

# Molecular docking and molecular dynamics simulation

Sakshi Singh<sup>1</sup>, Qanita Bani Baker<sup>2</sup> and Dev Bukhsh Singh<sup>3</sup>

<sup>1</sup>Department of Molecular and Human Genetics, Institute of Science, Banaras Hindu University, Varanasi, India, <sup>2</sup>Department of Computer Science, Jordan University of Science and Technology, Irbid, Jordan, <sup>3</sup>Department of Biotechnology, Institute of Biosciences and Biotechnology, Chhatrapati Shahu Ji Maharaj University, Kanpur, India

## 18.1 Introduction

Computer-aided drug design methods have been applied in the area of drug discovery for the past two decades. Molecular modeling, docking, and simulation strategies relied on a rigid view of the receptor–ligand interaction mechanism with the help of computational resources (Meng, Zhang, Mezei, & Cui, 2011; Morris & Lim-Wilby, 2008). The concept of the molecular docking approach is to how two or more molecular structures (protein or enzyme, nucleic acid, and small lead/drug) fit together (Ferreira, Dos Santos, Oliva, & Andricopulo, 2015). Protein–ligand (small molecule), protein–nucleic acid, and protein–protein docking play an essential role in predicting the ligand’s orientation and binding affinity in the active site of the target protein. The molecular docking studies revealed the intramolecular interaction of small molecules at a binding site of target protein (receptor) (Singh, Chaturvedi, & Rai, 2020). The prediction of binding pocket deduces the functionality and active site domain of the protein, which can obtain the crucial interaction information for computer-aided drug designing (Agoni, Olotu, Ramharack, & Soliman, 2020; Singh et al., 2020). The properties of binding affinity characterize the binding strength of a ligand with a target molecule. Protein–protein docking has been applied to predict the complex structure from known structures of the individual proteins (Vakser, 2014). Protein–DNA or protein–RNA interaction play important role in the biological process, replication, transcription, splicing, and protein synthesis. The results of molecular docking depend on binding energies, the number of hydrogen bonds, and potential hits found in the protein–ligand complex structure (Singh, 2020). (Wójcikowski, Ballester, & Siedlecki, 2017; Singh, 2020). However, docking is unable to capture the conformational changes and flexibility this remains a major challenge for structure-based drug designing (Huang & Zou, 2010). Molecular dynamics (MD) simulation approach can address this challenge and forecast the time-dependent behavior of a molecular system (Adcock & McCammon, 2006). It is a thermodynamic-based method, which is widely used to understand the receptor–ligand interaction complex, or docked complex by conformational detail at the atomic level (Di Nola, Roccatano, & Berendsen, 1994; Durrant & McCammon, 2011; Singh & Pathak, 2020). This method is useful in studying the energy landscape of protein–ligand interaction and defining their conformational changes, which are not usually available by high-resolution experiments. It is also valuable for the structural refinements of postdocking complexes, which identify the complementarity between the receptor–ligand complexes and enhance the complex state, allowing rescoring of the docked complex. MD simulation could investigate the biomolecular processes, that is, protein–ligand, protein–nucleic acid, and protein–peptide binding, and their conformation changes reveal the orientation of all-atomic resolution at femtosecond (fs) (Dror, Jensen, Borhani, & Shaw, 2010; Hollingsworth & Dror, 2018). The simulations could expose the involvement of biological disorders (including the impact of a mutation, effect of phosphorylation, protonation state, and the presence or absence of a ligand in binding site) depending on the protein–protein, protein–ligand, protein–peptide, or protein–nucleic acid complex (Hollingsworth & Dror, 2018).

## 18.2 Molecular docking

### 18.2.1 Algorithms

There are several algorithms behind the various developed molecular docking programs (Dias, Timmers, Caceres, & de Azevedo, 2008). In this section, we give an overview of the most known algorithms and the advantage of each one. Moreover, we present the main programs that are built based on each algorithm. One or more specific search algorithms can be used in docking programs. Summary of the algorithms, programs based on these algorithms, advantages, and limitations of some docking methods is listed in Table 18.1.

#### 18.2.1.1 Matching algorithm

A matching algorithm is one of the simplest algorithms that consider the geometric overlap between two molecules (Meng et al., 2011). Several approaches are deployed to make several alignments between receptor and ligand. The matching algorithm approach is also widely used as a part of flexible docking methods such as in the DOCK program (Ewing et al., 2001). For example, the DOCK program's first steps are defining the regions of the binding site where the potential ligand could be located and called sphere centers that are identified using a Matching algorithm. This approach is deployed also in many other well-known programs, such as DOCK (Ewing et al., 2001). It is also called shape-matching algorithms.

#### 18.2.1.2 Monte Carlo algorithm

Monte Carlo (MC) methods are used in many docking software, such as AutoDock, MCDOCK, AutoDockVina, QXP, and ROSETTALIGAND (Meiler & Baker, 2006). The basic concept behind MC is the acceptance or the rejection of the random changes by using the Metropolis criteria (Hastings, 1970). It uses a conformational searches approach with potentials of molecular affinity for a ligand to dock with a structure (Goodsell & Olson, 1990). The MC methods are based on the MC algorithm of the Metropolis, which provides an acceptance criterion in the docking search evolution. A random variation of the ligand degrees of freedom is done at any iteration of the algorithm. Then, the change is accepted when the energy score of the binding pose is increased; otherwise, it is accepted as per the probability ( $P$ ) expressed in the following equation:

$$P \sim \exp[-(E_1 - E_0)/k_B T]$$

where  $k_B$  is the Boltzmann constant,  $T$  is the system temperature, and  $E_1$  and  $E_0$  are the energy score before and after system modification. Most MC simulation-based programs provide an accurate and precise performance (Essex, Severance, Tirado-Rives, & Jorgensen, 1997). However, the MC algorithm is not suitable for time-dependent approaches, such as in MD simulations.

#### 18.2.1.3 Genetic algorithm

The genetic algorithm (GA) is one of the well-known optimizing techniques used in several applications and it is widely used to solve the docking problem (Li et al., 2015; Whitley, 1994). The genetics algorithm is adapted to solve the docking problem (Jones et al., 1997; Tüzün, Yavuz, & Saripinar, 2018). The GA technique was used to find the best parameters that affect the activity of the studied drug molecule and to complete the calculations quicker. The idea behind GA is inspired by Darwin's theory of natural evolution where a genetic operator is used to combine two chromosomes (parents) to produce a new chromosome that could be better than parents (Katoch, Chauhan, & Kumar, 2020). This process includes many scoring functions (SFs) along with a set of parameters, such as crossover rate and mutation rate.

#### 18.2.1.4 Particle swarm optimization

Kennedy and Eberhart (1995) invented the particle swarm optimization (PSO) algorithm. Swarm optimization techniques are one metaheuristic used in molecular docking used as a search approach for the docking problem where the docking problem is formulated as a parameter optimization problem that is associated with a well-defined SF (García-Godoy, López-Camacho, García-Nieto, Nebro, & Aldana-Montes, 2015). The optimization approach aims to find the docked conformation of a ligand with minimum energy. PSO has emerged as a fast and accurate approach used to solve a complex search problem. There are several swarm-based docking programs as SODOCK (Chen, Liu, Huang, Hwang, & Ho, 2007). PSO principle is well suited for the molecular docking problem where the ligand value needs to be minimized according to the SF.

**TABLE 18.1** Docking algorithms, tools, their advantages, and limitations.

Algorithms	Main advantages	Limitations	Programs example
Matching algorithm	Advantage of speed	It requires a prior knowledge of receptor coordinates details. The lack of molecular flexibility affects the accuracy.	DOCK 4.0 (Ewing, Makino, Skillman, & Kuntz, 2001); Ph4DOCK (Goto, Kataoka, & Hirayama, 2004); GM-DockZn (Wang et al., 2020)
Monte Carlo (MC)	Accurate and precise performance	It is not suitable for time-dependent approaches such as in molecular dynamics simulations.	MCDOCK (Liu & Wang, 1999); PRODOCK (Trosset & Scheraga, 1999); RiboDock (Mihai, 2017); AutoDock (Goodsell, Sanner, Olson, & Forli, 2021)
Genetic algorithm (GA)	Applicable for rigid and flexible ligand, take the benefit of parallelization	It is not suitable for time-dependent approaches, such as in molecular dynamics simulations. The uncertainty of convergence is an issue in the GA.	GOLD 3.1 (Jones, Willett, & Glen, 1995); GasDock (Li et al., 2004); PSI-DOCK (Pei et al., 2006); Autodock 4.0 (Trott & Olson, 2010)
Particle swarm optimization (PSO)	A fast heuristic method is more reliable	The uncertainty of convergence is an issue in PSO.	Autodock (Goodsell et al., 2021); ClustMPSO (Janson, Merkle, & Middendorf, 2008)
Tabu search	Metaheuristic algorithm prevents the previously considered solutions from being revisited and enables new queries	It is used for the calculation of the mean square error.	SFDock (Fogel, Cheung, Pittman, & Hecht, 2008); Pro leads (Fogel et al., 2008)
Incremental construction	Fragments the ligand and separately docks it at the active site of the receptor	It is failed if ligands have more bonds that are rotatable.	eHiTS (Zsoldos, Reid, Simon, Sadjad, & Johnson, 2007); DOCK 4.0 (Pagadala, Syed, & Tuszynski, 2017); FlexX (Kellenberger, Rodrigo, Muller, & Rognan, 2004)
Simulated annealing	Flexibility and conformational analysis Specific kind of dynamic simulations Combined with the MC method give higher accuracy results	Time consuming due to the cycle of simulations. It is required a combination of GA and MC algorithms for accurate results.	MolDock (Thomsen & Christensen, 2006); AutoDock4 (Morris & Lim-Wilby, 2008); AutoDockVina (Trott & Olson, 2010); ROSETTA3 (Leaver-Fay et al., 2011)
Molecular dynamics	The stability of the docked complex is more effectively represented	Time consuming and computational power. Difficult over high barrier energy conformation.	AUTODOCK VINA (Trott & Olson, 2010)
Multiple Copy Simultaneous Search (MCSS)	Multiple copies of functional ligand docked with receptor binding site Energy minimization of functional ligands before docking.	MCSS is applicable for rigid receptors only.	HOOK (Eisen, Wiley, Karplus, & Hubbard, 1994); FlexX (Kellenberger et al., 2004)
Evolutionary programming (EP)	Based on heuristic search algorithm, and apply differential evolution for cavity prediction	Only for flexible docking.	MolDock (Thomsen & Christensen, 2006); GOLD (Jones, Willett, Glen, Leach, & Taylor, 1997); AutoDock (Goodsell et al., 2021); DIVALI (Clark, 1995); DARWIN (Taylor & Burnett, 2000); PSI-DOCK (Guedes, Pereira, & Dardenne, 2018; Pei et al., 2006); FLIPDock (Zhao & Sanner, 2007); Lead finder (Stroganov et al., 2008); EADock (Grosdidier, Zoete, & Michielin, 2011)

(Continued)

**TABLE 18.1** (Continued)

Algorithms	Main advantages	Limitations	Programs example
Fast Fourier transform algorithm	Rigid body protein docking Expressing the interaction energy in protein–ligand orientation Global docking search on a protein–ligand system	It provides a limited view of frequencies in the sense of signal processing.	ZDOCK server (Chen, Li, & Weng, 2003); Patchdock (Schneidman-Duhovny et al., 2005); HexServer (Macindoe et al., 2010)
LUDI	The fragment-based approach provides unbiased fragments to the active site of the receptor, utilizes the protein–ligand complex hydrogen bond	Finding a template docked with fragments is hard.	FlexX (Pagadala et al., 2017)

## 18.2.2 Scoring functions

The SF is used for predicting the docked orientation, which identifies the intermolecular complex structure. The SF is also used to rank how much ligand is relative to another. Designing the SF is a crucial aspect and is a fundamental step (Xu, Huang, & Zou, 2018). Scoring functions affect the correctness of ranking the candidate dockings (Guedes et al., 2018). The docking reliability depends on the accuracy of the SFs, which is used to define the mode of binding and the ligand site. The SF is also used to identify the potential drugs that lead to a protein target. While finding a rapid and accurate prediction is still a challenging task in molecular docking.

### 18.2.2.1 Force-field SFs

The force-field function is one of the SFs developed based on the physical interaction at the atomic level, such as bond angles, bond length, and torsions (Vanommeslaeghe & Guvench, 2014). Usually, force-field functions and related parameters are obtained from the mechanical calculation based on the physics principles along with the experimental data. AutoDock (Goodsell et al., 2021), DOCK (Ewing et al., 2001; Pagadala et al., 2017), GOLD (Pierce, Sandretto, & Bemis, 2002; Verdonk et al., 2004), and *D*-score (Ewing et al., 2001) are the examples of force-field SFs.

### 18.2.2.2 Empirical SFs

Empirical SFs are employed to identify the best molecular docking structure. It estimates the binding affinity by finding the summation of the important energetic factors in the protein ligand (Guedes et al., 2018). Several studies employed these empirical SFs, the study compared a set of proposed empirical SFs and using a program called POLSCORE (Dias, de Azevedo, & Walter, 2008) with the other two SF programs: DrugScore (Gohlke, Hendlich, & Klebe, 2000) and X-SCORE (Wang, Lai, & Wang, 2002). The overall comparison showed that POLSCORE is better in accurately predicting the docked position. Usually, the empirical SF approach has employed a training set for unknown binding affinities to find the best weights for the energetic factors. This can be done by using several optimization techniques, such as linear regression analysis (Ashtawy & Mahapatra, 2015; Guedes et al., 2018). *F*-Score (Rarey, Kramer, Lengauer, & Klebe, 1996); ChemScore (Essex et al., 1997), and LUDI (Böhm, 1994) are some other examples for empirical SFs.

## 18.2.3 Knowledge-based SFs

Knowledge-based SFs generally take advantage of the structural information in the known protein–ligand complex (Li, Sze, Lu, & Ballester, 2020). It is also known as statistical potential based SFs (Rykunov & Fiser, 2010). Knowledge-based SFs are energy potentials derived from the structural details collected from atomic structures that have been experimentally calculated. Several knowledge-based SFs were developed and used to predict protein structure and in protein–ligand. Based on several types of research, comparing knowledge-based SFs to the previously mentioned approaches [force field (Section 18.2.2.1) and empirical SFs], knowledge-based SFs prove their potential to have a good balance between the efficiency (speed) and the accuracy (Guedes et al., 2018; Nguyen & Wei, 2019). It also

shows robustness relative to the training set. PMF (Muegge, 2006), Smog (DeWitte & Shakhnovich, 1996), Bleep and DrugScore (Velec, Gohlke, & Klebe, 2005) are examples of knowledge-based SFs.

### 18.2.3.1 Machine learning-based SFs

There are various Machine Learning (ML)-based SFs (ML-SFs) that can be incorporated with docking algorithms and to build the SFs (Wójcikowski et al., 2017). ML-SFs generally outperform other classical SFs and they are used for rescoring and increasing the accuracy of SFs. ML-SFs depend on the training data sets when they are from a supervised type (Li et al., 2020). There are several ML approaches used to implement ML-SFs, such as Support Vector Machine (SVM), Random Forest (RF), and convolutional neural network (CNN) (Wang et al., 2017).

## 18.3 Docking methodologies

### 18.3.1 Flexible docking

Flexible docking based on the induced-fit model allows flexibility in the binding pose prediction of the protein–ligand, protein–protein, or peptide–protein interactions (Huang & Wong, 2009; Lexa & Carlson, 2012). As compared to rigid docking, flexible docking considers flexibility for a ligand to induce certain alterations/orientation changes in the side chains of residues present at the binding site of the target protein. During flexible docking, integration of multiple receptor conformations corresponds to highly multidimensional data sets. It enables the potential energy surface as multiple coordinate functions due to the many degrees of freedom related to residues of flexible proteins structure. There are two different algorithms used in this approach: systematic, incremental construction, and stochastic. Incremental construction develops a binding pose prediction based on ligand–receptor interaction. DOCK (Pagadala et al., 2017), LUDI (Böhm, 1994), FlexX (Kellenberger et al., 2004), Glide (Repasky, Shelley, & Friesner, 2007), Surflex (Jain, 2003), and Hammerhead programs (Welch, Ruppert, & Jain, 1996) apply this algorithm. Stochastic or probabilistic algorithms are used to selectively accept or reject configurations across the criteria spectrum under which computational efforts are optimized. AutoDock (Goodsell et al., 2021), MC (Pagadala et al., 2017), GOLD (Pagadala et al., 2017), and DARWIN (Taylor & Burnett, 2000) employ this algorithm. Recently, the DockThor program is available for highly flexible ligand docking on the protein–peptide dataset.

### 18.3.2 Semiflexible docking

In semiflexible docking, the range of protein conformation differs; therefore it is appropriate to deal with the protein and small molecule/ligand interactions (Andrusier, Mashiach, Nussinov, & Wolfson, 2008; Pagadala et al., 2017). Ligand is allowed to be flexible with protein around some torsional angles while the protein configuration is fixed in rigid docking (Andrusier et al., 2008). Small molecule structure configuration changes while macromolecules remain rigid or retain any of the rotatable amino acids' residues to ensure computational efficiency.

### 18.3.3 Virtual screening of high-throughput docking

Virtual screening is a set of computational strategies that allow analyzing the large collections of compounds or databases to identify the potential target candidates (Fig. 18.1) (Lionta, Spyrou, KVassilatis, & Cournia, 2014; Meng et al., 2011). It is divided into two groups: (1) structure/receptor based, which focuses on the known protein receptor coordinates; (2) ligand based, which relies on the information derived from one or more active ligands. Top scoring virtually screened compounds can be synthesized for their activity evaluation in in vitro conditions using inhibition assays (Singh & Dwivedi, 2019; Singh, Gupta, Kesharwani, & Misra, 2013). The structure of the protein–ligand complex can be determined using X-ray crystallography to decipher the information about interacting residues. The protein–ligand interaction information can be used to optimize the binding interaction/affinity using linkers to grow it on the lead to improve interactions in the unutilized region of the binding site. Application of ML plays an important role in drug discovery and has also been applied in virtual screening. Currently, deep learning (DL) algorithms are used for binding mode prediction that applies docking ranking as inputs in combination with docking structures.

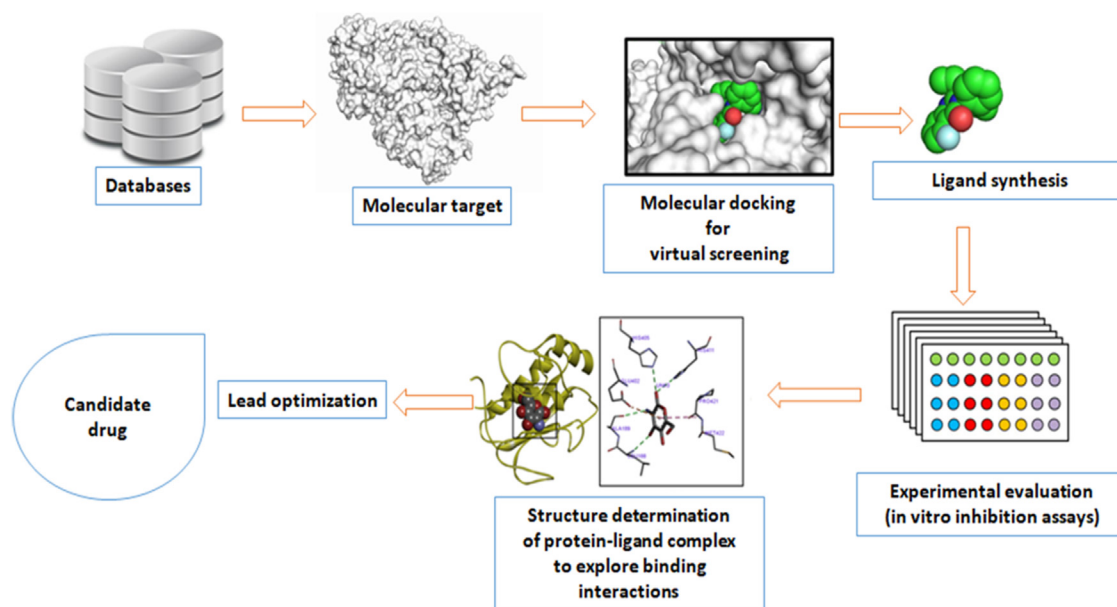


FIGURE 18.1 Ligand-based virtual screening, synthesis, and protein–ligand complex analysis for optimizing the binding interactions and affinity.

### 18.3.4 Fragment docking

Fragment-based drug designing is challenging because of false positive hits, incorrect scoring, compared with drug-like molecules, and problems in multiple binding modes (Bian & Xie, 2018; Istyastono et al., 2015). The development of fragment-specific approaches is an emerging research topic. Fragment docking assists the growth of fragment hits by predicting the binding pose of the proposed compounds. It is effective for high-throughput screening and opens a new avenue for drug designing (Jacquemard, Drwal, Desaphy, & Kellenberger, 2019). To identify correct alignments between a multi-fragment ligand and a set of docked poses of different types of fragments, the graph theory maximum clique algorithm is used. De novo modeling software takes advantage of an established binding mode of a fragment to suggest modified analogs with enhanced binding affinities, defined computationally or experimentally. For example, GANDI (Dey & Cafilisch, 2008), BREED (Pierce et al., 2004), LigBuilder (Wang et al., 2002), Autogrow (Durrant, Lindert, & McCammon, 2013), and ADAPT (Srinivas Reddy, Chen, & Zhang, 2013) tools are used for fragment docking methods.

### 18.3.5 Machine learning in docking

The accuracy of docking results depends upon the efficiency of the SF, which evaluates binding affinity prediction. The prediction of binding affinity utilizes the atomic coordinates of the receptor–ligand complex structure (Pantsar & Poso, 2018; Raschka & Kaufman, 2020). Consequently, the current reliability of the SF is not satisfactory enough. Thus an ML- or DL-based approach has been proposed several results to improve the SF (Raschka & Kaufman, 2020; Wang et al., 2017). ML is a field that focuses on developing algorithms that allow computers to learn from representative datasets. Supervised, unsupervised, and reinforcement learning are the three main subcategories (Patel, Shukla, Huang, Ussery, & Wang, 2020; Raschka & Kaufman, 2020). In supervised, the goal is to predict a category mark (classification) or score (regression analysis) from a large set of such classified instances (Guedes et al., 2018). It is concerned that supervised learning is about the functional mapping between so-called input features or observations (such as small molecule fingerprint representations) and a discrete or continuous target variable, such as an active/inactive mark or a binding affinity in the sense of a particular receptor.

Most likely, RF and CNNs allow improving the accuracy of binding affinity prediction of the docked complex (Zheng, Fan, & Mu, 2019). RF-Score is the first ML-SF to achieve a significant improvement in binding affinity prediction over classical SFs. RF-Score uses regression to refer its binding affinity to a structural overview of the complex. Based on the several results achieved, it was indicated that in predicting scores that correspond to experimentally derive binding affinities, their approach outperformed standard docking programs. In addition; they also showed that their SF prediction accuracy improved with the use of larger datasets for correction of the RF-based SF (Ashtawy & Mahapatra, 2015). In 2D and 3D image recognition tasks, CNNs were very effective, which prompted us to apply them to the

recognition of docking decoy hits. 3D CNNs are applied to drug-protein interaction scoring in the bioinformatics area, protein functional site analysis, consistency evaluation of single protein structure models, and secondary structure detection in cryo-EM maps. ML/DL techniques, such as SVM, RF, and CNN, have been widely adopted in various stages of drug development (Ashtawy & Mahapatra, 2015; Carpenter & Huang, 2018; Zheng et al., 2019). Novel drug targets and ligands, new binding site prediction, the relationship between targets and disease, lead optimization, de novo molecular design, pathways, and properties of modified molecules are the stages in which ML is applied.

### 18.3.6 Docking tools and their features

Docking tools are based on the algorithms (GA, simulated annealing, MC algorithm, and incremental construction), SFs (force field, empirical, and knowledge based), methodologies (rigid, flexible, and fragment), and several other factors. Summary of molecular docking tools along with the benchmark has been listed (Table 18.2).

## 18.4 Molecular dynamics simulation

The MD simulation addresses the velocities and position of atoms using Newton's motion laws (Dror et al., 2010). A simulation can identify interactions between all the system peripherals at the atomic level, functioning as a “computational microscope” (Ingólfsson, Arnarez, Periole, & Marrink, 2016). MD simulation uses classical equations of motion to understand the dynamic perturbation in the protein structure over a given period of time. Trajectories related to simulation are visualized in the form of a graph to interpret the different properties of the simulated system for a given time and residue. MD simulation is also used to interpret the conformational states of a protein at different times. It provides the dynamic and thermodynamic properties, such as binding-free energy, pressure, and temperature-related information (Adcock & McCammon, 2006; Hollingsworth & Dror, 2018). It is often used to identify equilibrium distributions among different conformations, determine novel conformations, and characterize the changes in structural conformational distribution, resulting in ligand binding or structural mutation (Fig. 18.2).

MD simulation study also provides insight into cryptic pockets or allosteric-binding sites present in a protein (Durrant & McCammon, 2011). Many force fields, such as AMBER, CHARMM, OPLS, and GROMOS, are widely

**TABLE 18.2** Some important docking tools and their descriptions.

Docking tool	Description	References
DOCK	The algorithm of this tool is to find the lowest-energy binding mode; including force-field-based scoring function, on-the-fly optimization, an upgraded rigid-body docking matching algorithm, and a flexible ligand-docking algorithm	Ewing et al. (2001) and Pagadala et al. (2017)
GOLD	Lead optimization, high-throughput virtual screening, and predict the correct binding mode of active molecules	Pierce et al. (2002) and Verdonk et al. (2004)
D-Score	Force-field-based scoring function	Ewing et al. (2001)
AutoDock	Automated docking: it is intended to predict how ligands, such as substrates or drug candidates, bind to a known 3D structure receptor	Goodsell et al. (2021)
DrugScore	Knowledge-based scoring function; evaluate protein–protein interaction complex	Gohlke et al. (2000)
ChemScore	Empirical scoring functions; protein–ligand docking based on the distance parameter	Eldridge, Murray, Auton, Paolini, and Mee (1997)
X-SCORE	Empirical scoring function; measures binding affinity of the protein–ligand complex	Wang et al. (2002)
F-Score	Fast flexible docking method using an incremental construction algorithm	Rarey et al. (1996)
LUDI	De novo design of enzyme inhibitors	Böhm (1994)
PMF Score	Flexible ligand docking; knowledge-based scoring function; give a score to binding pose generated by docking	Muegge (2006)
Smog	To determine protein–ligand interaction, the scoring function utilizes pair-wise atom potentials	DeWitte and Shakhnovich (1996)

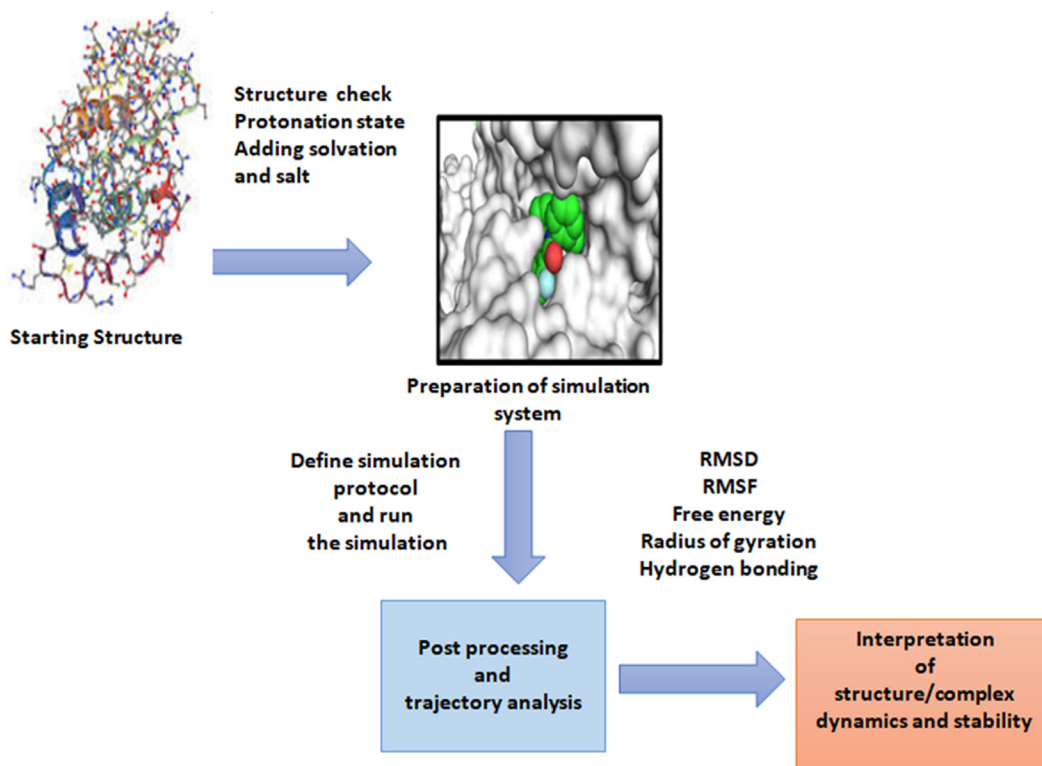


FIGURE 18.2 Major steps involved in molecular dynamics simulations analysis.

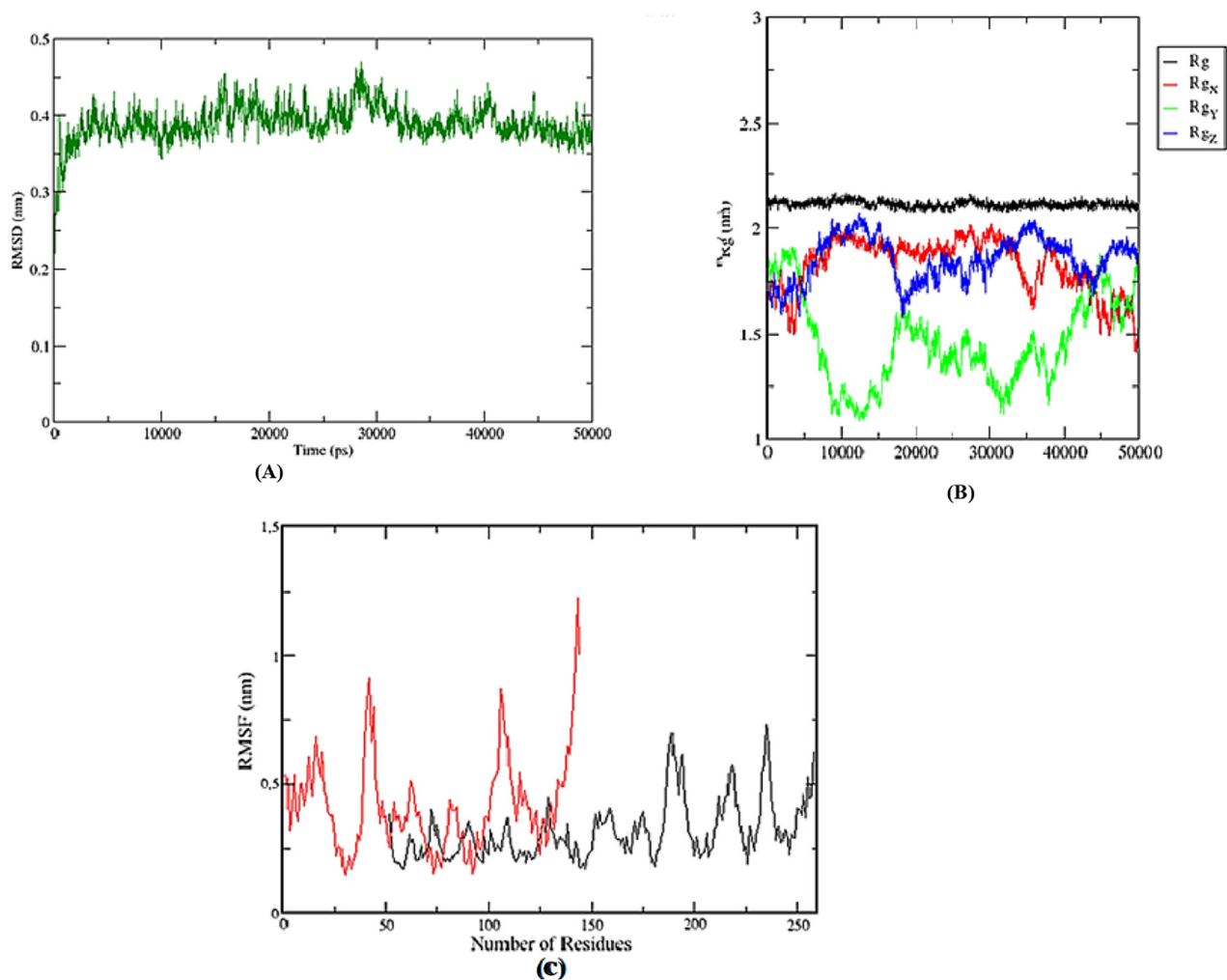
used in the MD simulation. The trajectories generated by simulation provide protein conformations that could be used for multiprotein conformations docking applications. The MD simulation trajectory analysis includes properties, such as root mean square deviation (RMSD), root mean square fluctuation (RMSF), free energy landscape (FEL), and principal component analysis, that determine the consistency of the structure of the protein (Fig. 18.3).

RMSD measures the average displacement of the atoms relative to a reference structure at the moment in simulation. It is effective in studying the time-dependent motions of the protein–ligand structure (Kufareva et al., 2011). This is often used to assess whether a protein–ligand system is stable or deviates from the initial coordinates within the time scale of simulations. The RMSF analysis measures the structural fluctuations (Fracalvieri, Pandini, Stella, & Bonati, 2011). The RMSF plot represents the residues wise fluctuation, and a high RMSF score for a residue indicates the high perturbation at that location.

The radius of gyration (Rg) measures the compactness of the protein or protein–ligand complex and can also be used to understand the conformational changes/variation in protein (Ivankov, Bogatyreva, Lobanov, & Galzitskaya, 2009). It represents the compact folding of secondary structures into the 3D protein structure, and a lower Rg score indicates that protein is very compact in structure (Rai, Pathak, & Singh, 2021). In a time-dependent manner, structural conformation changes explain the stability of the protein structure. The FEL represents the free energy of conformations at different stages of binding interaction and can be used to interpret the ligand binding and folding dynamics of a protein or protein–ligand complex (Chong & Ham, 2019; Singh & Dwivedi, 2016). The application of MD simulations is used to validate the results of molecular docking. However, the simulation of a ligand's binding with a protein under physiological conditions is a very flexible docking process. The possibility of examining the entire binding mechanism could provide insights into the metastable states searched by the ligand during the interaction, water position during binding, and conformational rearrangements prior to concurrent or consecutive to binding during simulation alternative binding sites.

#### 18.4.1 Postdocking refinement

This approach can optimize docking poses by molecular mechanics (MM), and MD simulations and estimate-binding affinities by using the binding-free energy calculation with MM-PBSA and MM-GBSA methods. Binding Estimation After Refinement (BEAR) is a postdocking processing approach used for the refinement of ligand-binding modes



**FIGURE 18.3** Molecular dynamics simulation trajectory analysis: (A) RMSD plot, (B) Gyration plot, and (C) RMSF plot.

predicted by docking tools (AutoDock and LibDock) and evaluates the binding-free energies performed with MM-PBSA and MM-GBSA algorithms. The computational workflow of the BEAR tool initially preprocessed the docked molecules by generating the ligand topology, receptor topology, and complex topology (Fig. 18.4). Then, refinement of the complex was performed using iterative three steps based on MM energy minimization and MD simulation. Afterwards, the estimation of binding-free energy of the docked complex was calculated with the use of MM-PBSA and MM-GBSA algorithms. In drug development, postdocking refinement and rescoring methods provide higher hit rates in virtual screening and facilitate better correlation with experimental results.

#### 18.4.2 Binding-free energy calculations: MM-GBSA/MM-PBSA

Binding affinity characterizes the efficiency of protein–ligand, protein–peptide, and protein–protein docking (Brandsdal et al., 2003). The absolute binding-free energies calculation requires several accurate approaches and an efficient computational system. The generalized born and solvent-accessible surface area solvation (MM-GBSA)/Poisson–Boltzmann surface area (MM-PBSA) MD model can be represented as a free-energy endpoint form, as the first and last states of the system are measured to determine a change in free energy (Homeyer & Gohlke, 2012). The accuracy of the MM-GBSA/MM-PBSA depends on the relationships formed between statistical thermodynamics and end-point free-energy models (Fig. 18.5). This approach has some drawbacks, including the fact that water interactions in complex, insensitive to the trajectory, and susceptible to induced fit effects are not taken into account (Huang & Wong, 2009). MM-GBSA/MM-PBSA approach is the postanalysis approach to determining the binding-free energies of

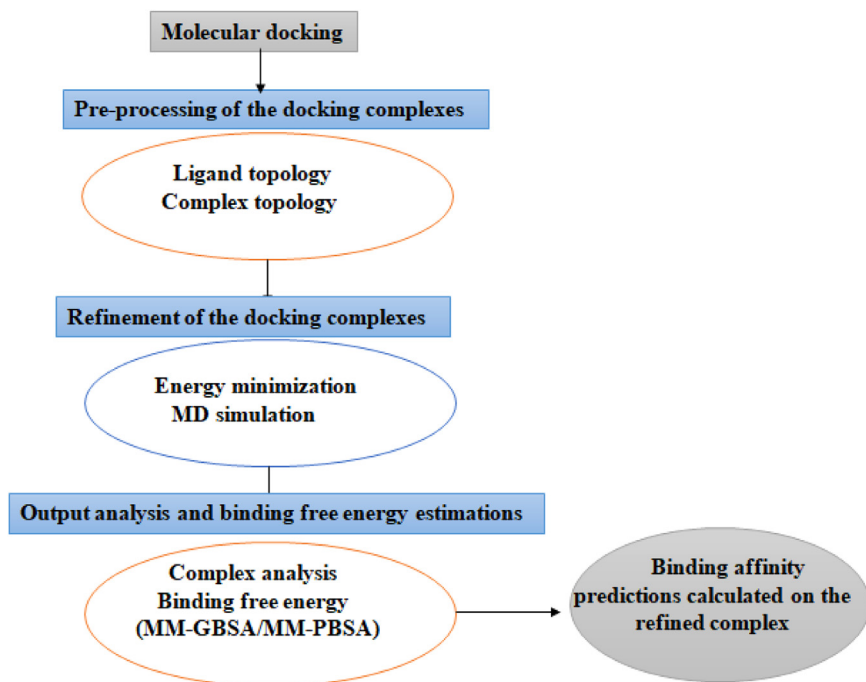


FIGURE 18.4 The computational workflow of binding affinity estimation after refinement.

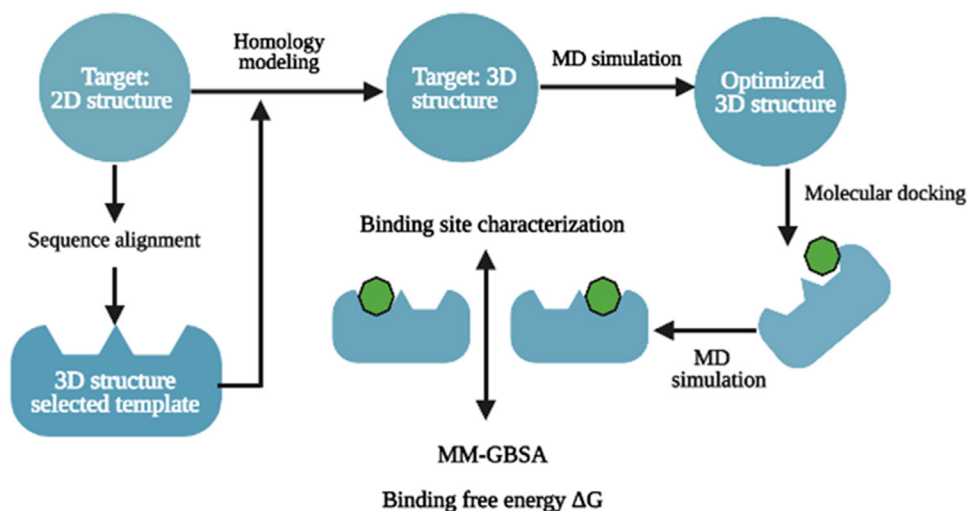


FIGURE 18.5 A simple flow representing the steps from modeling, docking, and molecular dynamics simulation for binding-free energy calculation.

molecular complexes (Genheden & Ryde, 2015). As given below, the method divides free energy into molecular mechanical energies, continuum solvation energies, and solvent entropy terms.

$$\Delta G_{\text{binding}} = G_{\text{complex}} - (G_{\text{receptor}} + G_{\text{ligand}})$$

where  $G_{\text{complex}}$  means the total free energy of the protein–protein complex and  $G_{\text{receptor}}$  and  $G_{\text{ligand}}$  are the receptor and ligand total free energies of a solvent, respectively.

## 18.5 Challenges in molecular docking and MD simulation techniques

The major challenge of molecular docking is the accuracy of the SF to evaluate the binding energy of receptor–ligand interaction. The fact is that some intermolecular interactions, such as solvation effect and entropy change, are hardly predicted or rarely considered in SFs. The second unresolved issue to deal with a water molecule in the binding pocket

is challenging and needs attention. For example, the X-ray crystal structures inefficient hydrogen coordinate information due to the lack of scattering of smaller atoms (Nittinger, Schneider, Lange, & Rarey, 2015). The exact position of hydrogen in structure is not known, which contributes to inaccuracies in the analysis of water molecules that could serve as a bridge molecule between receptor and ligand. Moreover, there is no valid theoretical method available to predict reliably how strongly water molecules affect receptor–ligand interaction. The other challenge in docking is to predict known protein binding poses. It is not straightforward to choose the docking tool that will provide the best results for the given target. The accurate prediction of the binding pose is a stringent task in molecular docking. The major issues in MD simulation are how to get useful information from a large amount of data. MD simulation is based on experimental models that provide the composition of a given system and initial coordinates from resolve molecular structures. There are challenges in the connection between simulation and experimental data, such as experimental conditions, and scales are quite different from what has been simulated. It is exciting to expand simulations to a cellular-scale system, but it poses a significant challenge in utilizing such efforts' vast data. Understanding these challenges will be effective in the data analysis and interpretation of the biological systems at a large scale.

A glance at the wide variety of studies we examined shows that docking is an important method that produces multiple success stories in the process of drug development and prediction of side effects. This complements the novel methods or can also be used to define potential targets that are unknown. Because of the constant improvements in computing power, the research is increasingly progressing and extending its functional applications, making docking facilities accessible online, thereby enabling the computation to be performed by remote computers, and allowing the user to visualize and access the effects of the docking. However, issues, such as developing target structure databases, computational efficiency, receptor flexibility, availability of efficient search algorithm, and accurate SF for docking, need to be solved.

## 18.6 Conclusion

Molecular docking and MD simulations are very important techniques to understand the binding interaction of a ligand molecule with a drug target. MD simulations have many additional advantages over docking as it considers the many physiological parameters important to predict the real mode of interactions. MD simulation is also important due to its capability to explain the dynamic perturbation in a molecular system. Several docking tools are available for the protein–ligand docking but all these tools do not lead to a common statistics-related binding analysis as they all use a different approach and SF for the binding energy and affinity calculations. MD simulation tools are widely used for understanding the binding dynamics of a ligand molecule with a protein target, protein unfolding problems, conformational/compactness analysis of the molecular system, and impact of mutations, or drug resistance.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

- Adcock, S. A., & McCammon, J. A. (2006). Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chemical Reviews*, 106, 1589–1615.
- Agoni, C., Olotu, F. A., Ramharack, P., & Soliman, M. E. (2020). Druggability and drug-likeness concepts in drug design: Are biomodelling and predictive tools having their say? *Journal of Molecular Modeling*, 26, 1–11.
- Andrusier, N., Mashiach, E., Nussinov, R., & Wolfson, H. J. (2008). Principles of flexible protein–protein docking. *Proteins: Structure, Function, and Bioinformatics*, 73, 271–289.
- Ashtawy, H. M., & Mahapatra, N. R. (2015). Machine-learning scoring functions for identifying native poses of ligands docked to known and novel proteins. *BMC Bioinformatics*, 16(6), 1–17.
- Bian, Y., & Xie, X. Q. S. (2018). Computational fragment-based drug design: Current trends, strategies, and applications. *The AAPS Journal*, 20, 59.
- Böhm, H. J. (1994). The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure. *Journal of Computer-Aided Molecular Design*, 8, 243–256.
- Brandsdal, B. O., Osterberg, F., Almlöf, M., Feierberg, I., Luzhkov, V. B., & Aqvist, J. (2003). Free energy calculations and ligand binding. *Advances in Protein Chemistry*, 66, 123–158.
- Carpenter, K. A., & Huang, X. (2018). Machine learning-based virtual screening and its applications to Alzheimer's drug discovery: A review. *Current Pharmaceutical Design*, 24(28), 3347–3358.

- Chen, H. M., Liu, B. F., Huang, H. L., Hwang, S. F., & Ho, S. Y. (2007). SODOCK: Swarm optimization for highly flexible protein-ligand docking. *Journal of Computational Chemistry*, 28, 612–623.
- Chen, R., Li, L., & Weng, Z. (2003). ZDOCK: An initial-stage protein-docking algorithm. *Proteins: Structure, Function, and Bioinformatics*, 52(1), 80–87.
- Chong, S. H., & Ham, S. (2019). Folding free energy landscape of ordered and intrinsically disordered proteins. *Scientific Reports*, 9, 1–9.
- Clark, K. P. (1995). Flexible ligand docking without parameter adjustment across four ligand–receptor complexes. *Journal of Computational Chemistry*, 16, 1210–1226.
- DeWitte, R. S., & Shakhnovich, E. I. (1996). SMOG: De novo design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence. *Journal of the American Chemical Society*, 118, 11733–11744.
- Dey, F., & Caflisch, A. (2008). Fragment-based de novo ligand design by multiobjective evolutionary optimization. *Journal of Chemical Information and Modeling*, 48(3), 679–690.
- Di Nola, A., Roccatano, D., & Berendsen, H. J. (1994). Molecular dynamics simulation of the docking of substrates to proteins. *Proteins: Structure, Function, and Bioinformatics*, 19, 174–182.
- Dias, R., de Azevedo, J., & Walter, F. (2008). Molecular docking algorithms. *Current Drug Targets*, 9, 1040–1047.
- Dias, R., Timmers, L. F., Caceres, R. A., & de Azevedo, W. F. Jr (2008). Evaluation of molecular docking using polynomial empirical scoring functions. *Current Drug Targets*, 9, 1062–1070.
- Dror, R. O., Jensen, M. Ø., Borhani, D. W., & Shaw, D. E. (2010). Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. *Journal of General Physiology*, 135, 555–562.
- Durrant, J. D., Lindert, S., & McCammon, J. A. (2013). AutoGrow 3.0: An improved algorithm for chemically tractable, semi-automated protein inhibitor design. *Journal of Molecular Graphics and Modelling*, 44, 104–112.
- Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biology*, 9, 1–9.
- Eisen, M. B., Wiley, D. C., Karplus, M., & Hubbard, R. E. (1994). HOOK: A program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins: Structure, Function, and Bioinformatics*, 19(3), 199–221.
- Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., & Mee, R. P. (1997). Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of Computer-Aided Molecular Design*, 11, 425–445.
- Essex, J. W., Severance, D. L., Tirado-Rives, J., & Jorgensen, W. L. (1997). Monte Carlo simulations for proteins: Binding affinities for trypsin – benzamidine complexes via free-energy perturbations. *The Journal of Physical Chemistry B*, 101, 9663–9669.
- Ewing, T. J., Makino, S., Skillman, A. G., & Kuntz, I. D. (2001). DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *Journal of Computer-Aided Molecular Design*, 15, 411–428.
- Ferreira, L. G., Dos Santos, R. N., Oliva, G., & Andricopulo, A. D. (2015). Molecular docking and structure-based drug design strategies. *Molecules (Basel, Switzerland)*, 20, 13384–13421.
- Fogel, G. B., Cheung, M., Pittman, E., & Hecht, D. (2008). Modeling the inhibition of quadruple mutant Plasmodium falciparum dihydrofolate reductase by pyrimethamine derivatives. *Journal of Computer Aided Molecular Design*, 22, 29–38.
- Fraccalvieri, D., Pandini, A., Stella, F., & Bonati, L. (2011). Conformational and functional analysis of molecular dynamics trajectories by Self-Organising Maps. *BMC Bioinformatics*, 12, 1–19.
- García-Godoy, M. J., López-Camacho, E., García-Nieto, J., Nebro, A. J., & Aldana-Montes, J. F. (2015). Solving molecular docking problems with multi-objective metaheuristics. *Molecules (Basel, Switzerland)*, 20, 10154–10183.
- Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion in Drug Discovery*, 10, 449–461.
- Gohlke, H., Hendlich, M., & Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions. *Journal of Molecular Biology*, 295, 337–356.
- Goodsell, D. S., & Olson, A. J. (1990). Automated docking of substrates to proteins by simulated annealing. *Proteins*, 8, 195–202.
- Goodsell, D. S., Sanner, M. F., Olson, A. J., & Forli, S. (2021). The AutoDock suite at 30. *Protein Science*, 30, 31–43.
- Goto, J., Kataoka, R., & Hirayama, N. (2004). Ph4Dock: Pharmacophore-based protein-ligand docking. *Journal of Medicinal Chemistry*, 47, 6804–6811.
- Grosdidier, A., Zoete, V., & Michielin, O. (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research*, 39(2), W270–W277.
- Guedes, I. A., Pereira, F. S., & Dardenne, L. E. (2018). Empirical scoring functions for structure-based virtual screening: Applications, critical aspects, and challenges. *Frontiers in Pharmacology*, 9, 1089.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications.
- Hollingsworth, S. A., & Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron*, 99, 1129–1143.
- Homeyer, N., & Gohlke, H. (2012). Free energy calculations by the molecular mechanics Poisson–Boltzmann surface area method. *Molecular Informatics (Silver Spring, Md.)*, 31, 114–122.
- Huang, S. Y., & Zou, X. (2010). Advances and challenges in protein-ligand docking. *International Journal of Molecular Sciences*, 8, 3016–3034.
- Huang, Z., & Wong, C. F. (2009). Docking flexible peptide to flexible protein by molecular dynamics using two implicit-solvent models: An evaluation in protein kinase and phosphatase systems. *The Journal of Physical Chemistry B*, 113, 14343–14354.
- Ingólfsson, H. I., Amarez, C., Periole, X., & Marrink, S. J. (2016). Computational ‘microscopy’ of cellular membranes. *Journal of Cell Science*, 129, 257–268.
- Istiyastono, E. P., Kooistra, A. J., Vischer, H. F., Kuijper, M., Roumen, L., Nijmeijer, S., et al. (2015). Structure-based virtual screening for fragment-like ligands of the G protein-coupled histamine H<sub>4</sub> receptor. *MedChemComm*, 6, 1003–1017.

- Ivankov, D. N., Bogatyreva, N. S., Lobanov, M. Y., & Galzitskaya, O. V. (2009). Coupling between properties of the protein shape and the rate of protein folding. *PLoS One*, *4*, e6476.
- Jacquemard, C., Drwal, M. N., Desaphy, J., & Kellenberger, E. (2019). Binding mode information improves fragment docking. *Journal of Cheminformatics*, *11*, 1–15.
- Jain, A. N. (2003). Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *Journal of Medicinal Chemistry*, *46*, 499–511.
- Janson, S., Merkle, D., & Middendorf, M. (2008). Molecular docking with multi-objective particle swarm optimization. *Applied Soft Computing*, *8*, 666–675.
- Jones, G., Willett, P., & Glen, R. C. (1995). Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of Molecular Biology*, *245*, 43–53.
- Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, *267*, 727–748.
- Katoch, S., Chauhan, S. S., & Kumar, V. (2020). A review on genetic algorithm: Past, present, and future. *Multimedia Tools and Applications*, 1–36.
- Kellenberger, E., Rodrigo, J., Muller, P., & Rognan, D. (2004). Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins: Structure, Function, and Bioinformatics*, *57*, 225–242.
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks* (Vol. 4, pp. 1942–1948). IEEE.
- Kufareva, I., Rueda, M., Katritch, V., Dock, G. P. C. R., Stevens, R. C., & Abagyan, R. (2011). Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure (London, England: 1993)*, *19*, 1108–1126.
- Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., et al. (2011). ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods in Enzymology*, *487*, 545–574.
- Lexa, K. W., & Carlson, H. A. (2012). Protein flexibility in docking and surface mapping. *Quarterly Reviews of Biophysics*, *45*, 301.
- Li, H., Li, C., Gui, C., Luo, X., Chen, K., Shen, J., et al. (2004). GAsDock: A new approach for rapid flexible docking based on an improved multi-population genetic algorithm. *Bioorganic & Medicinal Chemistry Letters*, *14*, 4671–4676.
- Li, H., Sze, K. H., Lu, G., & Ballester, P. J. (2020). Machine-learning scoring functions for structure-based drug lead optimization. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *10*, e1465.
- Li, Z., Gu, J., Zhuang, H., Kang, L., Zhao, X., & Guo, Q. (2015). Adaptive molecular docking method based on information entropy genetic algorithm. *Applied Soft Computing*, *26*, 299–302.
- Lionta, E., Spyrou, G., KVassilatis, D., & Cournia, Z. (2014). Structure-based virtual screening for drug discovery: Principles, applications and recent advances. *Current Topics in Medicinal Chemistry*, *14*, 1923–1938.
- Liu, M., & Wang, S. (1999). MCDOCK: A Monte Carlo simulation approach to the molecular docking problem. *Journal of Computer-Aided Molecular Design*, *13*(5), 435–451.
- Macindoe, G., et al. (2010). HexServer: An FFT-based protein docking server powered by graphics processors. *Nucleic Acids Research*, *38*(2), W445–W449.
- Meiler, J., & Baker, D. (2006). ROSETTALIGAND: Protein–small molecule docking with full side-chain flexibility. *Proteins: Structure, Function, and Bioinformatics*, *65*(3), 538–548.
- Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking: A powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design*, *7*(2), 146–157.
- Mihai, D. (2017). *RoboDK: An offline programming and 3D simulation software for industrial robots*. Smashing Robotics.
- Morris, G. M., & Lim-Wilby, M. (2008). *Molecular docking. Molecular modeling of proteins* (pp. 365–382). Humana Press.
- Muegge, I. (2006). PMF scoring revisited. *Journal of Medicinal Chemistry*, *49*, 5895–5902.
- Nguyen, D. D., & Wei, G. W. (2019). AGL-Score: Algebraic graph learning score for protein–ligand binding scoring, ranking, docking, and screening. *Journal of Chemical Information and Modeling*, *59*(7), 3291–3304.
- Nittinger, E., Schneider, N., Lange, G., & Rarey, M. (2015). Evidence of water molecule—A statistical evaluation of water molecules based on electron density. *Journal of Chemical Information and Modeling*, *55*, 771–783.
- Pagadala, N. S., Syed, K., & Tuszynski, J. (2017). Software for molecular docking: A review. *Biophysical Reviews*, *9*(2), 91–102.
- Pantsar, T., & Poso, A. (2018). Binding affinity via docking: Fact and fiction. *Molecules (Basel, Switzerland)*, *23*(8), 1899.
- Patel, L., Shukla, T., Huang, X., Ussery, D. W., & Wang, S. (2020). Machine learning methods in drug discovery. *Molecules*, *25*(22), 5277.
- Pei, J., Wang, Q., Liu, Z., Li, Q., Yang, K., & Lai, L. (2006). PSI-DOCK: Towards highly efficient and accurate flexible ligand docking. *Proteins*, *62*, 934–946.
- Pierce, A. C., Sandretto, K. L., & Bemis, G. W. (2002). Kinase inhibitors and the case for CHO hydrogen bonds in protein-ligand binding. *Proteins*, *49*, 567–576.
- Pierce, A. C., Rao, G., & Bemis, G. W. (2004). Breed: Generating novel inhibitors through hybridization of known ligands. application to cdk2, p38, and hiv protease. *J. Med. Chem*, *47*, 2768–2775.
- Rai, S. K., Pathak, R. K., Singh, D. B., et al. (2021). Chemo-informatics guided study of natural inhibitors targeting rho GTPase: A lead for treatment of glaucoma. *In Silico Pharmacology*, *9*, 4. Available from <https://doi.org/10.1007/s40203-020-00061-y>.
- Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, *261*, 470–489.

- Raschka, S., & Kaufman, B. (2020). Machine learning and AI-based approaches for bioactive ligand discovery and GPCR-ligand recognition. *Methods*.
- Repasky, M. P., Shelley, M., & Friesner, R. A. (2007). *Flexible ligand docking with glide*. *Current protocols in bioinformatics*, Chapter 8, Unit 8.12.
- Rykunov, D., & Fiser, A. (2010). New statistical potential for quality assessment of protein models and a survey of energy functions. *BMC Bioinformatics*, 11(1), 1–11.
- Schneidman-Duhovny, D., et al. (2005). PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research*, 33(2), W363–W367.
- Singh, D. B., & Dwivedi, S. (2016). Docking and molecular dynamics simulation study of inhibitor 2-Fluoroaristeromycin with anti-malarial drug target PfSAHH. *Network Modeling and Analysis in Health Informatics and Bioinformatics*, 5, 16.
- Singh, D. B., & Dwivedi, S. (2019). Computational screening and ADMET-based study for targeting Plasmodium S-adenosyl-L-homocysteine hydrolase: Top scoring inhibitors. *Network Modeling and Analysis in Health Informatics and Bioinformatics*, 8, 4.
- Singh, D. B., Gupta, M. K., Kesharwani, R. K., & Misra, K. (2013). Comparative docking and ADMET study of some curcumin derivatives and herbal congeners targeting  $\beta$ -amyloid. *Network Modeling and Analysis in Health Informatics and Bioinformatics*, 2(1), 13–27.
- Singh, D. B. (Ed.) (2020). *Computer-aided drug design*. Singapore. [https://doi.org/10.1007/978-981-15-6815-2\\_7](https://doi.org/10.1007/978-981-15-6815-2_7).
- Singh, D. B., & Pathak, R. K. (2020). *Computational approaches in drug designing and their applications*. *Experimental protocols in biotechnology* (pp. 95–117). New York, NY: Humana.
- Singh, S., Chaturvedi, N., & Rai, G. (2020). De novo modeling and structural characterization of IL9-IL9 receptor complex: A potential drug target for hematopoietic stem cell therapy. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 9(1), 1–10.
- Srinivas Reddy, A., Chen, L., & Zhang, S. (2013). Structure-based de novo drug design, in. In G. Schneider (Ed.), *De novo molecular design* (pp. 97–124). Weinheim: Wiley-VCH Verlag GmbH and Co. KGaA.
- Stroganov, O. V., et al. (2008). Lead finder: An approach to improve accuracy of protein – ligand docking, binding energy estimation, and virtual screening. *Journal of Chemical Information and Modeling*, 48(12), 2371–2385.
- Taylor, J. S., & Burnett, R. M. (2000). Darwin: A program for docking flexible molecules. *Proteins: Structure, Function, and Genetics* (41), 173V191.
- Thomsen, R., & Christensen, M. H. (2006). MolDock: A new technique for high-accuracy molecular docking. *Journal of Medicinal Chemistry*, 49(11), 3315–3321.
- Trosset, J. Y., & Scheraga, H. A. (1999). PRODOCK: Software package for protein modeling and docking. *Journal of Computational Chemistry*, 20(4), 412–427.
- Trott, O., & Olson, A. J. (2010). AutoDockVina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2), 455–461.
- Tüzün, B., Yavuz, S. Ç., & Sarpınar, E. (2018). 4D-QSAR analysis and pharmacophore modeling: Propoxymethylphenyloxasiazole derivatives by electron conformational-genetic algorithm method. *Journal of Physical & Theoretical Chemistry*, 14(2), 149–164.
- Vakser, I. A. (2014). Protein-protein docking: From interaction to interactome. *Biophysical Journal*, 107, 1785–1793.
- Vanommeslaeghe, K., & Guvench, O. (2014). Molecular mechanics. *Current Pharmaceutical Design*, 20(20), 3281–3292.
- Veleg, H. F. G., Gohlke, H., & Klebe, G. (2005). DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *Journal of Medicinal Chemistry*, 48, 6296–6303. Available from <https://doi.org/10.1021/jm050436v>.
- Verdonk, M. L., Berdini, V., Hartshorn, M. J., Mooij, W. T. M., Murray, C. W., Taylor, R. D., et al. (2004). Virtual screening using protein-ligand docking: Avoiding artificial enrichment. *Journal of Chemical Information and Modeling*, 44, 793–806.
- Wang, H., Zhou, Z., Li, Y., Chen, Z., Lu, P., Wang, W., et al. (2017). Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18F-FDG PET/CT images. *EJNMMI Research*, 7, 11.
- Wang, K., Lyu, N., Diao, H., Jin, S., Zeng, T., Zhou, Y., et al. (2020). GM-DockZn: A geometry matching-based docking algorithm for zinc proteins. *Bioinformatics (Oxford, England)*, 36(13), 4004–4011.
- Wang, R., Lai, L., & Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of Computer-Aided Molecular Design*, 16, 11–26.
- Welch, W., Ruppert, J., & Jain, A. N. (1996). Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites. *Chemistry & Biology*, 3(6), 449–462.
- Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and Computing*, 4(2), 65–85, Harvard.
- Wójcikowski, M., Ballester, P. J., & Siedlecki, P. (2017). Performance of machine-learning scoring functions in structure-based virtual screening. *Scientific Reports*, 7(1), 1–10.
- Xu, X., Huang, M., & Zou, X. (2018). Docking-based inverse virtual screening: Methods, applications, and challenges. *Biophysics Reports*, 4(1), 1–16.
- Zhao, Y., & Sanner, M. F. (2007). FLIPDock: Docking flexible ligands into flexible receptors. *Proteins: Structure, Function, and Bioinformatics*, 68(3), 726–737.
- Zheng, L., Fan, J., & Mu, Y. (2019). Onionnet: A multiple-layer intermolecular-contact-based convolutional neural network for protein–ligand binding affinity prediction. *ACS Omega*, 4(14), 15956–15965.
- Zsoldos, Z., Reid, D., Simon, A., Sadjad, S. B., & Johnson, A. P. (2007). eHits: A new fast, exhaustive flexible ligand docking system. *Journal of Molecular Graphics and Modelling*, 26, 198–212.