INTERNATIONAL RESEARCH JOURNAL OF PHARMACY

www.irjponline.com

ISSN 2230 – 8407

Review Article

MOLECULAR DOCKING SOFTWARE'S APPLICATIONS AND BASIC CHALLENGES FACED: A REVIEW

Shalini K. Sawhney *, Monika Singh

ITS College of Pharmacy, Muradnagar, Ghaziabad, Uttar Pradesh, India

*Corresponding Author Email: shaliniksawhney@its.edu.in

Article Received on: 06/01/20 Approved for publication: 09/02/20

DOI: 10.7897/2230-8407.110213

ABSTRACT

Molecular docking shows a very important role in the rational design of drugs. Method of molecular docking calculates the preferred orientation of one molecule with the second molecule when bound to each other to form a stable complex. Knowledge of the preferred orientation in turn may be used to predict the strength of association or binding affinity between two molecules. Docking is normally 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. Given the biological and pharmaceutical importance of molecular docking, considerable efforts have been directed towards improving the methods used to calculate docking. The docking process simulates the ligand-protein pairwise interaction and the energies are calculated. There are three types of Docking: Protein-Protein Docking, Protein-Ligand Docking and Protein-Protein and Protein-Ligand docking. Molecular docking software includes Auto dock, Flex-X, Glide, Gold, ZDOCK, RDOCK, LeDOCK, Dock, Auto dock Vina, Mdock etc. The present review accumulates the characteristics and applications of the different software used and the basic challenges faced in molecular docking studies.

Keywords: Drug discovery, computer aided drug design, computer-aided molecular modelling, computer-aided molecular design

INTRODUCTION

The discovery and development of drug involves the skills and proper knowledge of both an artist and the scientist.[1] The discovery of a single drug molecule requires investment of huge amount of resources by biotechnology and pharmaceutical companies. The process of drug discovery is very exclusive. As claimed by many such companies, the drug development process costs approximately between $800 million - $900 million with a time interval of around 12-15 years. In modern drug discovery, computer plays a vital role by saving both time and money, as the process is follows *in-silico* chemico biological approach.[2] The process of new drug development is a multi-step process which includes pre-clinical development and clinical trials.[3]

Chemicals available can be studied by *in vitro* high-through put screening (HTS) to know their ability in binding, inhibiting or activating the biomolecular targets[4,5]. Due to the large cost, researchers are putting their efforts towards improvement of computational tools for performing virtual screening (VS) before HTS.[6,7]

The aim of pre-clinical development is the identification and optimization of a lead compound for a targeted action in which the pharmacokinetic and pharmacodynamic investigation of the drug is involved and generally begins with *in silico* examinations which further leads to *in vitro* and ultimately *in vivo* studies.[8] After obtaining reasonable biological actions and safety and on defining a proper galenic formulation for the lead compound, the phases of clinical trial starts consecutively.[9]

Computational techniques are being used in the process of drug discovery and development.[10] Terms used in this field, includes computer-assisted drug design (CADD), computer-aided molecular modelling (CAMM), computer-aided molecular design (CAMD), computational drug design, rational drug design, computer-aided rational drug design, *in silico* drug design. CADD involves drug discovery and development using computational methods which includes[2]:

I. Utilizing the biological and chemical evidence about targets and/or ligands for identification and optimization of the new drugs.

II. Elimination of compounds with undesirable properties (poor activity and/or poor Absorption, Distribution, Metabolism, Excretion and Toxicity, (ADMET)) by using design of in-silico filters and choice of the most promising candidates. Fast development in this area has been made possible by improvements in software and hardware computational power and sophistication.

III. CADD is used for identification of active drug candidates (hits), selection of most likely candidates for further evaluation (leads), and for optimization of leads i.e. transformation of compounds with biologically activity into appropriate drugs by refining their ADMET/PK (pharmacokinetic), pharmaceutical, physicochemical properties.

IV. Virtual screening is employed for discovering new drug candidates by searching private 3-D chemical structure, public, or commercial databases. This aims for amplification molecular asset with desirable properties (active, drug-like, lead-like) and eliminate compounds with properties that are not desired (reactive, inactive, poor pharmacokinetic profile, toxic).

Virtual screening is growing very rapidly and the increased number of citations with keywords "virtual screening" from 4 to 302 in 1997 and in 2004 respectively is an evidence for this.[11] All these techniques are becoming very important part of drug discovery and development process and it is expected that the computational method can help in improving the efficiency for the industry by decreasing the requirement of resources.[12]

Molecular docking may be defined as an optimization process, which would describe the "best-fit" orientation of a ligand that binds to a particular protein of interest. However since both the ligand and the protein are flexible, a "hand-in-glove" analogy is more suitable than "lock-and-key". During the course of the process, the ligand and the protein adjust their conformation to achieve an overall "best-fit" and this kind of conformational modifications resulting in the overall binding is referred to as "induced-fit". The focus of molecular docking is to computationally stimulate the molecular recognition process. The purpose of molecular docking is to achieve an optimized conformation for both the protein and ligand and relative orientation between protein and ligand such that the free energy of the overall system is minimized.[13,14] Molecular modelling includes all the methods either theoretical or computational used for modelling or mimicking the molecule's behaviour. The techniques are employed in computational biology, drug design, materials science and computational chemistry for the study of molecular organisations which ranges from small chemical systems to large biological molecules and associations of material. Manual method can be used for simple calculations, but for carrying out the molecular modelling of larger sized system computers are needed.[10]

Three stages mainly involved in molecular modelling are[2]:

a) Selection of a model for describing the intra-molecular and inter- molecular interactions in the system. Quantum mechanics and molecular mechanics are the two most common models that are used in molecular modelling. These models facilitate the energy of any arrangement of the atoms and molecules in the system to be calculated and permit the modeler to determine how the energy of the system differs as the positions of the atoms and molecules change.

b) Calculations, like a conformational search, Monte Carlo simulation or a molecular dynamics, or an energy minimization.

c) Analysis of calculation is done, for calculating the properties as well as for scrutiny that whether the method has been performed appropriately or not.

Docking plays a vital role in the rational drug design. The method of molecular docking allow the searching in an automatic and more objectives way to assess the match between two molecules.[15] The estimation of available docking programs is a useful task, which helps in deciding the most appropriate docking algorithm for any distinct study.[2]

Binding of a small molecule ligand with an enzyme protein may either stimulate or inhibit the enzyme. In case when the protein is a receptor, this binding may either cause antagonism or agonism.[2] Docking may be applied to:

Hit Identification

In combination with scoring function, docking can be used for quick screening of large databases of desirable drugs *in-silico*, for identifying the molecules that can bind to selected protein target (Virtual Screening).[14]

Lead Optimization

Docking helps in, relative orientation of a ligand binding to the target protein in predicting (both site and type of orientation) which help in designing the analogs with more potency and selectivity.[2]

Bioremediation

Protein ligand docking is also used to predict contaminants that can be degraded by enzymes. It also helps in estimating the binding affinity.[16]

Searching for lead structures

Searching for lead structures for protein targets

Molecular Docking Software

Flexible-rigid docking has been extensively used. Since flexible docking is usually more accurate, the relevant investigates have become the most studying spot in recent years. Following is the broadly used molecular docking software and its algorithms, evaluation methods, features and application areas.

Molecular docking databases

The most prevalent protein structure database is the public database Protein Data Bank (PDB).[17] Also, the public databases such as Pub Chem Compound Database[18] and ZINC[19] are free to use. Besides, there are many important commercial databases, such as Compound Database (AcD)[20], Cambridge Structural Database (CSD).[21]

Software for molecular docking

Flex X

Fragmentation algorithm, Semi-empirical calculation on free energy; it is fast flexible-rigid docking. It can be used for virtual screening of small molecule databases by using incremental construction strategy[22]

Gold

GA (genetic algorithm) Semi-empirical calculation on free energy; it is fast flexible docking. It is a GA-based docking program. The accuracy and reliability of this software have been highly evaluated[23]

Glide

Exhaustive systematic search for semi-empirical calculation on free energy Medium Flexible docking; this software uses domain knowledge to narrow the searching range and has XP (extra precision), SP (standard precision) and high throughout virtual screen modes

Auto Dock

GA (genetic algorithm) LGA (Lamarckian genetic algorithm) Semi-empirical calculation on free energy Medium Flexible-rigid docking; this software is always used with Auto dock-tools and it is free for academic use[25]

ZDOCK

Geometric complement-arity and molecular dynamics Molecular force field Medium Rigid docking. Chen *et al*.[26] propose a new scoring function which[26] combines pair wise shape complementarily (PSC) with desolation and electrostatic and develop the ZDOCK server[27]

RDOCK

GA (genetic algorithm) MC (monte carlo) MIN (Simplex minimization)[28] Molecular force field Medium Rigid docking. The CHARMm-based procedure for refinement and scoring. Besides predicting the binding mode, it is especially designed for high throughput virtual screening (HTVS) campaigns

LeDOCK

Simulated annealing (SA) Genetic algorithm (GA); molecular force field Fast Flexible docking; LeDock is a new molecular docking program. From the results of the present study[41], since it is fast and exhibits a high accuracy, it is recommended for the virtual screen task[40]

Dock

Fragmentation algorithm Molecular force field Fast Flexible docking; it is widely applicable and is always used in docking between flexible proteins and flexible ligands[42]

Auto dock Vina

GA (genetic algorithm) Semi-empirical calculation on free energy; it is Fast Flexible-rigid docking. Auto Dock Vina employs an iterated local search global optimizer and it is faster than the Auto Dock 4[6]

Docking technique identifies the novel targets by assigning a single small-molecule ligand as the investigation to dock with multiple receptors to discover potential binding cavities. In this way, the potential targets of drug can be predicted. For example, Grinter *et al*.[29] explored the potential target oxidized squalene cyclase (OSC) of PRIMA-1 by using the reverse docking software package Mdock. Also, Chen *et al*.[30] applied reverse docking technique to discover targeted proteins of marine compounds with anti-tumour activity. Furthermore, Chen *et al*[30] also indicated that reverse docking can be complementary with *in vitro* assays as an operative method of target fishing. Finally, we reflected that exploring relevant mechanism of action or side effect profile by structural biology analysis[31] such as the pocket analysis[32]; could significantly benefit the novel drug design.

Applications of molecular docking

Virtual screening to discover the lead compound and hit compound

Virtual screening[33] is to discover the lead compound and hit compound from the molecular databases according to the scoring function, which has greatly improved the screening competence compared with the traditional screen method. The applications of virtual screening are commonly used. Notably, given the exponential growth of high through put[34], high-performance computing[35], machine learning[36] and deep learning[37] techniques, the integrated method flourishes quickly. For example, Pereira *et al*.[38] applied deep learning approach in virtual screening, which extracts relevant features from molecular docking data to create the distributed vector representations for protein-ligand complexes. Also, Pyzerknapp *et al*.[39] proposed the virtual high through put screening

Prediction of potential targets

It is noted that the above-mentioned methods are all general docking methods which use the different ligands in the database to dock with the same receptor. However, current commonly used reverse docking technique is different from them.

Basic challenges in molecular docking

Certain basic challenges include:

Ligand chemistry

The ligand preparation has noticeable effect on the docking results because the ligand identification by any bio molecule depends on 3-dimensional orientation and electrostatic interaction. This assures that the conformation of both the ligand as well as ligand preparation is important. Earlier, keeping estimated pKa values, the structure being most likely optimized by eliminating or adding hydrogens but the tautomeric and protomeric states of the molecules which are to be docked, still continued a major discrepancy. Since almost all data bases keep molecules in their neutral forms but under physiological conditions they are actually ionized. Hence it is compulsory to ionize them prior to docking. But in different programs, the standard ionization is easy to achieve; regarding the issue of tautomers, the question still remains there, that which tautomer one should use or to use all possible tautomers.[44]

Receptor flexibility

This is a major challenge in docking i.e., treating flexible protein. A bio molecule/protein adopts different conformations depending upon the ligand to which it binds. This confirms that docking done with a rigid receptor will give a single conformation of receptor. But, when the docking is done with flexible receptor, the ligands may require many receptor conformations to get attached. In molecular docking studies, usually the most ignored aspect is different conformational states of proteins. Since the protein flexibility is significant as it accounts for better affinity to be achieved between a given a drug and target. Another aspect of target flexibility is active site water molecules. Water molecules must be rectified to avoid using artefact waters in the docking process.[45]

Scoring function

Another challenge in docking is deficiency in scoring function. Just like search algorithm is having potential to give optimum conformation, scoring function should also be able to distinguish true binding modes from all the other parallel modes. A potential scoring function would be computationally much economical, unfavourable for analysing several binding modes. When there is accuracy, scoring functions make number of proposals to evaluate ligand affinity. The physical phenomenon i.e., entropy and electrostatic interactions are disregarded in scoring schemes. Therefore the deficiency of suitable scoring function, in terms of accuracy and speed, is the main concern in molecular docking programming.[46] It is also necessary to take thermodynamic features into observation[47], or use retrospective verification to estimate the reliability of the prediction of affinity.[48]

CONCLUSION

The methodology of drug design is very useful in providing valuable information for identification and validation of the target, small-molecular screening and optimization and selection of the lead molecule. Molecular docking method screens large databases of compounds at a lower cost than experimental techniques like high through put screening is principally notable. New approaches are required to ease, accelerate and streamline the process of drug discovery and development with advantage of saving resources, money and time. But it is necessary to take thermodynamic features into account and use retrospective verification to evaluate the reliability of the prediction of affinity.

In the distant future, we are improving the conformational search algorithm by taking more flexible bonds, solvent states and integrating recent biological data mining algorithms in consideration. In general, we trust that molecular docking technique will become such a reliable drug-design tool that assimilates the big biological data by optimizing the scoring function and upgrading the appropriate search algorithms. The potential docking technique is done after thoroughly screening the target, ligands and docking method performance. The ligand flexibility however is almost resolved and does not create much problem however protein flexibility needs to be improved. Water molecules should be included to consider the hydrogen bonding with non-aqueous residues. In this short review, we have focused on types, approaches, applications and challenges of molecular docking in brief but accounting for flexibility and successful scoring remain significant challenges.

ACKNOWLEDGEMENT

Authors thank ITS College of Pharmacy, Muradnagar, Ghaziabad, India for the support to publish the work.

REFERENCES

1. Mullasseril A. Computer assisted designing (CADD), *in silico* pharmacological studies followed by actual synthesis and characterization of Garcicillin a novel semi synthetic antibiotic. Der Pharma Chemica 2012; 4(5): 1991-96.
2. Eweas AF, Maghrabi IA, Namarneh AI. Advances in molecular modeling and docking as a tool for modern drug discovery. Der Pharma Chem 2014; 6: 211–28.
3. Moran M. Cost of Bringing New Drugs to Market Rising Rapidly. Psychiatry 2003; News 38: 25.
4. Good AC, Krystek SR, Mason JS. High-through put and virtual screening: core lead discovery technologies move towards integration. Drug Discovery Today 2000; 5: 61–9.
5. Carnero A. High throughput screening in drug discovery. Clinical Translation Oncology 2006; 8: 482–90.
6. Bajorath J. Integration of virtual and high-through put screening. Nature Reviews Drug Discovery 2002; 1: 882–94.
7. Abagyan R, Totrov M. High-through put docking for lead generation. Current Opinion Chemistry Biology 2001; 5: 375–82.
8. Frantz S. Pharma faces major challenges after a year of failures and heated battles. Nature Reviews Drug Discovery 2007; 6: 5–7.
9. Danzon PM, Epstein A, Nicholson S. Mergers and Acquisitions in the Pharmaceutical and Biotech Industries, National Bureau of Economic Research; 2004.
10. Kapetanovic IM. Computer-aided drug discovery and development (CADDD): *in silico*- chemico-biological approach. Chemico- Biological Interaction 2008; 171(2): 165–76.
11. Pozzan A. Molecular descriptors and methods for ligand based virtual high through put screening in drug discovery. Current Pharmaceutical Design 2006; 12(17): 2099–110.
12. Kumar N, Hendriks BS, Janes KA, de Graaf D, Lauffenburger DA. Applying computational modeling to drug discovery and development. Drug Discovery Today 2006; 11(17–18): 806–811.
13. Lengauer T, Rarey M. Computational methods for biomolecular docking. Current Opinion in Structural Biology 1996; 6(3): 402.6.
14. Gaba M, Gaba P, Singh S, Gupta GD. An overview on molecular docking. International Journal of Drug Development and Research 2010; 2(2): 219-31.
15. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. Annual Review of Bio physics and Bio molecular Structure 2003; 32: 335–73.
16. Suresh PS, Kumar A, Kumar R, Singh VP. An *in silico* approach to bio remediation: as a case study. Journal of Molecular Graphics and Modelling 2008; 26(5): 845.9.
17. Berman H, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov I and Bourne P. The Protein Data Bank. International Tables for Crystallography, Eds. Rossmann, MG and Arnold; 1999. E. 67. p. 675–684.
18. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA. *et al*. Pub Chem substance and compound databases. Nucleic Acids Research 2016; 44: D1202–D1213.
19. Irwin JJ and Shoichet BK. ZINC–a free database of commercially available compounds for virtual screening. Journal of Chemical Information and Modeling 2005; 45: 177–182.
20. Martin GE, Hadden CE, Russell DJ, Kaluzny BD, Guido JE, Duholke WK, Stiemsma BA, Thamann TJ, Crouch RC, Blinov K *et al*. Identification of degradants of acomplex alkaloid using NMR cryoprobe technology and ACD/structure elucidator. Journal of Heterocyclic Chemistry 2010; 39: 1241–1250.
21. Groom CR, Bruno IJ, Lightfoot MP and Ward SC. The Cambridge Structural Database. Acta Crystallographyca section B Structural Science Crystal Engineering and Material 2016; 72: 171–179.
22. Kramer B, Rarey M and Lengauer T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. Proteins 1999; 37: 228–241.
23. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW and Taylor RD. Improved protein-ligand docking using Gold. Proteins 2003; 52: 609–623.
24. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT and Banks JL. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. Journal of Medicinal Chemistry 2004; 47: 1750–1759.
25. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS and Olson AJ. Auto Dock 4 and Auto Dock Tools 4: automated docking with selective receptor flexibility. Journal of Computational Chemistry 2009; 30: 2785–2791.
26. Chen R, Li L and Weng Z. ZDOCK: an initial–stage protein–docking algorithm. Proteins 2003; 52: 80–87.
27. Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T and Weng Z. ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. Bio informatics 2014; 30: 1771–1773.
28. Li L, Chen R and Weng Z. RDOCK: refinement of rigidbody protein docking predictions. Proteins 2010; 53: 693–707.
29. Grinter SZ, Liang Y, Huang SY, Hyder SM and Zou X. An inverse docking approach for identifying new potential anti-

cancer targets. Journal of Molecular Graphics and Modelling 2011; 29: 795–799
30. Chen F, Wang Z, Wang C, Xu Q, Liang J, Xu X, Yang J, Wang C, Jiang T and Yu R. Application of reverse docking for target prediction of marine compounds with anti-tumour activity. Journal of Molecular Graphics and Modelling 2017; 77: 372–377.
31. Xie H, Lee MH, Zhu F, Reddy K, Huang Z, Kim DJ, Li Y, Peng C, Lim DY, Kang, S, *et al*. Discovery of the novel mTOR inhibitor and its antitumor activities *in vitro* and *in vivo*. Molecular Cancer Therapeutics 2013; 12: 950–958.
32. Liu ZP, Liu S, Chen R, Huang X and Wu LY. Structure alignment-based classification of RNA-binding pockets reveals regional RNA recognition motifs on protein surfaces. BMC Bioinformatics 2017; 18, 27.
33. Kitchen DB, Decornez H, Furr JR and Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nature Reviews Drug Discovery 2004; 3: 935–949.
34. Joseph-McCarthy D, Baber JC, Feyfant E, Thompson DC and Humblet C. Lead optimization via high-through put molecular docking. Current Opinion in Drug Discovery and Development 2007; 10: 264–274.
35. Ge H, Wang Y, Li C, Chen N, Xie Y, Xu M, He Y, Gu X, Wu R, Gu Q *et al*. Molecular dynamics-based virtual screening: accelerating the drug discovery process by high performance computing. Journal of Chemical Information and Modeling 2013; 53: 2757–2764.
36. Melville JL, Burke EK and Hirst JD. Machine learning in virtual screening. Combinatorial Chemistry and High Through put Screening 2009; 12: 332–343.
37. Gawehn E, Hiss JA and Schneider G. Deep learning indrug discovery. Molecular Informatics 2016; 35: 3–14.
38. Pereira JC, Caffarena ER and Dos Santos CN. Boosting docking-based virtual screening with deep learning. Journal of Chemical Information and Modeling 2016; 56: 2495–2506.
39. Pyzerknapp EO, Suh C, Gómezbombarelli R, Aguileraiparraguirre J and Aspuruguzik A. What is high-through put virtual screening? A perspective from organic materials discovery. Annual Review of Materials Research 2015; 45, 45: 195–216.
40. Zhao H and Caflisch A. Discovery of ZAP70 inhibitorsby high-through put docking into a conformation of its kinasedomain generated by molecular dynamics. Bioorganic and Medicinal Chemistry Letters 2013; 23: 5721–5726.
41. Wang Z, Sun H, Yao X, Li D, Xu L, Li Y, Tian S and Hou T. Comprehensive evaluation of ten docking programs on adiverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. Physical Chemistry Chemical Physics 2016; 18: 12964–12975.
42. Therese PL, Brozell SR, Sudipto M, Pettersen EF, Meng EC, Veena T, Rizzo RC, Case DA, James TL, Kuntz ID. DOCK 6: combining techniques to model RNA-small molecule complexes. RNA 2009; 15: 1219–1230.
43. Trott O and Olson AJ. Auto Dock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multi threading. Journal of Computational Chemistry 2010; 31: 455–461.
44. Elokely KM, Doerksen RJ. Docking Challenge: Protein Sampling and Molecular Docking Performance. Journal of Chemical Information and Modeling 2013; 53: 1934-1945.
45. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: A practical alternative. Current Opinion in Structural Biology 2008; 18: 178-184.
46. Cho AE, Rinaldo D. Extension of QM/MM docking and its applications to metallo proteins. Journal of Computational Chemistry 2009; 30: 2609-2616.
47. Tame JR. Scoring functions: a view from the bench. Journal of Computer Aided Molecular Design 1999; 13: 99–108.
48. Yabuuchi H, Niijima S, Takematsu H, Ida T, Hirokawa T, Hara T, Ogawa T, Minowa Y, Tsujimoto G and Okuno Y. Analysis of multiple compound-protein interactions reveals novel bioactive molecules. Molecular Systems Biology 2011; 7: 472.

Cite this article as:

Shalini K. Sawhney and Monika Singh. Molecular Docking Software's applications and Basic Challenges Faced: A Review. Int. Res. J. Pharm. 2020;11(2):15-19 http://dx.doi.org/10.7897/2230-8407.110213

Source of support: Nil, Conflict of interest: None Declared