \Delta_{ij}$

## Different Type of Software Docking

Docking is frequently used to predict the binding orientation of small molecule drug candidates to their protein targets in order to predict the affinity and activity of the small molecule. Hence docking plays an important role in the rational design of drugs. Two approaches are particularly popular within the molecular docking community. One approach uses a matching technique that describes the protein and the ligand as complementary surfaces. The second approach simulates the actual docking process in which the ligand-protein pairwise interaction energies are calculated. There are three types of Docking: Protein-Protein Docking, Protein-Ligand Docking and Protein-Protein & Protein-Ligand Docking. Softwares used in protein-protein docking include Affinity, Auto dock, Combibuild, Dock vision, Fred, Flexi dock, Flex-X, Glide and Gold. Some Software is discussed below:

### Dock

DOCK (Dedicator of cyto-kinesis) is a family of related proteins involved in intracellular signaling networks. Studies to date suggest that this family act as guanine nucleotide exchange factors for small G proteins of the Rho family, such as Rac and Cdc42. DOCK family proteins are categorized into four subfamilies based on their sequence homology: DOCK-A, DOCK-B, DOCK-C, DOCK-D

### Auto dock

Auto Dock is a suite of automatic docking tools. It has been predicted how small molecules, such as substrates or drug

candidates, stopping with a receptor known 3D structure. In addition to using them for docking, the atomic affinity grids can be visualized. This can help, for example, to guide organic synthetic chemists design better binders. Auto dock is current distribution includes two generations of software: Auto dock 4 and Auto dock Vina. Auto dock has now been distributed to more than 29000 users around the world. It is being used in academic, governmental, non-profit and commercial settings. In January of 2011, a search of the ISI Citation Index showed more than 2700 publications have cited the primary Auto Dock methods papers. Auto dock [40] uses the Lamarckian genetic algorithm, which allows favorable phenotypic characteristics to become inheritable. Auto dock was also used to identify inhibitors of RNA Editing Ligase-1 enzyme of T. brucei, the causative agent of human African trypanosomiasis [41].

### Flog

FLOG (Flexible Ligands Oriented on Grid), that searches a database of 3D coordinates to find molecules complementary to a macromolecular receptor of known 3D structure. The philosophy of FLOG is similar to that reported for dock [42]. In common with that system, we use a match center representation of the volume of the binding cavity. This new algorithm includes the concept of 'essential points', match centers that must be paired with a ligand atom. Also, we introduce the use of a rapid simplex-based rigid-body optimizer to refine the orientations. We demonstrate, using dihydrofolate reductase as a sample receptor, that the FLOG system can select known inhibitors from a large database of drug-like

compounds [43]. Fully automated “anchor and grow” methods have been implemented in several methods such as flog.

### Gold

Genetic Optimization for Ligand Docking (GOLD) [44] explores full ligand flexibility with partial target flexibility using a genetic algorithm. The GOLD algorithm optimizes rotatable dihedrals and ligand-target hydrogen bonds. The fitness of a generation is evaluated based on a maximization of intermolecular hydrogen bonds. The fitness function is the sum of a hydrogen bonding term, a term for steric energy interaction between the protein and the ligand and a Lennard-Jones potential for internal energy of ligand.

Genetic Optimization and Ligand Docking, uses multiple subpopulations of ligand. Force-field based scoring function includes three terms: H-bonding term, intermolecular dispersion potential, and intermolecular potential. It is 71% success in identifying experimental binding mode in 100 protein complexes.

### Auto dock vina

Auto Dock Vina is a new generation of docking software from the Molecular Graphics Lab. It achieves significant improvements in the average accuracy of the binding mode predictions, while also being up to two orders of magnitude faster than Auto Dock 4.

Auto Dock 4.2 is faster than earlier versions, and it allows side chains in the macromolecule to be flexible. As before, rigid docking is blindly fast, and high-quality flexible docking can be done in around a minute. Up to 40,000 rigid dockings can be done in a day on one cpu.

Auto Dock 4.2 now has a free-energy scoring function that is based on a linear regression analysis, the AMBER force field, and an even larger set of diverse protein-ligand complexes with known

inhibition constants than we used in Auto Dock 3.0. The best model was cross-validated with a separate set of HIV-1 protease complexes, and confirmed that the standard error is around 2.5kcal/mol. This is enough to discriminate between leads with milli-, micro- and nano-molar inhibition constants.

### Beta dock

Beta Dock is molecular docking simulation software based on the theory of Beta-complex. We claim that Beta Dock is superior to any contemporary docking software while it requires less human intervention. Current version of Beta Dock is also an intermediate deliverable of the on-going research. Prioritizing shape complementarity, based on the theory of b-complex and the Voronoi diagram. Rigid bodies (both receptor and ligand) [45].

### Flex X

Flex X is a software package to predict protein-ligand interactions. For a protein with known three-dimensional structure and a small ligand molecule, Flex X accurately predicts the geometry of the protein-ligand complex within a few seconds. We docked a set of known active compounds with standard Flex X and derived three sets of target-specific receptor-based pharmacophore constraints by statistical analysis of the predicted placements. Applying these receptor-based constraints in a virtual screening protocol utilizing Flex X-Pharm led to significantly improved enrichments.

Base fragment is picked up and docked using “pose-clustering” algorithm. Clustering algorithm is implemented to merge similar ligand transformations into active site. Flexible fragments are added incrementally using “MIMUMBA” and evaluated using overlap function, followed by energy calculations till the ligand is completely built. Final evaluation through Bohm’s scoring function that includes H-bonds, ionic, aromatic and lipophilic terms.

**Table 8**

Program	Year Published	Organisation	Description
Dock	1988	University of California-San Francisco	This program was first discovered. This program are based on Geometric Matching Algorithm
Auto Dock	1990	The Scripps Research Institute	Automated docking of ligand to macromolecule by Lamarckian Genetic Algorithm and Empirical Free Energy Scoring Function
Dock vision	1992	Dockvision	Base on Monte Carlo, genetic algorithm, and database screening docking algorithm
Divali	1995	University of California, San Francisco	Base on AMBER type potential function and genetic algorithm
Flog	1994	Merck Research Laboratories	Rigid body docking programme using database of pre-generated conformations
Gold	1995	Collaboration between the University of Sheffield, GlaxoSmithKline plc and CCDC	Genetic algorithm based, flexible ligand, partial flexibility for protein
Blaster	2009	University of California-San Francisco	Combines ZINC database with DOCK to find ligand for target protein
Auto dock Vina	2010	The Scripps Research Institute	New generation of Auto dock
Beta Dock	2011	Hanyag University	Based on Vroni Diagram

Flex Pep Dock	2010	The Hebrew University	Modeling of peptide-protein complex, implemented within the Rosetta framework
GPCR auto model	2012	INRA	Automates the homology modeling of mammalian olfactory receptors (ORs) based on the six three-dimensional (3D) structures of G protein-coupled receptors (GPCRs) available so far and performs the docking of odorants on these models
FlexX	2015	University of Sherbrooke	Target side-chain flexibility and soft scoring function, based on surface complementarity
Mols 2.0	2016	University of Madras	Rigid protein-small-molecule docking, Flexible protein-peptide docking

A number of protein-ligand docking programs currently available is high and has been steadily increasing over the last decades. The following list presents an overview of the most common programs, listed alphabetically, with indication of the

corresponding year of publication, involved organisation or institution, short description, availability of a web service and the license. This table is comprehensive but not complete (Table 8).

### Comparison Studies of Docking Software (Table 9)

**Table 9:** <sup>a</sup>Top performing programs/scoring functions indicated by bold font.

Programs/Functions Compared	Test Set	Pose Prediction	Affinity Prediction/VS Performance	Performance Measure	Refs.
DOCK, FlexX, Glide, ICM, PhDOCK, Surflex, Surflex Ringflex	Astex with additional kinase and nuclear receptor target	Y	N	RMSD, GARD	162
DOCK, FlexX, Glide, GOLD, Surflex, Chem Score, D-Score, F-Score, G-Score, Glide Score, Gold Score, Grid-Score, PMF-Score, X-Score, Surflex-Score	Crystal structure of dihydropteroate synthase (DHPS). Ligand decoy sets: Schrodinger, ZINC, and ACD, seeded with 65 known actives	Y	Y	RMSD, enrichment factors, and ROC curves	88
Kang's GA, GOLD, Glide, Surflex, DOCK6	GOLD data set of 134 crystal structures	Y	N	RMSD	59
DOCK, FlexX, Glide, ICM, PhDOCK, Surflex	A set of 68 diverse high resolution crystal structures; DUD	Y	Y	RMSD, ROC factors, and ROC AUCs	47
Auto Dock, Glide, FlexX, Surflex	Crystal structure of hDHFR with carboranyl ligands	Y	Y	RMSD	179
ASP, Chem Score, D-Score, Drug Score, F-Score, G-Score, Glide Score, Gold Score, Jain, Lig Score, LUDI, PLP, PMF, PMF-Score, X-Score	A set of 195 diverse high resolution crystal structures; binding constants from PDBbind database	Y	Y	RMSD, binding affinity	85
RMSD, binding affinity	Trypsin, rhinovirus, HIV protease, carboxypeptidase, estrogen receptor- $\alpha$ ; Maybridge ligand database	Y	Y	RMSD, enrichment Factors	91
Chem Score, Dock Score, Drug Score, eHiTS, FlexX, Glide, GOLD, Hammerhead, Lig Score, PLP, PMF, Rank Score (FITTED), Surflex, X-Score	58 complexes (HIV protease, thrombin, trypsin, and matrix metalloproteases)	N	Y	58 complexes (HIV protease, thrombin, trypsin, and matrix metalloproteases)	71
QM/MM combinations of: AMBER, Gold Score, Chem Score; HF/6-31G*, AM1d, PM3	Six HIV protease-inhibitor Complexes	Y	N	RMSD	66

### Conclusion

Docking methodologies were developing day after day gets more better result for drug designing. The scoring function is one of the most important components in structure-based drug design. Despite considerable success, accurate and rapid prediction of protein-ligand interactions is still a challenge in molecular docking. A number of protein-ligand docking programs currently available

is high and has been steadily increasing over the last decades and many companies producing newer software to overcome the challenges observed in old one [46-69].

### References

1. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 3(11): 935-949.

2. Vieth M, Hirst JD, Dominy BN, Dailer H, Brooks CL III (1998) Assessing search strategies for flexible docking. *J Comp Chem* 19(14): 1623-1631.
3. Blaney JM, Dixon JS (1993) A good ligand is hard to find: Automated docking methods. *Perspectives in Drug Discovery and Design* 1(2): 301-319.
4. Kuntz ID, Meng EC, Shoichet BK (1994) Structure-Based Molecular Design. *Acc Chem Res* 27(5): 117-123.
5. Gschwend DA, Good AC, Kuntz ID (1996) Molecular Docking Towards Drug Discovery. *J Mol Recog* 9(2): 175-186.
6. Vieth M, Hirst JD, Kolinski A, Brooks CL III (1998) Assessing energy functions for flexible docking. *J Comp Chem* 19: 1612-1622.
7. Bissantz C, Folkers G, Rognan D (2000) Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J Med Chem* 43(25): 4759-4767.
8. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, et al. (2006) A critical assessment of docking programs and scoring functions. *J Med Chem* 49(20): 5912-5931.
9. Stahl M, Rarey M (2001) Detailed analysis of scoring functions for virtual screening. *J Med Chem* 44(7): 1035-1042.
10. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Rev Drug Discov* 3(11): 935-949.
11. Cummings MD, Desjarlais RL, Gibbs AC, Mohan V, Jaeger EP (2005) Comparison of automated docking programs as virtual screening tools. *J Med Chem* 48(4): 962-976.
12. Bursulaya BD, Totrov M, Abagyan R, Brooks CL (2003) Comparative study of several algorithms for flexible ligand docking. *J Comput Aided Mol Des* 17(11): 755-763.
13. Perola E, Walters WP, Charifson, PS (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* 56(2): 235-249.
14. Kellenberger E, Rodrigo J, Muller P, Rognan D (2004) Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* 57(2): 225-242.
15. Sheng YH, Sam ZG, Xiaoqin Z (2010) Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys Chem Chem Phys* 12(40): 12899-12908.
16. Meng EC, Shoichet KB, Kuntz DI (1992) Automated docking with grid-based energy Evaluation. *J Comput Chem* 13(4): 505-524.
17. Weiner JS, Kollman AP, Case AD (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *J Am Chem Soc* 106(3): 765-784.
18. Weiner JS, Kollman AP, Nguyen TD, Case AD (1986) An all atom force field for simulations of proteins and nucleic acids. *J Comput Chem* 7(2): 230-252.
19. Muegge I (2006) PMF scoring revisited. *J Med Chem* 49(20): 5895-5902.
20. Bohm JH (1994) 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* 8(3): 243-256.
21. Bohm JH (1998) 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* 12(4): 309-323.
22. Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, et al. (1995) Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem Biol* 2(5): 317-324.
23. Gehlhaar DK, Bouzida D, Rejto PA (1999) Reduced Dimensionality in Ligand-Protein Structure Prediction: Covalent Inhibitors of Serine Proteases and Design of Site-Directed Combinatorial Libraries. *Rational Drug Design Chapter ACS Symposium Series* 19: 292-311.
24. Eldridge MD, Murray CW, Auton TR, Paolini GV, MeeN RP (1997) 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* 11(5): 425-445.
25. Wang R, Liu L, Lai L, Tang Y (1998) A New Empirical Method for Estimating the Binding Affinity of a Protein-Ligand Complex. *J Mol Model* 4(12): 379-394.
26. Wang R, Lai L, Wang S (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des* 16(1): 11-26.
27. Yin S, Biedermannova L, Vondrasek J, Dokholyan NV (2008) MedusaScore: An Accurate Force Field-Based Scoring Function for Virtual Drug Screening. *J Chem Inf Model* 48(8): 1656-1662.
28. Raub S, Steffen A, Kamper A, Marian CM (2008) AIScore-Chemically diverse empirical scoring function employing quantum chemical binding energies of hydrogen-bonded complex. *J Chem Inf Model* 48(7): 1492-1510.
29. Sottriffer CA, Sanschagrin P, Matter H, Klebe G (2008) SFCscore: Scoring functions for affinity prediction of protein-ligand complex. *Proteins* 73(2): 395-419.
30. Huang SY, Zou X (2006) An iterative knowledge-based scoring function to predict protein-ligand interactions: I. Derivation of interaction potentials. *J Comput Chem* 27(15): 1865-1875.
31. Jain AN (1996) Scoring noncovalent protein-ligand interactions: A continuous differentiable function tuned to compute binding affinities. *J Comput Aided Mol Des* 10(5): 427-440.
32. Head DR, Smythe ML, Oprea IT, Waller LC, Green MS, et al. (1996) VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. *J Am Chem Soc* 118(16): 3959-3969.
33. Lange G, Lesuisse D, Deprez P, Schoot B, Loenze P, et al. (2003) Requirements for specific binding of low affinity inhibitor fragments to the SH2 domain of (pp60)Src are identical to those for high affinity binding of full length inhibitors. *J Med Chem* 46(24): 5184-5195.
34. Miyazawa S, Jernigan LR (1985) Estimation of effective inter residue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* 18(3): 534-552.
35. Lu H, Skolnick J (2003) Application of statistical potentials to protein structure refinement from low resolution ab initio models. *J Biopolymers* 70(4): 575-584.
36. Lu H, Lu L, Skolnick J (2003) Development of unified statistical potentials describing protein-protein interactions. *J Biophys* 84: 1895-1901.
37. Kortemme T, Morozov AV, Baker DJ (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J Mol Biol* 326: 1239-1259.
38. Yang YC, Wang R, Wang S (2006) M-Score: A Knowledge-Based Potential Scoring Function Accounting for Protein Atom Mobility. *J Med Chem* 49(20): 5903-5911.
39. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, et al. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comp Chem* 19(14): 1639-1662.
40. Durrant JD, Hall L, Swift RV, Landon M, Schnauffer A, et al. (2010) Novel naphthalene-based inhibitors of Trypanosoma brucei RNA editing ligase 1. *PLoS Negl Trop Dis* 4(8): e803.
41. Shoichet BK, Kuntz ID, Bodian DL (1992) Molecular Docking Using





- Shape Descriptors. *J Comput Chem* 13(3): 380-397.
42. Miller MD, Kearsley SK, Underwood DJ, Sheridan RP (1994) FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. *J Comput Aided Mol Des* 8(2): 153-174.
43. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 267(3): 727-748.
44. Kim DS, Kim CM, Won CI, Kim JK, Ryu J, et al. (2011) BetaDock: shape-priority docking method based on beta-complex. *J Biomol Struct Dyn* 29: 219-242.
45. Shoichet KB, Leach RA, Kuntz ID (1999) Ligand solvation in molecular docking. *Proteins* 34(1): 4-16.
46. Wei BQ, Baase WA, Weaver LH, Matthews BW, Shoichet B (2002) A model binding site for testing scoring functions in molecular docking. *J Mol Biol* 322(2): 339-355.
47. Zou X, Sun, Kuntz ID (1999) Inclusion of Solvation in Ligand Binding Free Energy Calculations Using the Generalized-Born Model. *J Am Chem Soc* 121(35): 8033-8043.
48. Liu HY, Kuntz ID, Zou X (2004) Pairwise GB/SA Scoring Function for Structure-based Drug Design. *J Phys Chem*. 108(17): 5453-5462.
49. Liu HY, Zou X (2006) Electrostatics of Ligand Binding: Parametrization of the Generalized Born Model and Comparison with the Poisson-Boltzmann Approach. *J Phys Chem B* 110(18): 9304-9313.
50. Morris MG, Goodsell SD, Huey R, Olson JA (2007) A semiempirical free energy force field with charge-based desolvation. *J Comput Chem* 28(6): 1145-1152.
51. Rarey M, Kramer B, Lengauer T, Klebe G (1996) A Fast Flexible Docking Method using an Incremental Construction Algorithm. *J Mol Biol* 261(3): 470-489.
52. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, et al. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 47(7): 1739-1749.
53. Abagyan R, Totrov M, Kuznetsov D (1994) ICM-A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J Comput Chem* 15(5): 488-506.
54. Jain NA (2003) Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine. *J Med Chem* 46(4): 499-511.
55. Virtual Biosphere and Materials. BIOVIA Cerius2, version 4.6; Accelrys Inc.
56. Huang SY, Zou X (2006) An iterative knowledge-based scoring function to predict protein-ligand interactions: II. Validation of the scoring function. *J Comput Chem* 27(15): 1876-1882.
57. Huang SY, Zou X (2010) Inclusion of Solvation and Entropy in the Knowledge-Based Scoring Function for Protein-Ligand Interactions. *J Chem Inf Model* 50(2): 262-273.
58. Muegge I, Martin CY (1999) A General and Fast Scoring Function for Protein-Ligand Interactions: A Simplified Potential Approach. *J Med Chem* 42(5): 791-804.
59. Gohlke H, Hendlich M, Klebe G (2000) Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* 295(2): 337-356.
60. Veale HFG, Gohlke H, Klebe G (2005) Drug Score<sup>CSD</sup> -knowledge-based scoring function derived from small molecule crystal data with superior. *J Med Chem* 48(20): 6296-6303.
61. Zhang C, Liu S, Zhu Q, and Zhou Y (2005) A Knowledge-Based Energy Function for Protein-Ligand, Protein-Protein, and Protein-DNA Complexes. *J Med Chem* 48(7): 2325-2335.
62. DeWitte SR, Shakhnovich IE (1996) SMOG: de Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence. *J Am Chem Soc* 118(47): 11733-11744.
63. Ishchenko VA, Shakhnovich IE (2002) Small Molecule Growth 2001 (SMoG2001): An improved knowledge-based scoring function for protein-ligand interactions. *J Med Chem* 45(13): 2770-2780.
64. Mitchell OBJ, Laskowski AR, Alex A, Thornton MJ (1999) BLEEP-potential of mean force describing protein-ligand interactions: I. Generating potential. *J Comput Chem* 20(11): 1165-1176.
65. Mitchell OBJ, Laskowski AR, Alex A, Forster JM, Thornton MJ (1999) BLEEP-potential of mean force describing protein-ligand interactions: II. Calculation of binding energies and comparison with experimental data. *J Comput Chem* 20(11): 1177-1185.
66. Mooij TW, Verdonk LM (2005) General and targeted statistical potentials for protein-ligand interactions. *Proteins* 61(2): 272-287.
67. Zhao X, Liu X, Wang Y, Chen Z, Kang L, et al. (2008) An improved PMF scoring function for universally predicting the interactions of a ligand with protein, DNA, and RNA. *J Chem Inf Model* 48(7): 1438-1447.
68. Wang R, Lu Y, Wang S (2003) Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *J Med Chem* 46 (12): 2287-2303.
69. Montserrat Vaqué, Anna Ardévol, Cinta Bladé, M Josepa Salvadó, Mayte Blay, et al. (2008) Protein-ligand Docking: A Review of Recent Advances and Future Perspectives. *Current Pharmaceutical Analysis* 4(1): 1-19.