MANTECH Publications

Journal of Pharmaceutical Analysis and Drug Research
Volume 4, Issue 2, May-August, 2022

An Overview of the Ways in which Molecular Docking can be Used in the Drug Design Process- Review

Sarvesh Kumar[1], Suryakant Prasad[2]

Associate Professor[1], Student[2]

Department of Pharmacy

Magadh Institute of Pharmacy College

Corresponding Author's Email: - *suryakantprasad35@gmail.com[2]*

Abstract

The computer modelling of the structures of complexes generated by two or more interacting molecules is known as molecular docking. Docking is a method in molecular modelling that predicts the preferred orientation of one molecule to another when they are linked together to create a stable complex. Using scoring functions, knowledge of the preferred orientation may be used to predict the strength of the connection or binding affinity between two molecules. Because of its capacity to anticipate the binding-conformation of small molecules ligands to the proper target binding site, molecular docking is one of the most often utilised techniques in structure-based drug design.

Keywords: *Molecular Modelling, Scoring Functions*

INTRODUCTION

The known three-dimensional structures can be depicted to provide various viewpoints of the structures. It is feasible to superimpose one structure on another using complicated molecular mechanics algorithms. The same method is used to superimpose a prospective drug's three-dimensional structure on its probable target location. Docking is the name given to this operation, which is frequently automated. Ligand is a tiny molecule that generally interacts with protein binding sites. Binding sites are protein regions that are known to be involved in the formation of complexes with ligands. There are numerous mutual conformations that might result in binding. These are

frequently referred to as "binding modes. It also forecasts the intensity of the binding, the complex's energy, the sort of signal produced, and the binding affinity between two molecules using scoring functions. The most intriguing situation is the type of protein-ligand interaction, which has medical significance. Signal transduction is heavily reliant on the interactions of physiologically relevant components such as proteins, nucleic acids, carbohydrates, and lipids. Additionally, the relative directions of the two interacting partners may influence the sort of signal generated. As a result, docking may be used to forecast both the strength and kind of signal produced.

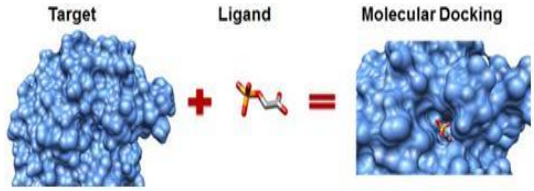


Figure 1: Protein and ligand docked complex

BASIC CONCEPT

The creation of a protein-ligand complex is known as docking. The aim is to predict the structure of the resultant complex, given the structures of a ligand and a protein. This is known as the docking problem. Docking is the energy-optimizing issue because the native shape of a complex may be considered to reflect the global minimum of binding free energy.

As a result, heuristic approximations are frequently required to make the problem tractable in a reasonable amount of time. Docking method development is thus concerned with making the correct assumptions and identifying acceptable simplifications that still provide a sufficiently accurate and predictive model for protein-ligand interactions.

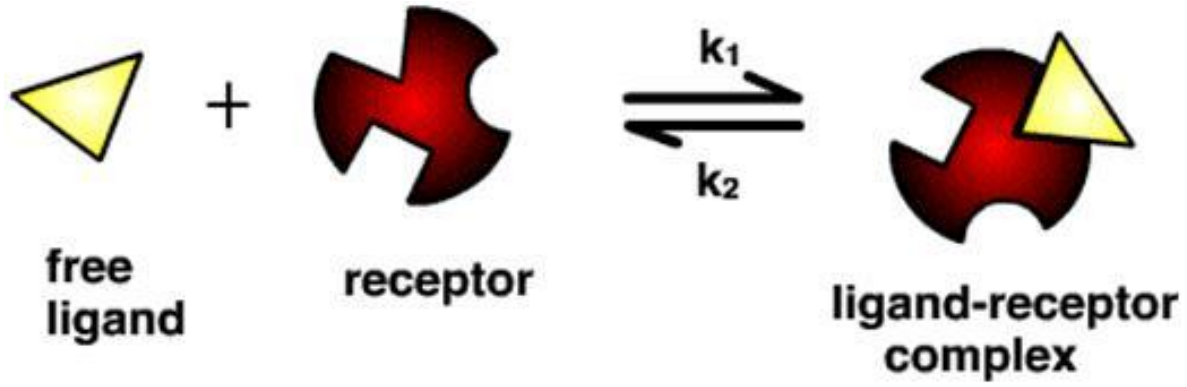


Figure 2: Protein ligand complex

MOLECULAR DOCKING APPROCHES

Within the molecular docking field, two techniques are particularly prominent. One method employs a matching strategy in which the protein and ligand are described as having complementary surfaces. The second method replicates the real docking process by calculating the ligand-protein pairwise interaction energies. Both techniques offer substantial advantages.

Shape complementarity

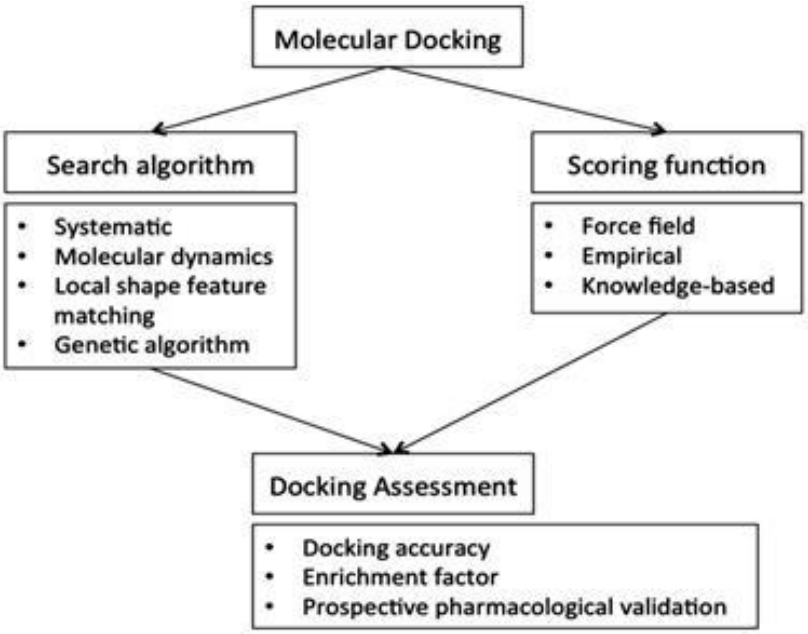
In this example, the molecular surface of the receptor is defined in terms of its solvent-accessible surface area, and the molecular surface of the ligand is characterised in terms of its corresponding surface description. The complementarity of the two surfaces is a shape matching description that may aid in determining the complementary posture of docking the target and ligand molecules. Shape complementarity approaches can scan through thousands of ligands in seconds and determine if they can bind at the protein's active site, and they are typically scalable to even protein-protein interactions. Because they employ geometric representations of the ligands to determine optimum binding, they are also considerably more suited to pharmacophore-based techniques.

Simulation

Simulating the docking procedure as such is even more difficult. In this strategy, the protein and the ligand are separated by some physical distance, and the ligand makes its way into the protein's active site after a specific number of "moves" in its conformational space. The motions include rigid body transformations like translations and rotations, as well as internal ligand structure modifications like torsion angle rotations. Each of these ligand conformational changes results in a total energy cost for the system. As a result, the total energy of the system is estimated after each step. The apparent benefit of docking simulation is that ligand flexibility may be easily integrated, whereas form complementarity approaches need innovative methods to incorporate ligand flexibility. Furthermore, it more properly represents reality, whereas complementary approaches are more abstract. Simulating a large energy landscape is obviously computationally costly. Docking simulation has become more realistic thanks to grid-based methodologies, optimization methods, and greater computer performance.

MECHANICS OF DOCKING

The success of a docking program depends on two components: the search algorithm and the scoring function.



SEARCH ALGORITHM

The search space in theory consists of all possible orientations and conformations of the protein paired with the ligand. However, in practice with current computational resources, it is impossible to exhaustively explore the search space this would involve enumerating all possible distortions of each molecule (molecules are dynamic and exist in an ensemble of conformational states) and all possible rotational and translational orientations of the ligand relative to the protein at a given level of granularity. Most docking programs in use account for the whole conformational space of the ligand (flexible ligand), and several attempt to model a flexible protein receptor. Each "snapshot" of the pair is referred to as a pose.

A variety of conformational search strategies have been applied to the ligand and to the receptor. These include:

- Systematic or stochastic torsional searches about rotatable bonds
- Molecular dynamics simulations
- Genetic algorithms to "evolve" new low energy conformations and where the score of each pose acts as the fitness function used to select individuals for the next iteration

The algorithm should create an optimum number of configurations that admit by experimentation method determining binding modes. The following are the various algorithms applied for docking analysis such as Point complementary, Monte Carlo, Fragment-based, Genetic algorithms, Systematic searches, Distance geometry etc.

Molecular dynamics (MD) simulation

In this approach, proteins are typically held rigid, and the ligand is allowed to freely explore their conformational space. The generated conformations are then docked successively into the protein, and an MD simulation consisting of a simulated annealing protocol is performed. This is usually supplemented with short MD energy minimization steps, and the energies determined from the MD runs are used for ranking the overall scoring. Although this is a computer-expensive method (involving potentially hundreds of MD runs), it has some advantages: for example, no specialized energy/scoring functions are required. MD force fields can typically be used to find poses that are reasonable and can be compared with experimental structures.The Distance Constrained Essential Dynamics method (DCED) has been used to generate multiple structures for docking, called eigen structures. This approach, although avoiding most of the costly MD calculations, can capture the essential motions involved in a flexible receptor, representing a form of coarse-grained dynamics.

Shape – complementarity methods

The most common technique used in many docking programs, shape-complementarity methods focus on the match between the receptor and the ligand in order to find an optimal pose. Programs include DOCK, FRED,[4] GLIDE,[5] SURFLEX, eHiTS and many more. Most methods describe the molecules in terms of a finite number of descriptors that include structural complementarity and binding complementarity. Structural complementarity is mostly a geometric description of the molecules, including solvent- accessible surface area, overall shape and geometric constraints between atoms in the protein and ligand. Binding complementarity takes into account features like hydrogen bonding interactions, hydrophobic contacts and van der Waals interactions to describe how well a particular ligand will bind to the protein. Both kinds of descriptors are conveniently represented in the form of structural templates which are then used to quickly match potential compounds (either

from a database or from the user-given inputs) that will bind well at the active site of the protein. Compared to the all-atom molecular dynamics approaches, these methods are very efficient in finding optimal binding poses for the protein and ligand.

Genetics algorithm

Two of the most used docking programs belong to this class: GOLD and AutoDock Genetic algorithms allow the exploration of a large conformational space – which is basically spanned by the protein and ligand jointly in this case – by representing each spatial arrangement of the pair as a "gene" with a particular energy. The entire genome thus represents the complete energy landscape which is to be explored. The simulation of the evolution of the genome is carried out by cross-over techniques similar to biological evolution, where random pairs of individuals (conformations) are "mated" with the possibility for a random mutation in the offspring. These methods have proven very useful in sampling the vast state-space while maintaining closeness to the actual process involved.

Although genetic algorithms are quite successful in sampling the large conformational space, many docking programs require the protein to remain fixed, while allowing only the ligand to flex and adjust to the active site of the protein. Genetic algorithms also require multiple runs to obtain reliable answers regarding ligands that may bind to the protein. The time it takes to typically run a genetic algorithm in order to allow a proper pose may be longer, hence these methods may not be as efficient as shape complementarity-based approaches in screening large databases of compounds. Recent improvements in using grid-based evaluation of energies, limiting the exploration of the conformational changes at only local areas (active sites) of interest, and improved tabling methods have significantly enhanced the performance of genetic algorithms and made them suitable for virtual screening applications.

SCORING FUNCTION

In the fields of computational chemistry and molecular modelling, scoring functions are fast approximate mathematical methods used to predict the strength of the non-covalent interaction (also referred to as binding affinity) between two molecules after they have been docked. Most commonly one of the molecules is a small organic compound such as a drug and the second is the drug's biological target such as a protein receptor.

Scoring functions have also been developed to predict the strength of other types of intermolecular interactions, for example between two proteins or between protein and DNA.

Force field

Affinities are estimated by summing the strength of intermolecular van der Waals and electrostatic interactions between all atoms of the two molecules in the complex using a force field. The intramolecular energies (also referred to as strain energy) of the two binding partners are also frequently included. Finally since the binding normally takes place in the presence of water, the desolvation energies of the ligand and of the protein are sometimes taken into account using implicit salvation methods such as GBSA or PBSA.

Empirical

Based on counting the number of various types of interactions between the two binding partners.[6] 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 (ΔSASA) 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 (unfavourable conformational entropy contribution).

Knowledge-based (also known as statistical potentials) – 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 favourable and therefore contribute favourably to binding affinity.

Machine-learning

Unlike these classical scoring functions, machine-learning scoring functions are characterized by not assuming a predetermined functional form for the relationship between binding affinity and the structural features describing the protein-ligand complex. In this way, the functional form is inferred directly from the data.

Machine-learning scoring functions have consistently been found to outperform classical scoring functions at binding affinity prediction of diverse protein-ligand complexes. This has also been the case for target-specific complexes, although the advantage is target-dependent and mainly depends on the volume of relevant data available. When appropriate care is taken, machine-learning scoring functions perform at least as well as classical scoring functions at the related problem of structure-based virtual screening.

MAJOR STEPS INVOLVED IN MECHANICS OF DOCKING

Molecular Docking is the process in which the intermolecular interaction between 2 molecules was studied in In-silico. In this process, the Macromolecule is the protein receptor. The micro molecule is the Ligand molecule which can be acted as an inhibitor. So, the Docking process involves the following steps:

Step I –

Preparation of protein: Three dimensional structure of the Protein should be retrieved from Protein data bank (PDB); afterward the retrieved structure should be pre-processed. This should admit removal of the water molecules from the cavity, stabilizing the charges, filling the missing residues, generation the side chains etc. according to the parameters available.

Step II –

Active site prediction: After the preparation of protein, the active site of protein should be predicted. The receptor might possess lots of active sites merely the one of the concern should be picked out. Mostly the water molecules and hetero atoms are removed if present.

Step III –

Preparation of ligand: Ligands can be retrieved from several databases such as ZINC, Pub Chem or can be sketched applying Chem sketch tool. While picking out the ligand, the LIPINSKY'S RULE OF 5 should be utilized. Lipinski rule of 5 assists in discerning amongst non-drug like and drug like candidates. It promises high chance of success or failure due to drug likeness for molecules abiding by with 2 or more than of the complying rules. For choice of a ligand allowing to the

LIPINSKY'S RULE:

Step IV-

Docking: Ligand is docked against the protein and the interactions are analyzed. The scoring function gives score on the basis of the best docked ligand basis is picked out.

TYPES OF DOCKING

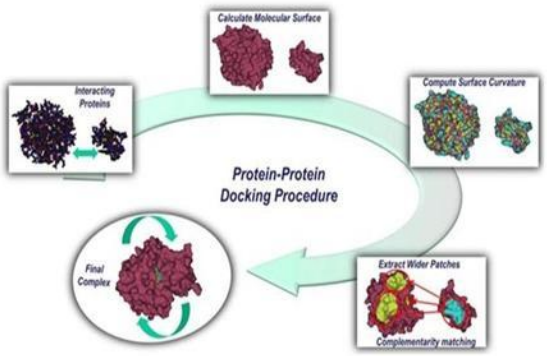
Molecular docking is the technique that is used to study molecular binding and how molecules bind. The term "docking" is mostly related to protein molecule interactions. There are several types of molecular docking for protein interactions:

Advantages:

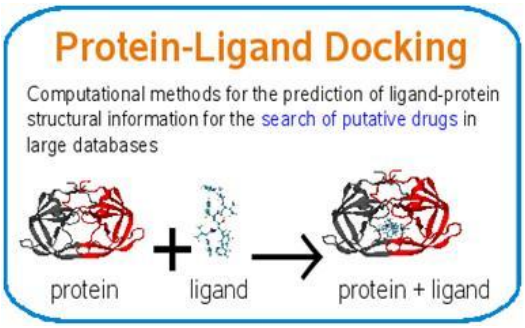
One of the major advantages conferred by docking is that it allows researchers to quickly screen large databases of potential drugs which would otherwise require tedious and prolonged work in the lab using traditional drug discovery procedures.

Protein- protein docking interaction

Protein-protein interactions occur between two proteins that are similar in size. The interface between the two molecules tends to be flatter and smoother than those in protein-ligand interactions. Protein-protein interactions are usually more rigid; the interfaces of these interactions do not have the ability to alter their conformation in order to improve binding and ease movement. Conformational changes are limited by steric constraints and thus are said to be rigid.



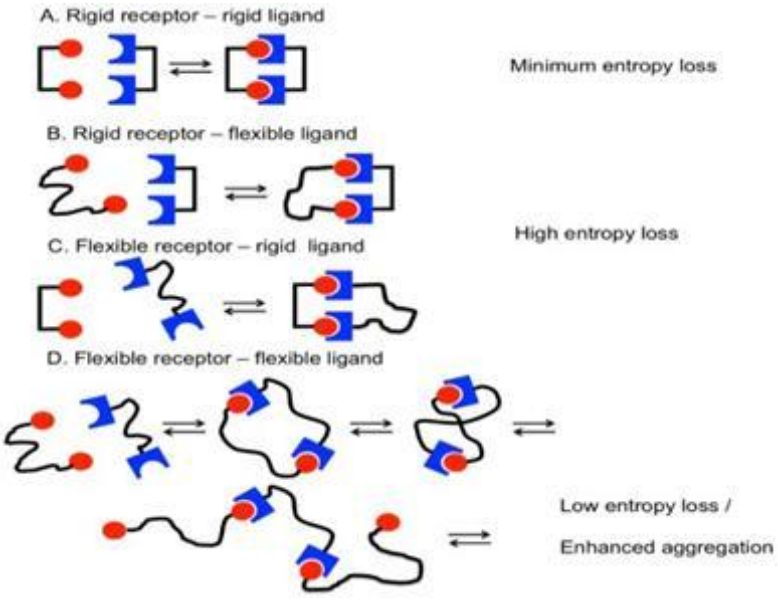
Protein receptor - ligand docking



Also known as the molecular docking technique, protein receptor -ligand docking is used to check the structure, position and orientation of a protein when it interacts with small molecules like ligands. Protein receptor-ligand motifs fit together tightly, and are often referred to as a lock and key mechanism. There is both high specificity and induced fit within these interfaces with specificity increasing with rigidity. Protein receptor-ligand can either have a rigid ligand and a flexible receptor, or a flexible ligand with a rigid receptor.

Rigid ligand with a flexible receptor

The native structure of the rigid ligand flexible receptor often maximizes the interface area between the molecules. They move within respect to one another in a perpendicular direction in respect to the interface. This allows for binding of a receptor with a larger than usual ligand. Normally when there is ligand overlap in the docking interface, energy penalties incur. If the van der Waals forces can be decreased, energy loss in the system will be minimized. This can be accomplished by allowing flexibility in the receptor. Flexible receptors allow for docking of a larger ligand than would be allowed for with a rigid receptor.



Flexible ligand with a rigid receptor

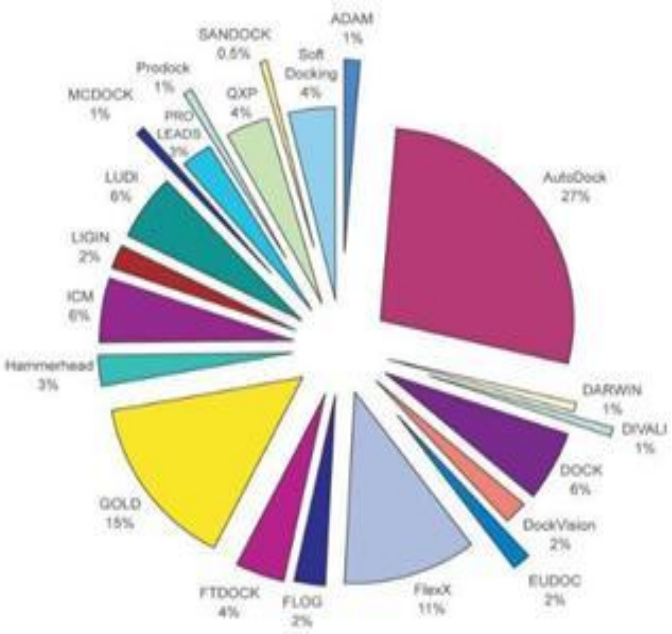
When the ligand-receptor fit does not need to be produced, the receptor can keep its stiffness while retaining the system's free energy. For effective docking, the ligand's properties must be consistent, and the ligand must be somewhat smaller in size than the receptor interface. However, no docking is totally stiff; there is inherent mobility that allows for little conformational modification for ligand binding. When the six degrees of freedom for protein mobility (three rotational and three translational) are considered, the amount of intrinsic flexibility offered by the receptor is significantly larger. This compensates for any energy difference between the receptor and the ligand, allowing for a simpler, more energetically advantageous interaction between the two.

APPLICATION

- A binding interaction between a small molecule ligand and an enzyme protein may result in activation or inhibition of the enzyme. If the protein is a receptor, ligand binding may result in agonism or antagonism.
- Docking is most commonly used in the field of drug design— most drugs are small organicmolecules, and docking may be applied to hit identification – docking combined with a scoring function can be used to quickly screen large databases of potential drugs in silicoto identify molecules that are likely to bind to protein target of interest
- lead optimization – docking can be used to predict in where and in which relative orientation a ligand binds to a protein (also referred to as the binding mode or pose). This information may in turn be used to design more potent and selective analogs.6
- Bioremediation – Protein ligand docking can also be used to predict pollutants that can be degraded by enzymes

LIST OF PROTEIN LIGAND DOCKING SOWTWARE

The number of docking programmes available today is large and has been gradually expanding over the past few decades. The following list provides an overview of the most popular protein-ligand docking systems, listed alphabetically with the year of release, participating organisation or institution, short description, availability of a web service, and licence. This table is extensive but not exhaustive.



Program	Year Published	Organisation	Description	Webservice	License
1-Click Docking	2011	Mcule	Docking predicts the binding orientation and affinity of a ligand to a target	Available →	Basic free version
AADS	2011	Indian Institute ofTechnology	Automated active site detection, docking,and scoring(AADS) protocol for proteinswith known structures based on Monte Carlo Method	Available →	Free to use Webservice
ADAM	1994	IMMD Inc.	Prediction of stable binding mode of flexible ligand molecule to target macromolecule	No	Commercial
AutoDock	1990	The Scripps Research Institute	Automated docking of ligand to macromolecule by Lamarckian Genetic Algorithm and Empirical Free Energy Scoring Function	No	Freeware →
AutoDockVina	2010	The Scripps Research Institute	New generation of Auto Dock	No	Open source →
BetaDock	2011	Hanyang University	Based on Voronoi Diagram	No	Freeware →
Blaster	2009	University of California San Francisco	Combines ZINC databases with DOCK to find ligand for target protein	Available →	Freeware

BSP-SLIM	2012	University of Michigan	A new method for ligand-protein blind docking using low-resolution protein structures	Available →	Freeware
DARWIN	2000	The Wistar Institute	Prediction of the interaction between a protein and another biological molecule by genetic algorithm	No	Freeware
DIVALI	1995	University of California-SanFrancisco	Based on AMBER-type potential functionand genetic algorithm	No	Freeware
DOCK	1988	University of California-SanFrancisco	Based on Geometric Matching Algorithm	No	Freeware for academic use →
DockingServer	2009	Virtua Drug Ltd	Integrates a number of computational chemistry software	Available →	Commercial
DockVision	1992	DockVision	Based on Monte Carlo, genetic algorithm, and database screening docking algorithms	No	Commercial →
DOLINA	2013	University of Basel	Combinatorial solution to local induced fit	No	Academic
EADock	2007	Swiss Institute of Bioinformatics	Based on evolutionary algorithms	Available →	Freeware →
eHiTS	2006	SymBioSys Inc	Exhausted search algorithm	No	Commercial →
EUDOC	2001	Mayo Clinic Cancer Center	Program for identification of drug interaction sites in macromolecules and drug leads from chemical databases	No	Academic
FDS	2003	University of Southampton	Flexible ligand and receptor docking with a continuum solvent model and soft-core energy function	No	Academic
FlexX	2001	BioSolveIT	Incremental build based docking program	No	Commercial →
FlexAID	2015	University of and soft scoring	Target side-chain flexibility	No	Open source →

CONCLUSION

Molecular docking is simple and useful for researching, understanding, explaining, and identifying molecular features in three dimensions. The structure of the intermolecular complex produced between two or more constituent molecules is attempted to be predicted through

molecular docking. Its involvement in intriguing new approaches such as computational enzymology, genomics, and proteomic search engines continues to grow. Molecular docking has identified interesting chemicals that might be future treatments in crucial areas of human health.

REFERENCES

1. Atkins PW (1997): Physical Chemistry, 6th edn. New York: W. H. Freeman & Co.
2. Nadendla Rama Rao, molecular modelling: A powerfull tool for drug design and molecular docking, Resonance. 2004; 51-60
3. Molecular modelling from Wikipedia, the free encyclopedia [update 22 Nov 2011.] Available from http://en.wikipedia
4. .org/wiki/molecular_modelling.
5. Lengauer T, Rarey M. Computational methods for bimolecular docking. Curr Opin Struct Biol. 1996; 6: 402-406.
6. 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: 935-949.
7. Pozzan A. Molecular descriptors and methods for ligand based virtual high throughput screening in drug discovery. Curr Pharm Des. 2006;
8. Schulz-Gasch T, Stahl M. Binding Site Characteristics in Structure-based Virtual Screening: Evaluation of Current Docking Tools. J Mol Model. 2003; 9: 47-57.
9. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking acuracy. J Med Chem. 2004; 47: 1739-1749.
10. Jones G, Wilett P, Glein RC, Leach AR, Taylor R. Development and Validation of Genetic Algorithm and an Empirical Binding Free Energy Function. J Mol Biol. 1997; 267: 727-748.
11. Venkatachalam CM, Jiang X, Oldfield T, Waldman M. LigandFit: A Novel Method for the Shape-directed Rapid Docking of Ligands to Protein Active Sites. J Mol Graphics Modell. 2003; 21: 289-307
12. Molecular docking using VLifeMDS: VLife product

documentation, tutorial: Biopredicta, VLife sciences Technologies Pvt.ltd.2011; 1-17.

13. Schapira M, Abagyan R, Totrov M. Nuclear hormone receptor targeted virtual screening. J Med Chem 2003;46: 3045-59.
14. Davis AM, Teague SJ, Kleywegt GJ. Application and Limitations of X-ray Crystallographic Data in Structure- Based Ligand and Drug Design. Angew Chem Int Ed Engl 2003; 42: 2718-36.
15. Jorgensen WL. Rusting of the lock and key model for protein Anderson AC, O'Neil RH, Surti TS, Stroud RM. Approaches to solving the rigid receptor problem by identifying a minimal set of flexible residues during ligand docking. Chem Biol 2001; 8: 445-57.
16. Nakajima N, Higo J, Kidera A, Nakamura H. Flexible docking of a ligand peptide to a receptor protein by multicanonical molecular dynamics simulation. Chem Phys Lett 1997; 278: 297-301.)
17. Abseher R, Horstink L, Hilbers CW, Nilges M. Essential spaces defined by NMR structure ensembles and molecular dynamics simulation show significant overlap. Proteins 1998; 31: 370-82.