**Molecular Docking: A Advance Bioinformatics Strategy for Structure-based Drug Designing**

 Poojaben M Prajapati2, Komal G Lakhani1, Bharat Maitreya2  and G. V. Marviya3

1Department of Biotechnology, College of Agriculture,

 Junagadh Agricultural University, Junagadh-362 001

2Department of Botany, Bioinformatics, Climate Change and Impact Management ,

School of Science, Gujarat University, Ahemdabad-380 009

3Krishi Vigyan Kendra, Junagadh Agricultural University, Targhadia (Rajkot)- 360 023

 Email: komallakhani2706@gmail.com

**Abstract:**

The computational modeling of structural complexes created by two or more interacting molecules is known as molecular docking. Prediction of an interesting three-dimensional structure is the aim of molecular docking. Software for molecular docking is mainly employed in drug development. Molecules and simple access to structural databases have harmed a vital mechanism. Several pricey tools for drug design and research are provided by molecular docking. Simple molecular prediction and quick access to structure databases are now essential tools on the desktop of a medicinal chemist. The major application of molecular docking is virtual screening. Different computational techniques can be employed to analyze docking gain, and numerous docking programs were utilized to visualize the molecule's three-dimensional structure. In structural molecular biology and computer-aided drug design, molecular docking is crucial. Docking may be used to perform virtual screening on huge libraries of compounds, score the outcomes, and provide structural hypotheses for how the ligands reduce the target, all of which are beneficial for lead optimization.

**Keyword:**

Molecular docking; molecular docking types; docking mechanism; docking assessment; application.

1. **Introduction**

Molecular docking is the process of finding the best alignment between the ligand and receptor molecules to create a stable complex [1]. Applying the scoring function, this orientation predicts binding affinity and the bond strength between a ligand and a protein. Drug-receptor interaction reveals molecular affinities and activities [2]. It is essential for both drug discovery and drug design. The system's overall free energy has been reduced. Finding and developing new drugs is a complicated process. The In-Silico approach aids in the development of novel drugs [3]. Computer-based drug design should be used to accelerate the drug discovery process. It is helpful in computational drug design and the structural biology of molecules [4]. It's used to predict how molecules would look in three dimensions. The scoring approach is now used to execute the virtual screening for rank candidates docking for large libraries molecules [5].

1. **COMPUTER-AIDED DRUG DESIGN**
* The term "CADD Computer Aided Drug Design" describes a computer-based approach used in computational chemistry to discover, enhance, or investigate medicines and related physiologically active compounds. It is especially helpful when creating novel drugs.
* It reveals the chemical and biological characteristics of ligands and targets..
* It is employed to discover and enhance innovative medications.
* Development of in-silico filters to forecast unfavorable characteristics of pharmacological compounds, such as poor activity, poor pharmacokinetics, and toxicity.
* It is used for the optimization of novel drug targets.
* CADD is utilized to locate hits. Through the use of chemical scaffolds, new use of virtual screening for novel drug compounds [6].

**3. DESIGN OF DRUGS BASED ON STRUCTURE**

In order to calculate the interaction energies of each tested chemical, structure-based computer-aided drug design requires knowledge of the target protein structure [7]. The structural database has information on crystallized target proteins. Structure-based design seeks to produce materials that bond tightly and specifically to the target while requiring the least amount of energy [2]. Virtual high-throughput screening, a computer-based screening technique, enables the evaluation of a large number of chemical compounds that are similar to one another for a specific biological activity [8]. Quantitative structure-activity relationship (QSAR) models or pharmacophore mapping can be used to select compounds based on their predicted biological activity, and virtual docking of compounds against desired protein targets is another method of virtual high-throughput screening [9]. It is critical and practical to use computational methods during the lead optimization phase of drug development. Utilizing computational techniques for hit-to-lead optimization, fewer compounds will need to be synthesized and tested in vitro [10].



**Fig. 1. Model for commuting-aided drug design**

1. **DRUG DESIGN BASED ON LIGAND**

Utilizing the knowledge of recognized active and inactive chemicals, ligand-based searches for chemical similarity or quantitative structure-activity relation (QSAR) are conducted. Where the three-dimensional structure of the target proteins is unavailable, ligand-based techniques are ideal.

Computer-Aided Structural Drug Design:

The steps consist of:

* To prepare the target protein and chemical library for docking,
* Selecting a suitable binding position for each drug and ranking the docked molecule structures.

To forecast the orientations or conformations of a receptor-ligand combination, structurally based computer simulation technique known as molecular docking is performed [11]. It is also employed to calculate the complex's molecules' estimated binding affinities.



**Fig. 2. Drug design structure**

**4. MOLECULAR DOCKING TYPES**

* **Search Technique:** The number of configurations created and the binding modes are determined via experimentation. The Monte Carlo approach, fragment-based genetic searches, and systemic searches are used for docking analysis.
1. Rigid Docking
2. Flexible Docking
* **Rigid Docking:** During this docking, the ligand and receptor molecules are fixed. Docking occurs [12].
* **Flexible Docking:** Both the ligand and the receptor are mobile during this docking. It is flexible in conformation. The energy is determined for each rotation. Calculations are made for each conformation surface cell occupancy. The best possible binding stance is then chosen.
* **Scoring Function:** The binding affinity that the binding score directly corresponds to. The highest-rated ligands are the best binders. It may be experimental, based on information, or based on molecular mechanics. Docking A critical component of medication design is scoring:
1. Knowledge-based and
2. Energy component methods
3. Statistics of the observed inter contact frequencies in a sizable database of protein complex crystal structures, using a knowledge-based scoring system.High binding affinities are expected for molecular interactions near the maximum frequency of interactions in the database. Low binding affinity molecular interactions in databases will have low interaction rates [29].
4. Energy component scoring method based on the mathematical assumption that the free energy change that takes place when a ligand binds to a protein target (DG bind) is the sum of the free energies for the interaction between the protein and the ligand, the solvent, the conformational changes in the protein, and the motion of the ligand and protein target during complex formation [13].

**6. DOCKING MECHANICS FOR MOLECULES**

**STEPS**

The intermolecular interaction between two drug molecules was examined using an in silico approach. The macromolecule is the protein receptor. It had an inhibiting effect. The steps in the docking procedure are as follows.

**Step I – Preparation of protein and Ligand:**

Obtaining the protein's three-dimensional (3D) structure from the Protein Data Bank (PDB) of the Research Collaboratory for Structural Bioinformatics. After that, pre-processing should be applied to the downloaded structure. hydrogen atoms added, stabilization of charges, filling of residue spaces, creation of side chains, and removal of water molecules.

**Step II –Ligand Preparation:**

You can obtain molecules from PubChem Ligands using a number of databases, including ZINC. The Chem sketch tool in the Mol file can be used to design it. The ligand molecule was then evaluated using LIPINSKY'S RULE OF 5. The drug's like and unlike compounds are used with it. Due to the molecules' drug-like characteristics, it raises the success rate and lowers failure rates.

**Step III- Grid Generation:**

The location, rotating group, excluded volumes, and restrictions did not change. The primary influencing factor is the nature of the mathematical processes (crossover, migration, mutation). Predictions for the binding cavity must be made.

**Step IV –Prediction of Active site:**

Predictions should be made about the protein's active site. Following protein preparation, water molecules and the cavity is cleared of heteroatoms.

**Step V- Docking:** Analysis of ligand and protein interactions is done. The optimal docking result should be chosen [14].

**Application of Molecular Docking**



**Fig. 3. Application of molecular docking**

**Approaches of Molecular Docking**

There are two techniques utilised for molecular docking.

**Simulation approach**

This process physically separates the ligand and target before allowing the ligand to bind into the target groove after "definite times of moves" in its conformational space. Internal (torsional angle rotations) and outward (rotations and translations) alterations to the ligand's structure are also possible. "Total Energy" is released each time the ligand moves inside the conformational limit. This method is better since it allows for ligand flexibilityIt is also more accurate to assess the molecular recognition between the ligand and the target. This approach takes longer to determine which conformation is best-docked because each conformation dissipates a substantial amount of energy. Recently, this shortcoming has fundamentally been revolutionised to make simulation approaches more approachable through the use of fast optimisation algorithms and grid-based tools [24].

**Shape complementarity approach**

This method uses the surface structure characteristics of the ligand and target to offer their molecular interaction. The molecular surface of the ligand is illustrated in terms of the surface area of the target that is accessible to solvents. Searching the complementary groove for a ligand on the target surface is made easier by the complementarity between two surfaces based on the shape-matching illustration. For instance, the hydrophobicity of protein target molecules depends on the number of twists in the main chain. Our approach rapidly scans thousands of ligands in a matter of seconds to assess a ligand's potential binding capabilities on the target molecular surface [25].

**Fundamental Challenges in Molecular Docking**

Under the following headings, several fundamental docking and scoring difficulties are covered.

**Ligand chemistry**

The way the ligand is prepared significantly affects the docking results because any biomolecule must recognise the ligand in three dimensions and through electrostatic contact. This illustrates the importance of ligand production and ligand structure. When preserving approximate pKa values, the structure was optimised by removing or adding hydrogens. However, there was still a significant gap between the tautomeric and protomeric states of the molecules that needed to be docked. Almost all databases store molecules in their neutral states, even though they are ionized in a physiological setting. As a result, ionizing them before docking is required. However, several programs make it simple to obtain standard ionisation. The question of which tautomer to employ or whether to use all available tautomers still exists in the topic of tautomers [26].

**Receptor flexibility**

An important hurdle in docking is the management of flexible proteins. Depending on the ligand it binds to, the conformation of a biomolecule or protein can change. This illustrates that just one receptor shape may be produced by docking with a stiff receptor. However, the ligands might need to bind to many receptor conformations if a flexible receptor is being employed for docking. The various protein conformational states are frequently the aspect of molecular docking research that is most ignored. Protein flexibility is important because it allows for increased affinities between a particular drug and target. Target flexibility is also influenced by active site water molecules. Water molecules must be adjusted in order to avoid using artefact waters during docking [27].

**Scoring function**

Another challenge for docking is imperfections in the scoring mechanism. A scoring system should be able to discriminate between genuine binding modes and all other parallel modes, much like a search approach may yield the optimal conformation. The analysis of different binding methods would suffer because a hypothetical scoring algorithm would be computationally far more effective [30]. When accuracy is present, several recommendations for determining ligand affinity are made via scoring functions. Systems for calculating scores disregard important concepts like entropy and electrostatic interactions. As a result, the need for an accurate and rapid scoring function is the main bottleneck in molecular docking programming [28].

**Molecular docking:**



**Fig. 4:** An ordinary docking process. The essential processes shared by all docking procedures are shown in this flowchart. The target macromolecule and the small molecule's 3D structures must be selected first, and each structure must then be built in accordance with the docking method being employed. The results of the docking must be examined, and the best-scoring binding modes must be chosen.

Figure 4 depicts the fundamental docking operations that all techniques share. Finding the most beneficial binding mode(s) of a ligand to the target of interest is referred to as docking. The binding mode of a ligand to a particular receptor can be determined using the state variables of the ligand. There are three of these: the ligand's position (represented by the x, y, and z translations), orientation (represented by the axis-angle, the Euler angle, or a quaternion), and if the ligand is flexible, its conformation (expressed by the torsion angles for each rotatable bond) [31]. In a multidimensional search space, each state variable determines one degree of freedom, and its bounds define the scope of the search. When using rigid body docking as opposed to treating the ligand as flexible, the search area is significantly less. Even yet, the chances of discovering a complementary match are reduced if the ligand's shape needs to be altered.

All docking strategies require a scoring system to evaluate the numerous binding modes that could be used, as well as a search method to look into the state variables [32]. Systematic and stochastic search techniques fall into two broad types, while empirical, force field-based, or knowledge-based scoring systems might be used. Deterministic systematic search methods sample the search space at regular intervals. Because dynamic search methods randomly update the state variables throughout each iteration until a user-defined termination criterion is reached, the search results may vary [15]. Sousa et al. study these families of algorithms in more detail. Additionally, the breadth of the search space for each search method can be classified as local or global. Local search methods frequently find the nearest or local minimum energy to the existing conformation, while global search methods hunt for the best or global minimum energy over the preset search space. In comparison to global methods, hybrid global-local search approaches have been shown to be more productive and capable of locating lower energies [16].

For instance, Solis and Wets [17] and Pattern Search [18] are two local search approaches; Monte Carlo (MC) simulated annealing [19] and the genetic algorithm [20] are global search techniques; and the Lamarckian GA (LGA) is a hybrid global-local search strategy. All of these search methods are included in AutoDock 4. In order to compare the chemical characteristics of the ligand and the inverse model of the binding site, DOCK conducts a methodical search. FlexX couples ligand properties with complimentary interaction sites. The GOLD technique to global search makes advantage of GA. The search method used by ICM combines a biassed MC methodology with a local energy reduction strategy.



**Figure 5. Molecular docking process.**

**Molecular docking software**

The three primary categories of molecular docking software. The usage of flexible-rigid docking is common. Flexible docking is typically more precise, however. Therefore recent years have seen a surge in research into this area. The frequently used molecular docking software is included in Table 1, along with its algorithms, evaluation techniques, features, and application domains.

**Molecular docking databases**

The public database Protein Data Bank (PDB) is the most widely used protein structural database. Additionally, it is free to use public databases like ZINC and the PubChem Compound Database. The Compound Database (AcD) and Cambridge Structural Database (CSD) are only examples of the numerous significant commercial databases available.

**Table 1: Representative molecular docking software**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name**  | **Search algorithm** | **Evaluation method** | **Speed** | **Features & Application areas** |
| Flex X | Fragmentation algorithm | Semi-empirical calculation on free energy | Fast | Flexible-rigid docking. It can be used for the virtual screening of small molecule databases using an incremental construction strategy. |
| Glide | Exhaustive systematic search | Semi-empirical calculation on free energy | Medium | flexible docking. This software has XP (extra precision), SP (standard precision), and high throughput virtual screen modes and leverages domain knowledge to restrict the search range. |
| AutoDock | GA (genetic algorithm) LGA (lamarckian genetic algorithm) | Semi-empirical calculation on free energy | Medium | rigid-flexible docking. This programme is free for academic use and is always used with Autodock-tools. |
| ZDOCK | Geometric complement-arity and molecular dynamics | Molecular force field | Medium | Rigid docking. Chen et al. [21] propose a new scoring function that combines pairwise shape complementarity(PSC) with desolvation and electrostatic and develop the ZDOCK server [22] |
| Autodock Vina | GA (genetic algorithm) | Semi-empirical calculation on free energy | Fast | Flexible-rigid docking. AutoDock Vina employs an iterated local search global optimizer and it is faster than AutoDock 4 [23] |

**References:**

1. Satpathy, R. (2020). Application of molecular docking methods on endocrine disrupting chemicals: A review. *Journal of Applied Biotechnology Reports*, *7*(2), 74-80.
2. Bhagat, R. T., Butle, S. R., Khobragade, D. S., Wankhede, S. B., Prasad, C. C., Mahure, D. S., & Armarkar, A. V. (2021). Molecular docking in drug discovery. *Journal of Pharmaceutical Research International*, 46-58.
3. Ekins S, Mestres J, Testa B. In silico pharmacology for drug discovery: applications to targets and beyond. Br J Pharmacol. 2007 Sep;152(1):21-37. doi: 10.1038/sj.bjp.0707306. Epub 2007 Jun 4. PMID: 17549046; PMCID: PMC1978280.
4. Staszak, M., Staszak, K., Wieszczycka, K., Bajek, A., Roszkowski, K., & Tylkowski, B. (2022). Machine learning in drug design: Use of artificial intelligence to explore the chemical structure–biological activity relationship. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *12*(2), e1568.
5. Blanes-Mira, C., Fernández-Aguado, P., de Andrés-López, J., Fernández-Carvajal, A., Ferrer-Montiel, A., & Fernández-Ballester, G. (2022). Comprehensive survey of consensus docking for high-throughput virtual screening. *Molecules*, *28*(1), 175.
6. Bharatam, P. V. (2021). Computer-aided drug design. *Drug Discovery and Development: From Targets and Molecules to Medicines*, 137-210.
7. Seidel, T., Schuetz, D. A., Garon, A., & Langer, T. (2019). The pharmacophore concept and its applications in computer-aided drug design. *Progress in the Chemistry of Organic Natural Products 110: Cheminformatics in Natural Product Research*, 99-141.
8. Tripathi, N. M., & Bandyopadhyay, A. (2022). High throughput virtual screening (HTVS) of peptide library: Technological advancement in ligand discovery. *European Journal of Medicinal Chemistry*, *243*, 114766.
9. Mishra, K., Jain, S. K., & Pant, R. Rational Drug Design and Optimization of New Leads using Modern Quantitative structure-activity relationship (QSAR) Techniques.
10. Bajusz, D., & Keserű, G. M. (2022). Maximizing the integration of virtual and experimental screening in hit discovery. *Expert Opinion on Drug Discovery*, *17*(6), 629-640.
11. Ajjarapu, S. M., Tiwari, A., Ramteke, P. W., Singh, D. B., & Kumar, S. (2022). Ligand-based drug designing. In *Bioinformatics* (pp. 233-252). Academic Press.
12. Fan, J., Fu, A., & Zhang, L. (2019). Progress in molecular docking. *Quantitative Biology*, *7*, 83-89.
13. Raval, K., & Ganatra, T. (2022). Basics, types and applications of molecular docking: A review. *IP International Journal of Comprehensive and Advanced Pharmacology*, *7*(1), 12-16.
14. Corso, G., Stärk, H., Jing, B., Barzilay, R., & Jaakkola, T. (2022). Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*.
15. Sousa, S.F., Fernandes, P.A., and Ramos, M.J. (2006) Protein-ligand docking: Current status and future challenges. Proteins, 65(1): 15–26.
16. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., and Olson, A.J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem, 19: 1639–1662.
17. Solis, F.J. andWets, R.J.-B. (1981) Minimization by random search techniques. Mathematical Operations Research, 6: 19–30.
18. Conn, A.R., Gould, N.I.M., and Toint, P.L. (1991) A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds. SIAM Journal on Numerical Analysis, 28(2): 545–572.
19. Kirkpatrick, S., C. D. Gelatt, J., and Vecchi, M.P. (1983) Optimization by simulated annealing. Science, 220(4598): 671–680.
20. Lambora, A., Gupta, K., & Chopra, K. (2019, February). Genetic algorithm-A literature review. In *2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon)* (pp. 380-384). IEEE.
21. Chen, R., Li, L., & Weng, Z. (2003). ZDOCK: an initial‐stage protein‐docking algorithm. *Proteins: Structure, Function, and Bioinformatics*, *52*(1), 80-87.
22. Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., & Weng, Z. (2014). ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics*, *30*(12), 1771-1773.
23. Huey, R., Morris, G. M., & Forli, S. (2012). Using AutoDock 4 and AutoDock vina with AutoDockTools: a tutorial. *The Scripps Research Institute Molecular Graphics Laboratory*, *10550*(92037), 1000.
24. Lamb ML, Jorgensen WL (1997) Computational approaches to molecular recognition. Curr Opin Chem Biol 1: 449-457.
25. Gschwend DA, Good AC, Kuntz ID (1996) Molecular docking towards drug discovery. J Mol Recognit 9: 175-186.
26. Elokely KM, Doerksen RJ (2013) Docking Challenge: Protein Sampling and Molecular Docking Performance. J Chem Inf Model 53: 1934-1945.
27. Totrov M, Abagyan R (2008) Flexible ligand docking to multiple receptor conformations: A practical alternative. Curr Opin Struct Biol 18: 178-184.
28. Cho AE, Rinaldo D (2009) Extension of QM/MM docking and its applications to metalloproteins. J Comput Chem 30: 2609-2616.
29. Gohlke, H., Hendlich, M., & Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions. Journal of molecular biology, 295(2), 337-356.
30. Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. Nature reviews Drug discovery, 3(11), 935-949.
31. Morris, G. M., & Lim-Wilby, M. (2008). Molecular docking. Molecular modeling of proteins, 365-382.
32. Sousa, S. F., Fernandes, P. A., & Ramos, M. J. (2006). Protein–ligand docking: current status and future challenges. Proteins: Structure, Function, and Bioinformatics, 65(1), 15-26.