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

Poojaben M Prajapati2, Komal G Lakhani1,, Bharat B 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 foretelling and quick access to structural databases have become crucial elements on the medicinal chemist's desktop. Virtual screening is the most significant use of molecular docking. Numerous docking programs were used to visualize the molecule’s three-dimensional structure, and different computational techniques can be used to analyze docking gain. In structural molecular biology and computer-aided drug design, molecular docking is crucial. Docking is helpful for lead optimization because it can be used to do virtual screening on vast libraries of compounds, rate the results, and provide structural hypotheses for how the ligands lower the target.

**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. The drug-receptor interaction indicates the affinity and activity of molecules [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. Currently, rank candidates docking for big libraries compound execute the virtual screening using the scoring method [5].

1. **COMPUTER-AIDED DRUG DESIGN**

"CADD Computer Aided Drug Design" refers to a computer-based method used in computational chemistry to find, improve, or research drugs and related physiologically active molecules. It is most useful in new drug design.

* It provides knowledge about ligands and targets' chemical and biological properties.
* It is used to find and improve novel drugs.
* Discovery of in-silico filters to predict undesirable properties like poor activity and poor Pharmacokinetic, and Toxicity of drug molecules.
* It is used for the optimization of novel drug targets. CADD is being used to find hits
* By using chemical scaffolds to find out novel Virtual screening is applied for new
* drug molecules [6].

1. **STRUCTURE-BASED DRUG DESIGN**

Knowing the target protein structure is necessary for structure-based computer-aided drug design to determine the interaction energies of every molecule that has been tested [7]. Target proteins that have been crystallized are available in the structural database. The structure-based design aims to create substances that bind securely and specifically to the target with the least energy consumption [2]. A computer-based screening technology known as virtual high-throughput screening allows for screening a huge library of chemical compounds comparable to each other for a certain biological activity [8]. Virtual high-throughput screening can take many forms, such as searching for chemically similar compounds, choosing compounds based on their predicted biological activity using quantitative structure-activity relationship (QSAR) models or pharmacophore mapping, and virtually docking compounds against desired protein targets [9]. Using computational techniques at the lead optimization stage of drug development is important and cost-effective. Reduce the number of compounds that need to be synthesized and evaluated in vitro using computational methods in hit-to-lead optimization [10].



**Fig. 1. Commuter-aided drug design model**

1. **LIGAND-BASED DRUG DESIGN**

Ligand-based searches for chemical similarity or quantitative structure-activity relation (QSAR) take advantage of the knowledge of known active and inactive compounds. Ligand-based approaches are ideal where the target proteins' three-dimensional structure is unavailable. Computer-Aided Structural Drug Design:

Steps include:

* For docking Preparation of the target protein and compound library,
* Determining a Proper binding pose for each compound, and
* Ranking the docked structures of molecules.

Molecular docking, a structurally based computer simulation method, is used to predict the orientations or conformations of a receptor-ligand complex [11]. It is also used to estimate the binding affinities between the molecules in the complex.



**Fig. 2. Drug design structure**

1. **TYPES OF MOLECULAR DOCKING**
* **Search Algorithm:** 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:** Both the receptor and the ligand molecule are fixed during this docking. Docking takes place [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. **Knowledge-based scoring function** the statistics of the observed inter contact frequencies in a significant database of protein complex crystal structures. High binding affinities are expected for molecular interactions near the maximum frequency of interactions in the database [80-85]. Low binding affinity molecular interactions in databases will have low interaction rates.
4. **Energy component scoring method** based on the mathematical supposition that the free energy change that occurs when a ligand binds to a protein target (DG bind) is the sum of the free energies for the interaction between the ligand and the protein, the solvent, the conformational changes in the protein, and the motion of the ligand and protein target during complex formation [13].

**6. MOLECULAR DOCKING MECHANICS**

**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:**

Downloading the 3D structure of the protein from the Research Collaboratory Structural Bioinformatics Protein Data Bank (PDB). The downloaded structure should then go through pre-processing. The formation of side chains with hydrogen atoms added, stabilization of charges, filling empty residue spaces, and elimination of water molecules.

**Step II –Ligand Preparation:**

PubChem Ligands molecules can be retrieved utilizing several databases, such as ZINC. It can be drawn using the Mol file's Chem sketch tool. 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 site, rotatable group, excluded volumes, and limitations remained constant. 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 heteroatoms are removed from the cavity.

**Step V- Docking:** Ligand and protein interactions are analyzed. The best docking score should be selected [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**

In this instance, the ligand and target are physically separated before the ligand is permitted to bind into the target groove following "definite times of moves" in its conformational space. Modifying the ligand's structure comprises internal (torsional angle rotations) and external (rotations and translations) changes. Every time the ligand moves within the conformational limit, "Total Energy" is released. This approach is more beneficial because it can accept ligand flexibility. Evaluating the molecular recognition between the ligand and the target is also more accurate. Due to the significant energy dissipation for each conformation, this method takes longer to predict the best-docked conformer. Fast optimization techniques and grid-based tools have recently essentially revolutionized this flaw to make simulation methods more approachable [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 number of twists in the main chain determines the hydrophobicity of protein target molecules. To determine the potential binding capabilities of a ligand on the target molecular surface, our method quickly scans thousands of ligands in a matter of seconds [25].

**Fundamental Challenges in Molecular Docking**

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

**Ligand chemistry**

Because ligand recognition by any biomolecule depends on 3-dimensional orientation and electrostatic interaction, the preparation of the ligand has a significant impact on the docking findings. This demonstrates the significance of ligand structure and ligand preparation. 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**

The handling of flexible proteins presents a significant barrier in docking. A biomolecule's or protein's conformation changes based on the ligand it binds to. This demonstrates that docking with a stiff receptor will only result in one receptor shape. However, when a flexible receptor is used for docking, the ligands may need to bind to many receptor conformations. Different protein conformational states are typically the most overlooked factor in molecular docking research. Because improved affinity can be achieved between a given medication and target, protein flexibility is significant. Active site water molecules are a further factor in target flexibility. To prevent employing artefact waters during docking, water molecules must be corrected [27].

**Scoring function**

Imperfections in the scoring function present another difficulty for docking. Similar to how a search method may provide the best possible conformation, a scoring function should be able to distinguish between actual binding modes and all other parallel modes. A hypothetical scoring function would be computationally considerably more efficient, which would be detrimental to analyzing multiple binding methods. When accuracy is present, scoring functions offer a variety of recommendations for determining ligand affinity. Scoring systems ignore fundamental phenomena like entropy and electrostatic interactions. Therefore, the primary bottleneck in molecular docking programming is the need for an accurate and quick scoring function [28].

**Molecular docking:**



**Fig. 4:** A typical docking workflow. This flowchart shows the key steps common to all docking protocols. The 3D structures for the target macromolecule and the small molecule must first be chosen, and then each structure must be prepared by the requirements of the docking method being used. Following the docking, the results must be analyzed, selecting the binding modes with the best scores.

The essential docking processes shared by all procedures are shown in Figure 4. Docking is known as finding the most advantageous binding mode(s) of a ligand to the target of interest. The state variables of a ligand can be used to identify its binding mode about the receptor specifically. These include 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 (represented by the torsion angles for each rotatable bond). Each state variable specifies one degree of freedom in a multidimensional search space, and its bounds indicate the search size. The search area is substantially smaller for rigid body docking than for treating the ligand as flexible. Still, there is a lesser likelihood of finding a complementary fit if the conformation of the ligand needs to be corrected.

A scoring system to score the many possible binding modes is necessary for all docking techniques, as is a search strategy to investigate the state variables. Search methods can be divided into two main categories: systematic and stochastic, while scoring functions might be empirical, force field-based, or knowledge-based. Systematic search techniques are deterministic and sample the search space at predetermined intervals. The search results can vary because stochastic search methods iteratively update the state variables at random until a user-defined termination criterion is satisfied; Sousa et al. explore these families of algorithms in greater depth [15]. Additionally, each search method's scope of the search space can be categorized as local or global. While global search methods look for the best or global minimum energy throughout the predetermined search space, local search methods often find the closest or local minimum energy to the current conformation. It has been demonstrated that hybrid global-local search approaches outperform global methods by being more effective and able to identify lower energies [16].

For instance, two local search techniques, Solis and Wets [17] and Pattern Search [18], two global search techniques, Monte Carlo (MC) simulated annealing [19] and the genetic algorithm [20], and one hybrid global-local search technique, the Lamarckian GA (LGA), are available in AutoDock 4. DOCK employs a methodical search to compare chemical properties between the ligand and the inverse representation of the binding site. FlexX pairs complementary interaction sites with ligand characteristics. GA is used in GOLD's global search approach. ICM's search strategy combines a local energy reduction approach with a biased MC technique.



**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 software for molecular docking**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **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 uses domain knowledge to narrow the searching range and has XP(extra precision), SP (standard precision) and high throughput virtual screen modes |
| AutoDock | GA (genetic algorithm) LGA (lamarckian genetic algorithm) | Semi-empirical calculation on free energy | Medium | Flexible-rigid docking. This software is always used with Autodock-tools and it is free for academic use |
| 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.