MOLECULAR DOCKING: AN INNOVATIVE TOOL IN DRUG DESIGNING

Jeenu Joseph[1*], Akhila S[2], Presannakumaran P.N[3], Santhosh M. Mathews[4]

Professor, Professor, Professor, Professor,
[1]Department of Pharmacognosy, Pushpagiri College of Pharmacy, Thiruvalla, Kerala, India.

Abstract: Computer aided drug discovery (CADD) and development process is attaining rapid popularity, accomplishment and appreciation in drug discovery and potential drug development process. Molecular docking is a valid computational technique that enable to predetermine the best binding orientation amid the 3D structure of ligand and protein (drug molecule) with the aid of search algorithms. Scoring function is used to analyze the results by converting the overall interaction energy into numerical values namely docking scores This article furnish basic information regarding molecular docking, types and steps of molecular docking, search algorithms, scoring functions and applications of molecular docking.

Index Terms – Molecular docking, protein ligand complex, drug design, virtual screening

I. INTRODUCTION

Virtual screening techniques are widespread today for novel drug discovery. Virtual screening enable rational drug discovery leading to effective direct drug screening and has the advantage of low cost.[1-3] Virtual screening are categorized into ligand based methods and structure based methods. Ligand based methods are applied when there are informations about active ligand molecules and little or no structural informations attainable about the targets. Here ligand based methods like pharmacophore modeling and QSAR (Quatitative Structure Activity Relationships) methods are utilized. Molecular docking studies are most popular technique in structure based drug discovery ever since early 1980s.[4]

New drug discovery mainly depend on In-silico-chemico biological approach. Computer aided drug discovery (CADD) and development process is attaining rapid popularity, accomplishment and appreciation in drug discovery and potential drug development process.

Advantages of CADD are as follows:

1. Computational ability enables to modernize drug discovery and development process.
2. Biological and chemical information regarding the targets (proteins) and ligands aid in disclosing and optimizing novel drugs.
3. Development of *insilico* filters aid to eliminate chemical compounds with poor activity or poor absorption, distribution, metabolism, excretion and toxicity (ADMET) and most promising candidates can be sorted out.
4. Novel drug targets can be identified and retrieved through protein data bank (PDB) which consist of database of target protein structures. Computer aided drug discovery (CADD) enable the discovery of novel drug candidate. (hits)
5. Novel potential drug can be discovered by virtually screening the drug candidates from a database.[5,6]

Molecular docking

Molecular docking is a valid computational technique that enable to pre ascertain the interaction energy of two molecules to form a stable complex having overall minimum energy. Molecular docking method help to predetermine the best binding orientation amid the the 3D structure of ligand and protein (drug molecule) with the aid of search algorithms. Proteins cavities have active sites which when comes in contact with external compounds become active and the small molecules (ligands) fits within these protein cavities to make it active. Strength of binding affinity of ligand and protein can be calculated utilizing scoring function. [7,8]

In various stages of drug discovery, docking studies can be applied to:
a) Ascertain the binding mode of already ascertained ligands.
b) Discover novel, competent and potent ligands.
c) Predict the binding affinity of the ligand and the protein.[9]

Scoring function is used to analyze the results by converting the overall interaction energy into numerical values namely docking scores. Visualizing tools like Pymol, MGL tools, Rasmol etc help in visualizing the 3D pose of the bound ligand and 'best fit of ligand' can be inferred. The active site of the protein can be determined from the protein ligand interaction that further enable in protein annotation. Information regarding the free energy, binding energy and stability of the resultant complexes can be predicted by docking studies. Thus molecular docking is an attractive scaffold that plays a vital role in rational drug designing and drug discovery process. [10]

Types of Docking:

a) Rigid /Lock and Key Docking- Emil fischer (1894) proposed Lock and Key theory where specificity of an enzyme against its substrate in biological system rely on its complementary geometric shapes that aid in fitting precisely like a Lock and Key. In this docking, internal geometry of ligand and receptor is preserved.
b) Induced fit/ Flexible Docking- In this study, both the side chain of protein and ligand is maintained as flexible. The energy and surface cell occupancy of each conformation is also calculated. The main chain of the protein is also moved to integrate possible confirmations that occurs during the interactions of the protein and ligand. Various attainable confirmations can be effectively evaluated by this method. Though this method may be time consuming and expensive, it is efficient and trust worthy. [11,12]

Various types of molecular interaction forces:

There are mainly four types of interaction forces.
1. Electrostatic force- Interactions include charge-charge, dipole-dipole, charge-dipole
2. Elecrodynamic forcesVanderwaals interactions
3. Steric forces
4. Solvent related forces-constitute hydrophobic and hydrophilic (Hydrogen bonds) interactions. [13,4]

Steps in mechanics of molecular docking

Following are the various steps in molecular docking:

1. **Protein preparation**
   3D structure of the protein is retrieved PDB (protein data bank) this 3 D structure has to be pre processed by various processes like removing molecule, charge stabilization, missing residues filling, side chain generations etc.
2. **Prediction of active site**
   Numerous active sites may be present in the protein from which the concerned one should be sorted. The water molecules and other hetero atoms are removed.

3. **Ligand Preparation**

Retrieval of ligand can be made from databases such as Pub Chem, ZINC or it can be sketched with the aid of Chem sketch tool. Ligand can be selected as per LIPINSKI RULE OF 5 that help to discern drug like and non drug like candidates.

4. **Docking**

In docking, selected ligand is docked against protein and the resultant interactions are further analyzed with the aid of statistical scoring function that are converted to numerical values called as the docking score. [14, 15]

Search Algorithms

For a given complex (protein-protein and protein-ligand), all predictable optimum configurations consisting of relative position and orientation of the molecules are created by the algorithms. The energy of the individual interactions and resulting complex are also determined by these algorithms. Various algorithms for docking analysis are as follows: [16]

Random search or Stochastic methods

In this method random changes are produced to a single or group of ligands and thereafter evaluated by predefined probability function. Algorithms produce huge number of molecular confirmations depending on the probability criteria from which most favourable conformations are selected. The huge computational cost to generate all probable confirmations is a limitation. Random search or stochastic search that utilize various probability criterion of acceptance are involved in Genetic algorithm, Monte Carlo simulation.[17]

Monte carlo algorithm

Monte Carlo method employ random initial configuration of ligands in the active site that is thereafter scored depending on specific properties like energy. Monte carlo bring about ligand poses via bond rotation and rigid body rotation of translation. The resulting confirmation if passes after testing with an energy based selection criteria will be further saved and modified to create next confirmation. Compared to molecular dynamics the convenience of monte carlo is that consist of quite large changes that enable the ligind to easily cross the energy barriers potential energy surface.[18-20]

Genetic algorithm

Genetic algorithm rely on natural genetics language and principles of biological evolution. In molecular docking, the particular ligand and protein arrangement is outlined by set of values specifying translation, confirmation and orientation of ligind with respect to protein. In genetic algorithm these parameters are called 'state variable' which correspond to genes in a chromosome. These parameters are encoded and stochastically changed which thereafter evaluated using fitness function. Fitness value in molecular docking is the total interaction energy of ligand protein. Depending on the fitness value, the genes are inherited by new chromosomes from both parent. In random mutation one gene of some offspring are changed by random amount. The mutation with better fitness value is only accepted. Solutions that are poorly suited to the environment die, whereas better suited ones reproduce. Genetic operators alter genes to bring about new ligand structure. With the aid of scoring function, these new ligand structures will be assessed and the surviving ones are utilized for next generation. Genetic algoritham is utilized in AutoDock, DIVALI, DARWIN and GOLD. [21]

Matching algorithm

Matching algorithm rely on molecular shape and chemical information to chart a ligand into the protein active site. Proteins and ligand is denoted as pharmacophore. Distance of the pharmacophore inside protein and ligind is computed for a match. New ligand conformations are ruled by distance matrix within the pharmacophore and matching ligand.[22-24] Matching algorithm is employed in DOCK, LibDock, SANDOCK and FLOG.[25-27]

Incremental construction

In incremental construction method, ligand is employed in a fragmental and incremental fashion into active site. Rotatable bonds are ruptured to form several fragments out of which one frgament is chosen to dock into the active site of the protein. Remaining fragments are added incrementally. Various orientations are created to fit into the active site. [28-30]The incremental construction method is employed in DOCK 4.0, SLIDE, eHiTS, FlexX etc. [31-33]

Molecular Dynamics

Molecular dynamics is a potent simulation method applied in molecular modeling molecular dynamics simulation utilized both flexible ligind and protein more effectively compared to other algorithm but it has a disadvantage in progressing in very small steps that result in difficulties in breaking high energy conformational barriers that may result in inadequate sampling for local optimization molecular dynamics is more effective [34-36]

Exhaustive search algorithm

In exhaustive search algorithm ligand confirmation are demonstrated by systematically rotating all probable rotatable bonds during a given interval. Exhaustive systematic search is restricted in large confirmational space. Favourable initial ligand poses are spotted out by rough positioning and scoring methods. [37,38]

Scoring Function

After docking, the binding affinity which is the strength of non covalent interaction between two molecules can be predicted by mathematical methods namely scoring function. Design of consistent, accurate and reliable scoring function is vital for the virtual database screening. During docking studies, enormous ligand poses are generated and due to clashes with the proteins, some are instantly rejected. After conducting the docking studies between the ligand and the protein, the binding affinity (strength of the non-covalent interaction) between the ligand and the protein molecules are mathematically calculated. The experimental binding modes are distinguished from all other modes with the aid of search algorithms.

Scoring functions are based on physics that rely on molecular mechanics force fields. The total energy of the pose within the binding site are contributed by conformational changes occurs in the protein and ligand, internal rotations, solvent effect, free energy that arise due to protein ligand interactions, association energy of protein and ligand to form a single complex and free energy arise due to alterations in the vibrational modes. Strength of other inter molecular interactions (protein drug and protein DNA) are also calculated by scoring function. These scoring functions directly indicate the ligand-protein binding affinity, which means best scoring function indicate best binders. A stable system which means a likely binding interaction is denoted by a low or negative energy.

Scoring function is constituted by three expressions which are:
a) Ranking of the possible configurations
b) Ranking between various ligands and protein through virtual screening
c) According to the binding affinities one or more ligands are ranked against various proteins.[39]

Free energy estimations techniques are employed for generating scoring functions of the various ligands and protein docking complexes. It can be denoted by following equations:

$\Delta G_{bind} = \Delta G_{solvent} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{rot} + \Delta G_{trans/rot} + \Delta G_{vib}$

$\Delta G_{solvent}$ denote ineraction of ligand and protein with solvent

ΔG_{conf} denote effect of conformational changes in ligand and protein ΔG_{int}

ΔG_{rot} denote loss of free energy due to freezing rotatable bonds (entropic contribution)

ΔG_{int} denote free energy of specific ligand protein interaction.

$\Delta G_{trans/rot}$ denote loss of translational and rotational free energy due to the association of two bodies (ligand and protein) to form a single ligand and protein complex.

Scoring functions frame many assumptions and simplifications of the above terms. Scoring functions are of three types which are force-field based scoring, empirical scoring and knowledge based scoring.[38, 40]

Force-field based scoring

Force-field denote the energy of the system as a grand total of multiple non bonde terms (electrostatic interactions, van der Waals interactions and bond stretching or bond bending torsional forces)entailed in molecular recognition. Force-field methods avail variety force-field parameters to evolve Force-field based scoring function.[41]

Empirical scoring

Empirical scoring avail many intermolecular interactions calibrated with the aid of maximum experimentaldata. It utilize energy terms whose coefficients or weights rely on experimental datas generated fromregression analysis with the aid of x-ray structures and experimentally obtained binding energies.[42]

Knowledge based scoring

Knowledge based scoring are invented to reproduce structures than binding energies. Knowledge based scoring aim to absolutely capture binding effects which are difficult to model specifically. A good balance among accuracy and speed is obtained from knowledge based scoring compared to empirical scoring and force-field based scoring. [43]

Applications of Molecular Docking [8-10]

Interaction of small molecule (ligand) and enzyme protein lead to inhibition or activaton of enzyme. Docking studies can be beneficially used in the field of drug design and it applications include:

  a) **Hit identification**: Huge database of potential drugs can be rapidly screen *insilico* with the aid of docking conjointly with scoring function to detect protein target of interest.

  b) **Lead optimization**: Best possible orientation of the ligand molecule into the target protein can be predicted with docking. This enables to design more competent, selective and potent drug analogs.

  c) **Bioremediation**: Docking studies may ascertain pollutants that may be degraded by pollutants.

  d) **Drug DNA interaction**: Molecular docking predict binding property of a drug to nucleic acid. This help to explore the correlation amid structure of the drug along with its cyto toxicity. This help to implement structure modification to the drug to enable them to bind structure specifically to the target.

  e) **Receptor Preparation**
  Structure and binding site of the protein can be predicted
  Protein-protein interaction can be predicted

  f) **Preparation of ligand**
  PKa values are predicted for charged atom. Within a specific PH range, programmes are executed for all the attainable charge arrangements.

  g) **Modern drug development**
  Proposed medicine's precision against homologous proteins can be detected. Protein protein interaction can be identified by detecting the interaction of protein with other proteins like cytochrome P450, protease etc.

Discussion

Molecular docking is an *in silico* method for novel drug discovery based on 3D structural complexes of drug molecule. Best binding structural complexes of ligand and protein can be attained with molecular docking studies, search algorithms and scoring functions that can be further utilized for future investigation, identification and interpretation of molecular properties. Molecular docking studies are safe, convenient and economic tool widely accepted in various fields of computational chemistry and for virtual screening of biologically active molecules.

REFERENCES

1. Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br J Pharmacol.* 2008;153(Suppl 1):S7-26.
2. Shoichet BK, McGovern SL, Wei B, Irwin JJ. *Hits, leads and artifacts from virtual and high throughput screening.* 2002. Molecular Informatics: Confronting Complexity.
3. Bailey D, Brown D. High-throughput chemistry and structure-based design: survival of the smartest. *Drug Discov Today.* 2001;6(2):57–59.
4. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. *J Mol Biol.* 1982;161(2):269–288.
5. Pozzan A. Molecular descriptors and methods for ligand based virtual high throughput screening in drug discovery. Curr Pharm Des. 2006; 12: 2099-2110.
6. Green DV. Virtual screening of virtual libraries. Prog Med Chem. 2003; 41: 61-97.

7. Rohs R, Bloch I, Sklenar H, Shakked Z (2005)Molecular flexibility in ab-initio drug docking to DNA: binding-site and binding-mode transitions in all-atom Monte Carlo simulations. Nucl Acids Res 33: 7048-7057.

8. Guedes IA, de Magalhães CS, Dardenne LE (2014) Receptor-ligand molecular docking. Biophysical Reviews 6: 75-87.

7. Morris G.M., Goodsell D.S., Halliday R.S., Huey R., Hart W.E., Belew R.K., Olson A.J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 1998;19:1639–1662.

8. Agarwal S, Chadha D, Mehrotra R (2015) Molecular modeling and spectroscopic studies of semustine binding with DNA and its comparison with lomustine–DNA adduct formation. J Biomol Struct Dyn 33: 1653-1668.

9. Trosset JY, Scheraga HA. PRODOCK: Software Package for Protein Modeling and Docking. J Comput Chem. 1999; 20: 412-427.

10. A Tripathi, VA Bankait. Molecular docking: From Lock and key to combination lock. J Mol Med Clin Appl. 2017; 2(1): 10.16966/2575-0305.106.

11. Goodsell DS, Olson AJ. Automated Docking of Substrates to Proteins by Simulated Annealing, Proteins. 1990; 8: 195-202.

12. McMartin C, Bohacek RS. QXP: Powerful, Rapid Computer Algorithms for Structure-based Drug Design. J Comput Aid Mol Des. 1997; 11: 333-344.

13. Schnecke V, Kuhn LA. Virtual Screening with Solvation and Ligand-induced Complementarity, Perspect. Drug Discov. 2000; 20: 171-190.

14. 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.

15. Van Der Spoel D, Lindahl E, Hess B, et al. GROMACS: fast, flexible, and free. *J Comput Chem.* 2005;26(16):1701–1718.

16. 16. Christen M, Hunenberger PH, Bakowies D, et al. The GROMOS software for biomolecular simulation: GROMOS05. *J Comput Chem.* 2005;26(16):1719–1751.

17. 17. Liu M, Wang S. MCDOCK: A Monte Carlo simulation approach to the molecular docking problem. *J Comput Aided Mol Des.* 1999;13(5):435– 451.

18. 18. Totrov R, Abagyan R. Flexible protein–ligand docking by global energy optimization in internal coordinates. *Proteins.* 1997;(Suppl):215–220.

19. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct.* 2003;32:335–373.

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

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

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

23. Miller MD, Kearsley SK, Underwood DJ, Sheridan RP. FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. *J Comput Aided Mol Des.* 1994;8(2):153–174.

24. Diller DJ, Merz KM., Jr. High throughput docking for library design and library prioritization. *Proteins.* 2001;43(2):113–124.

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

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

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

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

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

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

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

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

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

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

35. 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(7):1739–149.

36. Umesh Yadava. Search algorithms and scoring methods in protein ligand docking. *Endocrinol Metab Int J.* 2018;6(6):359–367.

37. 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. (15)

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

39. 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(1):128–137.

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

41. Li X, Liang J. Knowledge-based energy functions for computational studies of proteins in Computational Methods for Protein Structure Prediction and Modeling. Xu Y, Xu D, Liang J, editors. New York, Springer.

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

43. Li X, Liang J. Knowledge-based energy functions for computational studies of proteins in Computational Methods for Protein Structure Prediction and Modeling. Xu Y, Xu D, Liang J, editors. New York, Springer.