Chapter

Development of Nucleic Acid Targeting Molecules: Molecular Docking Approaches and Recent Advances

Mohit Umare, Fai A. Alkathiri and Rupesh Chikhale

Abstract

Molecular docking is a widely used and effective structure-based computational strategy for predicting dynamics between ligands and receptors. Until now the docking software were developed for the protein-ligand interactions and very few docking tools were developed exclusively for the docking of small molecules on the nucleic acid structures like the DNA and RNA. The progress in algorithms and the need for deeper understanding of ligand-nucleic acid interactions more focused, and specialized tools are being developed to explore this hindered area of drug discovery. This chapter is focused on and discus in details about various tools available for docking with nucleic acids and how the rejuvenation of machine learning methods is making its impact on the development of these docking programs.

Keywords: nucleic acids, molecular docking, docking algorithms, machine learning, non-canonical DNA, RNA

1. Introduction

Computer-Aided Drug Design (CADD) has evolved as a cost-effective method of producing potential medications for the treatment of a wide range of diseases [1]. The use of the CADD technique in pharmaceutical research is becoming more common. Recently, there has been a trend in drug design to strategically create effective therapies with multi-targeting effects, better effectiveness, and tolerability, particularly in terms of toxic effects [2, 3]. To assist the exploration, a mix of modern computer approaches, biological research, and synthesizing molecules was developed, and this combinational methodology increased the scope of discoveries [4, 5].

CADD may be generally defined as encompassing both structure- and ligand-based drug design (SBDD and LBDD) [6]. SBDD approaches are based on evidence acquired

from an understanding of a target's three-dimensional structure, and they allow rating databases of compounds based on the affinity of ligands to a specific target [7, 8]. LBDD provides a generic technique for understanding links between the structural and compositional features of molecules and their bioactivities. When three-dimensional data for a protein of interest is lacking, this strategy is used [9]. The existing knowledge on molecules and their bioactivity are employed in this approach to produce new possible therapeutic molecules. In this regard, molecular docking is a widely used and effective structure-based computational based strategies for predicting dynamics between ligands and physiological receptors [10, 11].

The molecular docking procedure consists of two main stages: projection of a new molecular configuration including its pose inside the peptide-binding pocket, and evaluation of the pose quality using a scoring function. [11, 12]. Around 1975, high-throughput protein isolation, [13] nuclear magnetic resonance spectroscopy, and X-ray crystallography [14] have advanced, primarily leading to improved knowledge of the structural properties of ligand and molecule complex [15].

MD studies, along with many other *in silico* technologies, have grown more frequent and simpler to use in drug development; yet it is not wholly reliant on molecular libraries. Since its inception in the 1980s as among the most mostly utilized procedures, the experimental data collected by MD techniques has developed at an accelerating rate [16]. Nearly annually, programs configured using various methods for MD analysis are produced, considerably boosting pharmaceutical research. The scoring function calculates the binding affinities of produced poses, ranks them, and selects the most advantageous ligand and protein binding modes [17].

The scoring function of an optimum search algorithm should be capable of assessing the physical and chemical characteristics of compounds and the thermodynamics of interactions [18]. The earliest algorithms were created to deal with protein interactions [19]. Over the previous few decades, the progressive development of efficient and comprehensive algorithms with the inclusion of new variables has mirrored computing technical breakthroughs. Kuntz and colleagues at UCSF then utilized a shape pairing method algorithm to keep looking for alternative combinations based on the geometric length between the target and the ligand molecule [20].

The molecular docking technique has risen to prominence in the realm of drug development. Times over the past twenty years, molecular docking has developed as a vital tool for computational drug development, and it has been proved to be more systematic than conventional drug development approaches [16]. The enormous increase in computational capabilities and the rising access of molecule and protein libraries have considerably aided molecular docking. Several docking methodologies have been implemented over the last several years that may be used to dock proteins on peptides with diverse levels of accuracy. Molecular docking was initially intended to be done between a ligand and a target protein, but there is a significant focus on docking between proteins, and nucleic acid-protein-ligand docking, nucleic acid-ligand docking in the recent decade [21].

Methods for addressing the shortcomings of the docking approach are still being researched [22]. Results can be refined, for example, by employing consensus procedures, implementing more stringent scoring techniques to a portion of the filtered library, or employing filters that include interaction fingerprints [23]. Significant effort has also been undertaken to collect inputs from potential binding waters. Identified water molecules as critical for molecule recognition can be considered part of the binding pocket, and prediction can be enhanced by energy contribution by displacing water molecules [24].

2. Methods in molecular docking

2.1 Monte Carlo

In molecular docking studies, the Monte Carlo technique is the use in creation of a randomized conformation of a molecule in a targets active site. The advantage is that this method uses equilibrium statistical method. Rather than attempting to mimic a system's dynamics, it develops states based on the suitable Boltzmann distribution [25]. It determines the initial configuration value. Further, it generates and evaluates a new configuration. Through using Metropolis criteria, it assesses whether the new configuration should be preserved [26]. The Metropolis criteria states that if a new strategy provides better conformation than the previous one, it is recognized immediately. If the combination is not innovative, a probability assessment based on Boltzmann's law is used. If the conclusion passes the likelihood function test, it is approved, and the other arrangement is discarded [27].

2.2 Ligand fit

Ligand fit denotes to a rapid and accurate approach for docking small molecules into targets active sites while considering form as a complementarity. The technique of cavity identification is used in the procedure to discover and produce cavity in the protein as probable binding site locations [28]. For producing ligand poses that are compatible with the receptor binding site shape, a shape similarity screening is paired along with a Monte Carlo parametric analysis. A grid-based technique for analyzing energies between protein and ligand is used to reduce candidate poses with respect to the active site. A non-linear interpolation approach drastically reduces errors caused by grid interpolation [27, 28].

2.3 Point complimentary

Here on grounds of the complementarity of the interatomic contacts, a technique for docking a drug into a binding pocket in an enzyme is disclosed. Docking is accomplished by increasing a complementarity function that is reliant on the atomic surface area of contact as well as the elemental composition of the interacting atoms [29]. Although the target and ligand molecules are viewed as inflexible entities, mobility of a restricted range of residues bordering the binding site can also be considered. These techniques of molecular docking are focused on comparing the shapes and/or chemical properties of different molecules [26].

2.4 Fragment based

Fragment-based drug discovery (FBDD) is a novel strategy that is increasingly being used to improve hit recognition for previously thought intractable biological targets. FBDD, in specifically, uncovers small ligands (300 Da) capable of binding to pharmacologically important macromolecules with micromolar affinity [30].

2.5 Distance geometry

Even though it is primarily known as a tool for predicting the solution conformation of compounds from NMR data, distance geometry is a basic and effective tool for

generating approximation models of complicated chemical formations [31]. Distance geometry is a basic geometrical approach that builds structures directly to fulfill model requirements; this does not involve an initial conformational or force field variables. The approach simply handles flexible rings without any extra attention or adjustment. Distance geometry is also distinct in that it works well together with qualitative data: a significant number of estimated distance boundaries are more useful in creating a model than a limited handful of highly exact distances [12, 31].

3. Nucleic acid docking

Nucleic acids (NAs) are biological macromolecules which can be broken down into phosphoric acid, sugars, and mixture of organic bases like purines and pyrimidines [32]. These can occur in various forms and constitute the building blocks like the DNA and RNA. These are essential for various cellular process including cell division and protein synthesis [33, 34]. Due to their crucial role in cell division, DNA, RNA, and their alternate structures have become target of choice for drug discovery in case of cancer drug discovery, infectious diseases, and rare diseases [35–38]. The NA modulators act by interfering with DNA replication process which affect the cell proliferation, transcription and ultimately inhibition of gene expression [39]. These agents can modulate the functioning of the RNA resulting in altered transcription and translation processes [40]. These modulators could be small molecule ligands, peptide or macromolecules, these can interact with the NAs by various mechanisms like intercalation, molecular cross-linking, DNA or RNA strand cleavage, and interference at the site of NA-protein interactions (**Figure 1**) [40, 41].

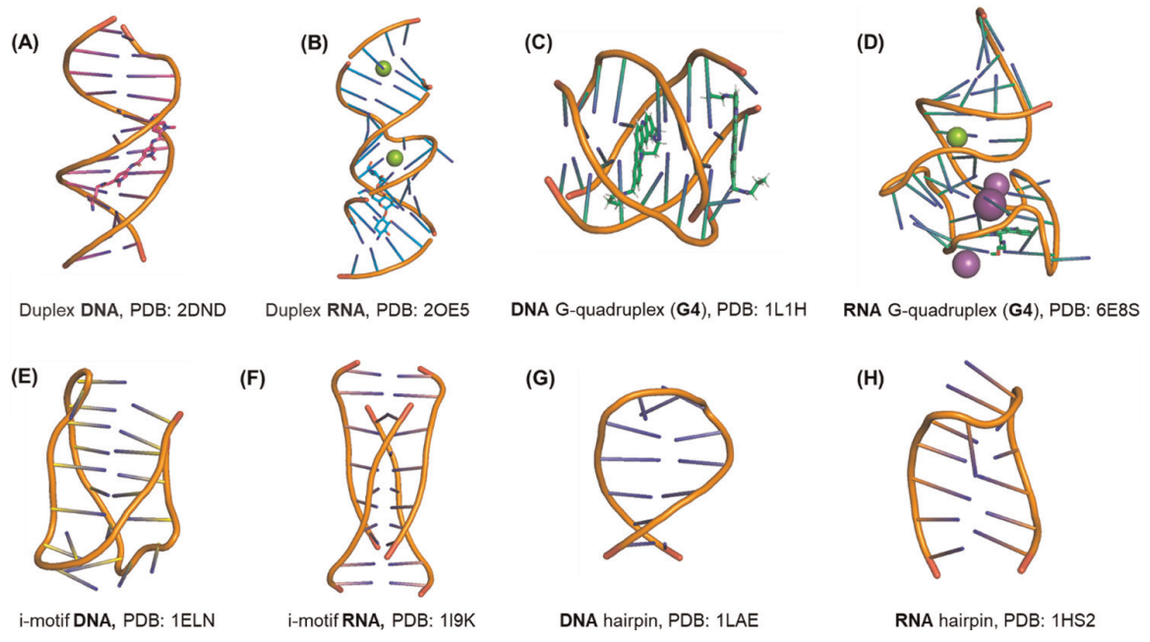

Figure 1.
The commonly known NA structures with and without bound ligands; (A) duplex DNA structure with a bound antitumour drug, distamycin, PDB: 2DND [42]; (B) duplex RNA structure with a bound aminoglycoside antibiotic, apramycin, PDB: 2OE5 [43]; (C) DNA G-quadruplex in complex with the di-substituted amino alkylamido acridine compound (G4), PDB: 1L1H [44]; (D) RNA G-quadruplex (G4) crystal structures of TO1-biotin complexes of mango-III, a structure-guided mutant mango-III (A10U), PDB: 6E8S [45]; (E) i-motif DNA, a fragment of the vertebrate telomere which folds intramolecularly, PDB: 1ELN [46]; (F) i-motif RNA, a oligodeoxynucleotides with stretches of cytidine residues associate into a four-stranded structure, PDB: 1I9K [47]; (G) DNA hairpin, solution structure of the PdG-containing hairpin PDB: 1LAE [48]; (H) RNA hairpin, solution structure of RNA hairpin loop, PDB: 1HS2 [49].

Recent advancement in crystallization techniques, oligonucleotide synthesis, methods for structure determination like the NMR, crystal diffraction and cryo-EM has allowed for enrichment of structural data for NAs [50, 51]. The protein data bank (PDB) is an open source repository where these structures are deposited and curated [52]. There are more than 730 DNA-ligand and 523 RNA-ligand co-crystallized structures in the PDB and these would keep increasing [53]. Structural data of NAs helps in the investigation of the possible binding of ligands into the target, a co-crystallized structure provides with a bound ligand which helps understand the binding or active site in the given NAs. These co-crystallized molecules offer an excellent opportunity to perform structure-based and ligand-based drug discovery experiments and apply various other computational methods for drug discovery of NAs therapeutics. The most widely used method in computational drug design is molecular docking studies. The algorithms available for performing molecular docking are basically made for ligand-protein docking. There are several similarities like the protein and NAs follow similar physicochemical binding principles. However, these algorithms often fail to lack of sufficient sampling of the conformation space in case of NA docking to reasons of non-specific scoring functions [54]. Most of the target protein molecules contain a hydrophobic binding site whereas, the NAs consist of a rather more solvent-exposed binding pocket with higher polarity and charge density [55]. These are the major differences between the proteins and NAs as targets in molecular docking. Most of these algorithms are focused on the protein target molecules and need to consider parameters that need to be included in the program for NAs docking. NAs particularly the RNAs are very flexible owing to their charge, intrinsic atomic arrangements, and movements due to the presence of ligands. This flexibility is not considered by most of the programs as they consider NAs as rigid bodies [56]. Some programs like MORDOR are available that allows for the flexibility of the NAs and the ligands [57]. It applies molecular mechanics minimisation restraints based on the data from the X-ray and NMR experimental data [58]. There are several shortfalls to these methods, they are marred by slow speed, minimisation stages are slow, and time consuming, and large library screening is not feasible. Other NA specific methods reported were ensemble docking based on structural information from the X-ray structures or NMR or structures from the normal-mode analysis of an MD simulation [59–61]. The presence of water molecules and metal ions add to the complications in NAs docking. The water molecules and metal ions are essential for the stability and functioning of the NAs, this makes their presence in any docking protocol imperative. The metal ions in case of NAs like the i-Motif and G-quadruplex are necessary for the formation and stability of the structure [62, 63]. Various algorithms that considers these challenges in NAs docking are discussed in the section scoring function.

4. Recent developments in docking tools for nucleic acid

There are several types of small molecules that interact with the NAs and its alternate forms. These can be subdivided into double stranded DNA/RNA (ds-DNA and ds-RNA) binding, G-quadruplex DNA/RNA (G4-DNA and G4-RNA) binding, i-Motif DNA/RNA (iM-DNA and iM-RNA) binding ligands and ligands interacting with other DNA structures like hairpins [62, 63]. These ligands can also be classified based on their mechanism of binding to the DNA, for example covalent binding and intercalators. Several review articles have discussed these ligands in more details in the past [64]. The lab-based experiments and further crystallization experiments are

costly and time consuming and hence to assist with these efforts molecular modeling and docking tools are used widely to find the most suitable ligand. Most of the available molecular docking tools have been developed for protein-ligand docking. These tools have been used for NA-ligand docking irrespective of the fact that these tools do not consider the NAs as flexible moieties and thus do not consider the most important feature of NAs. The other type of docking interaction that NA undergo is with the proteins, Protein-NAs docking [65]. There are several algorithms that are used to perform NA-protein docking as mentioned in the table number 1. Earlier reports in NA-ligand docking dealt with finding correct docking conformations based on RMSD to the native co-crystallized ligand. Autodock and Surflex were used to dock several ligands like pentamidine, daunorubicin, distamycin and ellipticine in the minor groove of the ds-DNA. It was observed that Surflex performed better over Autodock in speed of operation and results with lower reference RMSD [66]. Several algorithms have been published and are available for NAs-ligand docking like, GRAMM, FTDock, 3D-DOCK, HEX, Dot and DoT2, HADDOCK, PatchDock, SymmDock, ParaDock, GOLD, Glide [67], NPDcok and HDOCK (**Table 1**). The most recent NA-ligand docking tools are NLDock, LigandRNA and DOCK 6.

The DOCK algorithm developed by the Kuntz lab has been traditionally a protein-ligand docking program. However, the most recent development of the series is

Algorithms	Acronym	Principle	Reference
Geometric Recognition Algorithm to identify Molecular surface complementarity.	GRAMM	Rigid docking uses fast Fourier transformation, shape-based complementarity.	[68]
Fourier Transform rigid-body Docking	FTDock	Use and implementation of the biochemical and electrostatic information of the DNA and host protein or DNA.	[69]
Initial grid-based shape complementarity search	3D-Dock	Featured backbone refinement, side chain optimization and energy calculations.	[70]
Spherical polar Fourier correlations	HEX	Docking pairs of proteins by using spherical polar Fourier correlations to accelerate the search for candidate low-energy conformations.	[71]
Rapid computation of the electrostatic potential energy between two proteins or other charged molecules.	Dot and Dot2 (Daughter of Turnip)	Automated construction of improved biophysical models based on molecular coordinates, provides for flexibility with grid size and allows improved rescoring method. Uses Poisson-Boltzmann methods.	[72]
High Ambiguity Driven protein-protein Docking	HADDOCK	Uses Ambiguous Interaction Restraints (AIRs), takes up information form the biophysical, biochemical interactions found in the NMR or crystal structure.	[73]
Geometry-based molecular docking algorithm	PatchDock	Aims at finding good molecular shape complementarity.	[74]
Geometry-based docking algorithm for the prediction of a cyclically symmetric complex	SymmDock	It aims to find symmetric cyclic transformations.	[75]

Algorithms	Acronym	Principle	Reference
ab initio protein–DNA docking algorithm	ParaDock	Geometric complementarity-based docking.	[76]
Protein-Nucleic acid docking	NPDock	It predicts the protein–nucleic acid structures interactions by clustering the best-scored models and ranking the refined solutions.	[77]
Hybrid docking	HDOCK	Template based modeling and free docking.	[78]
Genetic Optimisation for Ligand Docking	Gold	Explores full range of ligand conformational flexibility, loosely bound water molecules in the binding site or the active site.	[79]
RNA − ligand interactions	DrugScore^{RNA}	Uses experimental structures as reference and applies distance-dependent pair potentials with reference.	[80]
Molecular Recognition with a Driven dynamics Optimize R	MORDOR	Explores the electrostatic, van der Waals forces. Takes consideration of dihedral angle, torsion angle, and bond lengths. CHARMM or AMBER based scoring functions and uses implicit solvent models.	[61]
Binding mode predictions	AutoDock AutoDock Vina	Uses simulated annealing method for docking, flexible ligand and some extent of receptor flexibility.	[81–84]
Fully automated flexible docking	Surflex	Uses surface-based molecular similarity method to generate suitable poses for molecular fragments.	[85]
RiboDock	rDock	It uses stochastic and deterministic search techniques and generates low energy ligand poses.	[86]
Nucleic acid-Protein Docking	NPDock	Makes use of clustering of best score models.	[77]
Nucleic acid-Ligand Docking	NLDock	ITScore-NL scoring function used, it makes the use of stacking and electrostatic potentials.	[87, 88]
RNA-Ligand docking	LigandRNA	Makes use of grid-based algorithm and potentials derived from experimentally solved RNA-ligand complexes.	[89]
Iterative knowledge-based scoring function for nucleic acid–ligand interactions	ITScore-NL	Physics based iterative methods used. Makes use of atomic and distance dependent pair potentials. Uses stacking interactions and electrostatic effects.	[88]
Ranking-based sampling algorithm	DOCK 6	Dominant electrostatics and charges from waters were considered.	[90]

Table 1.
List of NA-ligand docking tools with their names and principle of working and algorithms.

DOCK6 which has the special feature to dock small molecules on the NAs. DOCK6 have significant progress in ligand orientation and conformational sampling which has led to significant improvement in the accuracy of docking for the large and flexible molecules over the NAs. It uses a sampling algorithm 'anchor-and-grow' which allows a cluster-based pruning with controlled cut-off of 25 kcal/mol. This flexibility in the upper limit allows for ranked orientation and improves prediction near the binding site. DOCK 6 uses the MD parameters like the AMBER GB/SA and PB/SA for predicting and ranking the poses and the effect of presence of metal ions and the water molecules in the binding site. The NLDock developed by the Huang lab uses ITScoreNL which is an iterative knowledge-based scoring function. The ITScoreNL uses a statistical mechanics based interactive algorithm. It uses the information from a training set of experimentally determined structures in the protein data bank (PDB). This scoring function consist of atomic, distance dependent pair potential, stacking interaction, and electrostatic effects. Results from ITScoreNL significantly improve the performance in binding and affinity prediction for the NAs-ligand complex. Recent advances and enrichment of the RNA structures in the PDB let to the development of LigandRNA. It uses the 3D information from the available RNA structures. A potential is obtained using the inverse Boltzmann scheme which considers the ligand poses that are favorable and exhibit interactions fitting the maxima of the statistical distribution of RNA-ligand atom contacts derived from the RNA-ligand co-crystal structures. This method is dedicated to scoring and ranking ligand poses in their RNA three-dimensional structure with correct intramolecular interactions while maintaining high accuracy and precision. These recent tools have given larger momentum to screening of ligands for NAs with better accuracy and speed.

5. Scoring functions

Molecular docking is quickly becoming a valuable technique in drug development and molecular modeling fields. The precision of the selected scoring function, that can lead and identify ligand positions when hundreds of potential ligand positions are created, determines the effectiveness of molecular docking [11, 91, 92]. The scoring function can also be used to forecast binding affinity and discover possible drug candidates for a specific protein of interest, as well as to define the binding mode and location of a molecule [93]. In lead optimization, scoring functions serve three main purposes: first, they recognize the best location of a ligand's binding to a protein based on the scoring function; second, they estimate the absolute binding affinity between the protein and ligand; and third, they perform virtual screening, which can identify possible drug leads for a given target protein by finding a sizable molecule database. [93].

The most recent scoring functions for protein-ligand interactions using a new categorization that divides the scoring functions into force-field-based, empirical, and knowledge-based SFs. Ongoing study has drastically enhanced the research for scoring functions, particularly in protein-ligand interactions.

5.1 Physics-based scoring functions

Direct computation of the associations between both the atoms of a protein and a ligand is possible using physics-based SFs. Owing to the consideration of solvation, enthalpy, and entropy, physics-based SFs are suited to calculate binding free energy

among proteins and ligands with significantly improved prediction performance than other forms of SFs [94]. These are founded on solvation models, force fields, and quantum mechanics techniques. The van der Waals and electrostatic interactions between the protein and ligand atom pairs are added up in the conventional force field-based SF, which considers the energy-contributing role of enthalpy, to estimate the binding energy [95].

Pairwise atomic interactions between the ligand and protein are the focus of the fundamental equation in the classical method. R is the distance between atomic centres, q is the fractional charge on every atom, and e is the dielectric constant. The A and B parameters are determined for every pair of various atom type combinations [96].

$$\Delta G_{bind} = \sum_{i=1}^{ligand} \sum_{j=1}^{protein} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}} \right]$$

5.2 Empirical scoring functions

Empirical SFs calculate a complex's binding energy by adding up the essential energy components for binding affinity, such as hydrophobic effects, hydrogen bonds, steric conflicts, and so on. There are two study paths in empirical SFs. One approach is to use a usually high labeled training data to optimize protein complexes; the other is to pick appropriate energy terms using progressive parameters and methodical selection of the target molecule [92, 97].

5.3 Knowledge-based scoring functions

Predicated on the reverse Boltzmann statistic concept, knowledge-based SFs compute the appropriate pairwise potential in terms of 3D structures of a wide range of complexes. The rate of distinct atom pairs at different distances is thought to be connected to the interactions between two atoms, which translates the rate through the distance-dependent potential of mean force [18]. When tried to compare to physics and empirical SFs, knowledge-based SFs have the largest benefit in terms of processing cost and prediction accuracy. Unfortunately, knowledge-based SFs have a tough time locating the reference state [98].

5.4 DrugScoreRNA

Interactions of protein with protein, DNA, and ligand have all been studied using knowledge-based techniques. DrugScoreRNA is the first knowledge-based technique to scoring RNA-ligand complexes. Because of the small percentage of experimental measurements of RNA-ligand combinations, it was thought that obtaining statistically meaningful potentials was improbable [80].

The fact that the binding (free) energy landscape derived by such prospects is more focused than in the context of all other knowledge-based SFs or AutoDock may be taken into consideration as one of the factors contributing to DrugScoreRNA's effectiveness in docking [18]. This is anticipated to result in a quicker docking converging to a global solution, or, put another way, a lower probability that the configurational search would get stale in a local minimum. Reasonable correlation exists between experimental binding free energies and binding scores estimated by DrugScoreRNA. [99]

5.5 RiboDock

The growing understanding of the significance of RNA in fundamental biological processes has lately made them more appealing as prospective therapeutic targets. To find small compounds that may selectively bind to identified locations in RNA molecules and inhibit or otherwise modify their function, a greater number of scientifically confirmed RNA three- dimensional structures were available. This allowed for structure-based searches for these molecules [100]. The access to high resolution structures of RNA-ligand complexes substantially facilitates the investigation of the atomic intricacies of RNA-ligand contacts. Furthermore, it is difficult to determine the physical structure of RNA and its interactions, and it is now unable to do so in a high-throughput way. This is what inspired the creation of source code for simulating the configurations of RNA-ligand complexes based on the known structures of RNA targets. Many of these advancements were motivated by comparable strategies used earlier for protein-ligand complex modeling [89, 100].

One of the first to develop a scoring function specifically for RNA-ligand complexes was done in 2004 by Morley and Afshar. They added the empirical regression-based tool RiboDock (or rDock) to their own high-throughput docking tool to handle RNA-ligand structures [101]. This technique was, unfortunately, parameterized and tested on a small sample size of just 10 RNA molecules. Ligand intramolecular, intermolecular, site intramolecular, and external constraint factors are weighted together to form the rDock master score function. The major terminology of importance is S^{intra}, which stands for the RNA-ligand interaction score. According on the provided ligand configuration, S^{intra} provides the ligand's energy transfer. Similar to S^{site}, this term denotes the comparative energy of the active site's variable regions [100, 101].

5.6 LigandRNA

As discussed in the above section, the importance of RNA in fundamental biological processes has grown the scientific community interest in the research area of Nucleic Acid-Ligand docking. Another Scoring function developed for the similar function was LigandRNA [89].

The RNA-ligand complexes were computationally solved using the LigandRNA approach, which uses a grid-based algorithm and a knowledge-based SFs obtained from ligand-binding domains. LigandRNA requires two files as inputs: an RNA receptor file and a ligand poses file. It produces a list of poses ranked by their score as an output [100]. The potential is calculated using the inverse Boltzmann method, which assumes that only ligand poses with interactions that meet the maximum of the statistical distribution of RNA-ligand atom contacts generated from empirically established structures of RNA-ligand complexes are advantageous. Thus, according to their value, the supplied ligand poses are sorted, and this score would be used to assess the relative effectiveness of binding [89].

5.7 MM/PBSA and MM/GBSA

The molecular mechanics energies combined with the Poisson–Boltzmann or generalized Born and surface area continuum solvation (MM/PBSA and MM/GBSA) are the popular techniques for estimating the free energy of the binding of ligand

molecules to the target protein. In MM/PBSA, the free energy of a state, that is, P, L or PL in the following equation, is estimated from the following sum [102].

$G = E_{bnd} + E_{el} + E_{vdW} + G_{pol} + G_{np} - TS.$

$E_{bnd:}$ Bonded (bond, angle and dihedral) energy.

E_{el}: Electrostatic Energy.

E_{vdW}: van der Waals interactions.

G_{pol}: polar contribution to the solvation free energy.

G_{np}: non-polar contribution to the solvation free energy.

To calculate the MM/GBSA free energy, the system of relevance is first modeled either using Metropolis Monte Carlo or molecular dynamics (MD), with pose is being obtained at set intervals and for each pose the free energy is calculated by the above equation. The continuum-solvation technique, the dielectric constant, the charges, the sample selection, and the entropies have a significant impact on the outcomes. The approaches frequently exaggerate the differences between different ligand groups [103]. In actual use, it frequently produces outcomes of middling quality, frequently outperforming docking, and scoring. However, because of the findings' substantial reliance on the continuum solvation used, either the absolute affinities or the methodology is invalid [103, 104].

5.8 Molecular recognition with a driven dynamics optimizer (MORDOR)

The fixed nature of the protein target is drawback in most of the docking tools. To overcome this and to explore the dynamic nature of the target Molecular Recognition with a Driven dynamics Optimizer (MORDOR) tool was developed. MORDOR allows induced-fit type of docking algorithm. A new RNA stabilizing loop can be formed by the ligand, which could move bases [105].

MORDOR uses a unique conformational field search technique to achieve this goal, enabling a productive thorough search while docking. Utilizing a driving force to move the ligand, this method combines molecular minimization technique. By applying an extra RMSD kind of force, the ligand explores the receptor surface after beginning from any pose in and around the receptor. It is crucial to research induced fit with MORDOR when docking proteins, especially RNA. Drugs do not often bind a conventional form of nucleic acid, according to the architectures of nucleic acid-drug complexes. Also, more control over the docking process is provided by the allowance of an infinite number of restraints. Contrarily, it seems from known drug-nucleic acid binding structures that the small molecule ligands frequently replace bases, leading to a local restructuring of the nucleic acid. A drug development process will have a far better chance of being successful if flexible docking for RNA is used [61, 105].

5.9 Dock-RNA

Numerous biological activities, including the production and control of gene activity, depend on nucleic acid-ligand interactions. As a result, nucleic acid molecules like RNAs have grown in importance as pharmacological targets and knowing the structural characteristics of RNA-ligand complexes is essential to deriving treatment strategies. The nucleic acid-ligand docking method is divided into two stages: The model chooses a preliminary set of potential poses during the first stage using a different computer algorithm for the Born radiuses in the electrical charges; with in second stage, a stringent scoring function is utilized to arrange the poses to identify the top molecules [106].

The scoring function of the molecular docking program is dependent on the shift in free energy caused by RNA-ligand binding. It aggregates comparable ligand poses into clusters based on geometrical similarity and ranks the grouped poses based on the binding affinity. Because it separates itself from other models by sampling all potential interaction site and poses globally, the findings above highlight the relevance poses. Unfortunately, the RLDOCK approach is difficult to apply to big target and ligand sets. The time-consuming selection of the complex formation produces prohibitively small processing effectiveness of the approach in complexes with a big RNA such as ribosomal RNA or ligands with the more than 12 rotatable bonds [107, 108].

6. Role of machine learning and artificial intelligence

Machine learning (ML) specially the Deep learning methods (DL) and Artificial intelligence (AI) has rapidly developed and is being used in drug discovery. ML in drug discovery is used to improve the existing scoring functions or to develop a new scoring function for virtual screening studies. The existing scoring functions can be improved by refining their empirical function's weights. Most of the ML based scoring function improvements has been seen in the protein-ligand docking and their virtual screening domain. The ML methods being used are Random Forest methods [109], Gradient boosting trees method [110], Support vector machine methods [111], Multi-layer perceptron methods [112], Convolutional neural network methods [113], and Graph neural network [114]. The scoring functions for NAs-ligand interactions can be classified into force-field based, empirical, knowledge-based and machine learning based. The machine learning based scoring functions can capture intrinsic nonlinearities in the training set without imposing a predetermined functional form. The most important feature that separates the ML methods from others is that ML maps the ligands to a potential energy landscape, it is inherently flexible, and the mapping relationship works without the addition of extensive physicochemical knowledge. However, the use of ML in NAs binding ligands discovery comes with certain challenges as well. First, the mapping relationships generated by ML are not always interpretable and the second, ML models for NAs could find difficult to make accurate predictions for complexes out of the training sets.

For the NA-ligand complex interactions two ML based scoring functions were recently developed, RNAPoser [115] and AnnapuRNA [116]. The RNAPoser uses a set of 80 RNA-ligand experimental structures as dataset and investigates the 'nativeness' of the RNA-ligands poses. This program uses machine learning methods to train a set of pose classifiers that would estimate the position of the ligands in the experimental structures. These poses are defined as fingerprints which are encoded as local RNA environment surrounding the ligand. This method uses the leave-one-out training and testing approach where about 80% of the native poses were recovered within 2.5 Å. The classification is done based on ranking of ligands and scoring from machine learning classifiers, which were able to recover the native like poses. The validation set for the method returned recovery of native poses for more than 60% of the cases. These were found to be better than the poses with higher docking scores. Another recent development in the NA-ligand docking improvement is AnnapuRNA. It is a machine learning-based statistical scoring function which can evaluate the quality of RNA-Ligand complex structure predicted by a computational docking program and thus help in validation of the docking results. It uses the information like the initial ligand conformation, the docking program and the scoring function used by the

docking program. The training set is derived from the experimental data available on the PDB and it uses the *k*NN (*k*-Nearest Neighbors) and Deep Learning (multi-layer feedforward artificial neural network) as ML algorithms. This program supports a various docking program like the AutoDock, AutoDock Vina, Dock6, rDock, iDock, LigandRNA, and several other NAs specific programs.

7. Conclusion

In this chapter we have overviewed various important aspects in development of small molecule inhibitors for NAs and various docking software specific and non-specific for NAs-ligand docking. We have also reviewed various docking programs, algorithms and scoring functions, their advantages and lacune and challenges in the discovery of novel NAs binding ligands. Until recently most of the algorithms were focused on protein-ligand docking but now slowly programs specific for NAs are appearing in the molecular docking space. The progress in ML and AI has led to an advantage for development of NA specific algorithms. However, there is lot of scope for development of NA-docking specific programs, structural variations of NA also pose a challenge for the new programs. However, it is possible to convert these challenges into opportunities as the need for better NA targeting ligands are high in demand specifically due to the resurgence of viral infections and other infectious disease.

Acknowledgements

FA acknowledges the support provided by Department of Pharmaceutics, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia.

Author details

Mohit Umare[1], Fai A. Alkathiri[2] and Rupesh Chikhale[3]*

1 Tata Consultancy Services Limited, Pune, India

2 Department of Pharmaceutics, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia

3 UCL School of Pharmacy, London, UK

*Address all correspondence to: rupeshchikhale7@gmail.com

References

[1] Tang Y, Zhu W, Chen K, Jiang H. New technologies in computer-aided drug design: Toward target identification and new chemical entity discovery. Drug Discovery Today: Technologies. 2006;**3**: 307-313

[2] Kiriiri GK, Njogu PM, Mwangi AN. Exploring different approaches to improve the success of drug discovery and development projects: A review. Future Journal of Pharmaceutical Sciences. 2020;**6**(1):27

[3] Yu W, MacKerell AD Jr. Computer-aided drug design methods. Methods in Molecular Biology. 2017;**1520**:85-106

[4] Nicolaou KC. Organic synthesis: The art and science of replicating the molecules of living nature and creating others like them in the laboratory. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2014;**470**(2163):20130690

[5] Atanasov AG, Zotchev SB, Dirsch VM, Orhan IE, Banach M, Rollinger JM, et al. Natural products in drug discovery: Advances and opportunities. Nature Reviews Drug Discovery. 2021;**20**(3):200-216

[6] Aparoy P, Reddy KK, Reddanna P. Structure, and ligand-based drug design strategies in the development of novel 5-LOX inhibitors. Current Medicinal Chemistry. 2012;**19**(22):3763-3778

[7] Meng XY, Zhang HX, Mezei M, Cui M. Molecular docking: A powerful approach for structure-based drug discovery. Current Computer-Aided Drug Design. 2011;**7**(2):146-157

[8] Yasuo N, Sekijima M. Improved method of structure-based virtual screening via interaction-energy-based learning. Journal of Chemical Information and Modeling. 2019;**59**(3): 1050-1061

[9] Bacilieri M, Moro S. Ligand-based drug design methodologies in drug discovery process: An overview. Current Drug Discovery Technologies. 2006;**3**: 155-165

[10] Sivakumar KC, Haixiao J, Naman CB, Sajeevan TP. Prospects of multitarget drug designing strategies by linking molecular docking and molecular dynamics to explore the protein–ligand recognition process. Drug Development Research. 2020;**81**(6):685-699

[11] Silakari O, Singh PK. Chapter 6 - Molecular docking analysis: Basic technique to predict drug-receptor interactions. In: Silakari O, Singh PKBTC and EP of M and I in DD, editors. Concepts and Experimental Protocols of Modelling and Informatics in Drug Design. Academic Press; 2021. pp. 131-155

[12] Guedes IA, de Magalhães CS, Dardenne LE. Receptor-ligand molecular docking. Biophysical Reviews. 2013/12/21. 2014;**6**(1):75-87

[13] Koehn J, Hunt I. High-throughput protein production (HTPP): A review of enabling technologies to expedite protein production. Methods in Molecular Biology. 2009;**498**:1-18

[14] Jhoti H, Cleasby A, Verdonk M, Williams G. Fragment-based screening using X-ray crystallography and NMR spectroscopy. Current Opinion in Chemical Biology. 2007;**11**(5):485-493

[15] Kim Y, Bigelow L, Borovilos M, Dementieva I, Duggan E,

Eschenfeldt W, et al. Chapter 3. High-throughput protein purification for x-ray crystallography and NMR. Advances in Protein Chemistry and Structural Biology. 2008;**75**:85-105

[16] Pinzi L, Rastelli G. Molecular docking: Shifting paradigms in drug discovery. International Journal of Molecular Sciences. 2019;**20**(18):4331

[17] Stanzione F, Giangreco I, Cole JC. Use of molecular docking computational tools in drug discovery. Progress in Medicinal Chemistry. 1st ed. Vol. 60. Elsevier B.V.; 2021. pp. 273-343

[18] Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL 3rd. Assessing scoring functions for protein-ligand interactions. Journal of Medicinal Chemistry. 2004;**47**(12):3032-3047

[19] Hoskins J, Lovell S, Blundell TL. An algorithm for predicting protein-protein interaction sites: Abnormally exposed amino acid residues and secondary structure elements. Protein Science. 2006;**15**(5):1017-1029

[20] DesJarlais RL, Sheridan RP, Seibel GL, Dixon JS, Kuntz ID, Venkataraghavan R. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. Journal of Medicinal Chemistry. 1988 Apr;**31**(4):722-729

[21] Torres PHM, Sodero ACR, Jofily P, Silva-Jr FP. Key topics in molecular docking for drug design. International Journal of Molecular Sciences. 2019; **20**(18):4574

[22] Yuriev E, Ramsland PA. Latest developments in molecular docking: 2010–2011 in review. Journal of Molecular Recognition. 2013;**26**(5): 215-239

[23] Lionta E, Spyrou G, Vassilatis DK, Cournia Z. Structure-based virtual screening for drug discovery: Principles, applications and recent advances. Current Topics in Medicinal Chemistry. 2014;**14**(16):1923-1938

[24] Maurer M, Oostenbrink C. Water in protein hydration and ligand recognition. Journal of Molecular Recognition. 2019;**32**(12):e2810 e2810

[25] Zivkovic M, Zlatanovic M, Zlatanovic N, Golubović M, Veselinović AM. The application of the combination of Monte Carlo optimization method based QSAR modeling and molecular docking in drug design and development. Mini Reviews in Medicinal Chemistry. 2020;**20**(14): 1389-1402

[26] Zivkovic M, Zlatanovic M, Zlatanovic N, Golubović M, Veselinović AM. The Application of the Combination of Monte Carlo Optimization Method based QSAR. Modeling and Molecular Docking in Drug Design and Development. Mini Rev Med Chem. 2020;**20**(14):1389-402

[27] Salmaso V, Moro S. Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. Frontiers in Pharmacology. 2018;**9**. Available online From: https://www.frontiersin.org/articles/10.3389/fphar.2018.00923. DOI=10.3389/fphar.2018.00923. ISSN=1663-9812

[28] Venkatachalam C, Jiang X, Oldfield T, Waldman M. LigandFit: A novel method for the shape-directed rapid docking of ligands to protein active sites. Journal of Molecular Graphics & Modelling. 2003;**21**:289-307

[29] Sobolev V, Wade RC, Vriend G, Edelman M. Molecular docking using

surface complementarity. Proteins. 1996; **25**(1):120-129

[30] Murray CW, Rees DC. The rise of fragment-based drug discovery. Nature Chemistry. 2009;**1**(3):187-192

[31] Blaney JM, Dixon JS. Distance geometry in molecular modeling. In: Drug Discov Design. 1994. pp. 299-335. DOI: 10.1002/9780470125823.ch6. ISSN: 9780471188667

[32] Yamada Y. Nucleic acid drugs—Current status, issues, and expectations for exosomes. Cancers (Basel). 2021;**13**(19):5002. DOI: 10.3390/cancers13195002. PMID: 34638486; PMCID: PMC8508492

[33] The Structure of Nucleic Acids and Their Role in Protein Synthesis - PMC [Internet]. [cited 2022 Jul 7]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2603672/

[34] Carrey EA, Perrett D. Nucleic acids, purine, and pyrimidine nucleotides and nucleosides: Physiology, toxicology, and dietary sources. Encyclopedia of Human Nutrition. 2013;**3–4**:189-196

[35] Kiwerska K, Szyfter K. DNA repair in cancer initiation, progression, and therapy—A double-edged sword. Journal of Applied Genetics. 2019;**60**(3):329. [Internet] [cited 2022 Jul 7]. Available from: /pmc/articles/PMC6803590/

[36] Alhmoud JF, Woolley JF, al Moustafa AE, Malki MI. DNA damage/repair Management in Cancers. Cancers (Basel). 2020;**12**(4):1050. DOI: 10.3390/cancers12041050. PMID: 32340362; PMCID: PMC7226105

[37] Lee JH, Xiong F, Li W. Enhancer RNAs in cancer: Regulation, mechanisms and therapeutic potential. RNA Biology. 2020;**17**(11):1550-1559 [Internet] [cited 2022 Jul 7]. Available from: https://pubmed.ncbi.nlm.nih.gov/31916476/

[38] Knoch J, Kamenisch Y, Kubisch C, Berneburg M. Rare hereditary diseases with defects in DNA-repair. European Journal of Dermatology. 2012;**22**(4):443-455 [Internet]. [cited 2022 Jul 7]. Available from: https://pubmed.ncbi.nlm.nih.gov/22436139/

[39] Babur Ö, Demir E, Gönen M, Sander C, Dogrusoz U. Discovering modulators of gene expression. Nucleic Acids Research. 2010;**38**(17):5648-5656 [Internet]. [cited 2022 Jul 7]. Available from: https://academic.oup.com/nar/article/38/17/5648/1029365

[40] Zamani F, Suzuki T. Synthetic RNA modulators in drug discovery. Journal of Medicinal Chemistry. 2021;**64**(11):7110-7155 [Internet]. [cited 2022 Jul 7]. Available from: https://pubs.acs.org/doi/abs/10.1021/acs.jmedchem.1c00154

[41] Bhagavan NV, Ha CE. Structure and Properties of DNA, In: Bhagavan NV, Ha CE. editors. Essentials of Medical Biochemistry. Academic Press. 2011. pp. 275-286. Available from: https://www.sciencedirect.com/science/article/pii/B9780120954612000217. DOI: 10.1016/B978-0-12-095461-2.00021-7. ISBN 9780120954612

[42] Coll M, Frederick CA, Wang AH, Rich A. A bifurcated hydrogen-bonded conformation in the d(a.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. Proceedings of the National Academy of Sciences of the United States of America. 1987;**84**(23):8385-8389 [Internet]. [cited 2022 Jul 8]. Available from: https://www.pnas.org

[43] Hermann T, Tereshko V, Skripkin E, Patel DJ. Apramycin recognition by the

human ribosomal decoding site. Blood Cells, Molecules, and Diseases. 2007; **38**(3):193-198

[44] Haider SM, Parkinson GN, Neidle S. Structure of a G-quadruplex–ligand complex. Journal of Molecular Biology. 2003;**326**(1):117-125

[45] Trachman RJ, Autour A, Jeng SCY, Abdolahzadeh A, Andreoni A, Cojocaru R, et al. Structure and functional reselection of the mango-III fluorogenic RNA aptamer. Nature Chemical Biology. 2019;**15**(5):472-479 [Internet]. [cited 2022 Jul 8]. Available from: https://www.nature.com/articles/s41589-019-0267-9

[46] Phan AT, Guéron M, Leroy JL. The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere. Journal of Molecular Biology. 2000;**299**(1):123-144

[47] Snoussi K, Nonin-Lecomte S, Leroy JL. The RNA i-motif. Journal of Molecular Biology. 2001;**309**(1):139-153

[48] Weisenseel JP, Reddy GR, Marnett LJ, Stone MP. Structure of the 1, N2-propanodeoxyguanosine adduct in a three-base DNA hairpin loop derived from a palindrome in the salmonella typhimurium hisD3052 gene. Chemical Research in Toxicology. 2002;**15**(2): 140-152 [Internet]. [cited 2022 Jul 8]. Available from: https://pubs.acs.org/doi/abs/10.1021/tx010107f

[49] Zhang H, Fountain MA, Krugh TR. Structural characterization of a six-nucleotide RNA hairpin loop found in Escherichia coli, r(UUAAGU). Biochemistry. 2001;**40**(33):9879-9886 [Internet]. [cited 2022 Jul 8]. Available from: https://pubs.acs.org/doi/abs/10.1021/bi011226x

[50] Al-Hashimi HM. NMR studies of nucleic acid dynamics. Journal of Magnetic Resonance. 2013;237:191. [Internet]. [cited 2022 Jul 8]. Available from: /pmc/articles/PMC3984477/

[51] Wang X, Alnabati E, Aderinwale TW, Maddhuri Venkata Subramaniya SR, Terashi G, Kihara D. Detecting protein and DNA/RNA structures in cryo-EM maps of intermediate resolution using deep learning. Nature Communications. 2021; **12**(1):1-9 [Internet]. [cited 2022 Jul 8]. Available from: https://www.nature.com/articles/s41467-021-22577-3

[52] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Research. 2000;**28**(1):235-242 [Internet]. [cited 2022 Jul 8]. Available from: https://academic.oup.com/nar/article/28/1/235/2384399

[53] PDB Statistics: DNA-only Structures Released Per Year [Internet]. [cited 2022 Jul 8]. Available from: https://www.rcsb.org/stats/growth/growth-dna

[54] Feng Y, Yan Y, He J, Tao H, Wu Q, Huang SY. Docking and scoring for nucleic acid–ligand interactions: Principles and current status. Drug Discovery Today. 2022;**27**(3):838-847

[55] Sanderson BA, Sowersby DS, Crosby S, Goss M, Lewis LK, Beall GW. Charge density and particle size effects on oligonucleotide and plasmid DNA binding to nanosized hydrotalcite. Biointerphases. 2013;**8**(1):1-11 [Internet]. [cited 2022 Jul 10]. Available from: https://pubmed.ncbi.nlm.nih.gov/24706120/

[56] Bao L, Zhang X, Jin L, et al. Flexibility of nucleic acids: From DNA to RNA. Chinese Physics B. 2015;**25**(1): 018703 [Internet]. [cited 2022 Jul 10]. Available from: https://iopscience.iop.

org/article/10.1088/1674-1056/25/1/018703

[57] Mordor [Internet]. [cited 2022 Jul 10]. Available from: http://mondale.ucsf.edu/science/mordor.html.sav

[58] Stagno JR, Liu Y, Bhandari YR, Conrad CE, Panja S, Swain M, et al. Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. Nature. 2017;**541**(7636): 242-246 [Internet]. [cited 2022 Jul 10]. Available from: https://www.nature.com/articles/nature20599

[59] Lopéz-Blanco JR, Garzón JI, Chacón P. iMod: Multipurpose normal mode analysis in internal coordinates. Bioinformatics. 2011;**27**(20):2843-2850 [Internet]. [cited 2022 Jul 10]. Available from: https://academic.oup.com/bioinformatics/article/27/20/2843/202794

[60] Tessaro F, Scapozza L. How "Protein-Docking" Translates into the New Emerging Field of Docking Small Molecules to Nucleic Acids? Molecules. 2020;**25**(12):2749. DOI: 10.3390/molecules25122749. PMID: 32545835; PMCID: PMC7355999

[61] Guilbert C, James TL. Docking to RNA via root-Mean-Square-deviation-driven energy minimization with flexible ligands and flexible targets. Journal of Chemical Information and Modeling. 2008;**48**(6):1257-1268 [Internet]. [cited 2022 Jul 10]. Available from: http://pmc/articles/PMC2910576/

[62] Martella M, Pichiorri F, Chikhale RV, Abdelhamid MAS, Waller ZAE, Smith SS. i-motif formation and spontaneous deletions in human cells. Nucleic Acids Research. 2022;**50**(6, 55):3445 [Internet]. [cited 2022 Jul 10]. Available from: https://academic.oup.com/nar/article/50/6/3445/6543540

[63] King JJ, Irving KL, Evans CW, Chikhale RV, Becker R, Morris CJ, et al. DNA G-Quadruplex and i-motif structure formation is interdependent in human cells. Journal of the American Chemical Society. 2020;**142**(49): 20600-20604 [Internet]. [cited 2022 Jul 10]. Available from: https://pubs.acs.org/doi/abs/10.1021/jacs.0c11708

[64] Wang M, Yu Y, Liang C, Lu A, Zhang G. Recent Advances in Developing Small Molecules Targeting Nucleic Acid. International Journal of Molecular Sciences. 2016;**17**(6):779. DOI: 10.3390/ijms17060779. PMID: 27248995; PMCID: PMC4926330

[65] Krüger A, Zimbres FM, Kronenberger T, Wrenger C. Molecular modeling applied to nucleic acid-based molecule development. Biomolecules. 2018;**8**(3):83 [Internet]. [cited 2022 Jul 19]. Available from: https://www.mdpi.com/2218-273X/8/3/83/htm

[66] Holt PA, Chaires JB, Trent JO. Molecular Docking of Intercalators and Groove-Binders to Nucleic Acids Using Autodock and Surflex. [cited 2022 Jul 19]; Available from: https://pubs.acs.org/sharingguidelines

[67] Chikhale RV, Guneri D, Yuan R, Morris CJ, ZAE W. Identification of sugar-containing natural products that interact with i-motif DNA. Bioorganic & Medicinal Chemistry Letters. 2022;**73**:128886 [Internet]. [cited 2022 Jul 23]. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0960894X22003626

[68] Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA. Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. Proceedings of the National Academy of Sciences of the United States of America. 1992;**89**(6):

2195-2199 [Internet]. [cited 2022 Jul 19]. Available from: https://www.pnas.org

[69] Gabb HA, Jackson RM, Sternberg MJE. Modelling protein docking using shape complementarity, electrostatics and biochemical information. Journal of Molecular Biology. 1997;**272**(1):106-120

[70] Carter P, Lesk VI, Islam SA, Sternberg MJE. Protein–protein docking using 3D-dock in rounds 3, 4, and 5 of CAPRI. Proteins: Structure, Function, and Bioinformatics. 2005; **60**(2):281-288 [Internet]. [cited 2022 Jul 19]. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/prot.20571

[71] Protein Docking using Spherical Polar Fourier Correlations [Internet]. [cited 2022 Jul 19]. Available from: https://onlinelibrary.wiley.com/doi/epdf/10.1002/%28SICI%291097-0134%2820000501%2939%3A2%3C178%3A%3AAID-PROT8%3E3.0.CO%3B2-6

[72] Roberts VA, Thompson EE, Pique ME, Perez MS, ten Eyck LF. DOT2: Macromolecular docking with improved biophysical models. Journal of Computational Chemistry. 2013;**34**(20): 1743-1758 [Internet]. [cited 2022 Jul 21]. Available from: https://pubmed.ncbi.nlm.nih.gov/23695987/

[73] Dominguez C, Boelens R, AMJJ B. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. Journal of the American Chemical Society. 2003; **125**(7):1731-1737 [Internet]. [cited 2022 Jul 21]. Available from: https://pubs.acs.org/doi/full/10.1021/ja026939x

[74] Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: Servers for rigid and symmetric docking. Nucleic Acids Research. 2005;**33**(suppl_2):W363-W367 [Internet]. [cited 2022 Jul 21]. Available from: https://academic.oup.com/nar/article/33/suppl_2/W363/2505698

[75] Mashiach E, Schneidman-Duhovny D, Peri A, Shavit Y, Nussinov R, Wolfson HJ. An integrated suite of fast docking algorithms. Proteins: Structure, Function, and Bioinformatics. 2010;**78**(15): 3197-3204 [Internet]. [cited 2022 Jul 21]. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/prot.l22790

[76] Banitt I, Wolfson HJ. ParaDock: A flexible non-specific DNA—Rigid protein docking algorithm. Nucleic Acids Research. 2011;**39**(20):e135-e135 [Internet]. [cited 2022 Jul 21]. Available from: https://academic.oup.com/nar/article/39/20/e135/2409748

[77] Tuszynska I, Magnus M, Jonak K, Dawson W, Bujnicki JM. NPDock: A web server for protein–nucleic acid docking. Nucleic Acids Research. 2015;**43**(W1): W425-W430 [Internet]. [cited 2022 Jul 21]. Available from: https://academic.oup.com/nar/article/43/W1/W425/2467938

[78] Yan Y, Zhang D, Zhou P, Li B, Huang SY. HDOCK: A web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. Nucleic Acids Research. 2017;**45**(W1): W365-W373 [Internet]. [cited 2022 Jul 21]. Available from: https://academic.oup.com/nar/article/45/W1/W365/3829194

[79] Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. Journal of Molecular Biology. 1997;**267**(3):727-748

[80] Pfeffer P, Gohlke H. DrugScore RNA - knowledge-based scoring

function to predict RNA - ligand interactions. Journal of Chemical Information and Modeling. 2007;**47**(5): 1868-1876 [Internet]. [cited 2022 Jul 27]. Available from: https://pubs.acs.org/doi/full/10.1021/ci700134p

[81] Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry. 2010;**31**(2): 455-461 [Internet]. [cited 2022 Jul 21]. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.21334

[82] Morris GM, Ruth H, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of Computational Chemistry. 2009;**30**(16): 2785. [Internet]. [cited 2022 Jul 21]. Available from: /pmc/articles/PMC2760638/

[83] Morris GM, Goodsell DS, Huey R, Olson AJ. Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. Journal of Computer-Aided Molecular Design. 1996;**10**(4):293-304 [Internet]. [cited 2022 Jul 21]. Available from: https://pubmed.ncbi.nlm.nih.gov/8877701/

[84] Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. Proteins. 1990; **8**(3):195-202 [Internet]. [cited 2022 Jul 21]. Available from: https://pubmed.ncbi.nlm.nih.gov/2281083/

[85] Jain AN. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. Journal of Medicinal Chemistry. 2003;**46**(4):499-511 [Internet]. [cited 2022 Jul 21]. Available from: https://pubmed.ncbi.nlm.nih.gov/12570372/

[86] Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S. rDock: A fast, versatile and open source program for docking ligands to proteins and nucleic acids. PLoS Computational Biology. 2014;**10**(4): 1003571 [Internet]. [cited 2022 Jul 22]. Available from: www.ploscompbiol.org

[87] Feng Y, Zhang K, Wu Q, Huang SY. NLDock: A fast nucleic acid−ligand docking algorithm for modeling RNA/DNA−ligand complexes. Journal of Chemical Information and Modeling. 2021;**61**:4771-4782. [Internet]. [cited 2022 Jul 21]. DOI: 10.1021/acs.jcim.1c00341

[88] Feng Y, Huang S. ITScore-NL: An Iterative Knowledge-Based Scoring Function for Nucleic Acid-Ligand Interactions. Journal of Chemical Information and Modeling. 2020;**60**(12): 6698-6708

[89] Philips A, Milanowska K, Łach G, Bujnicki JM. LigandRNA: Computational predictor of RNA–ligand interactions. RNA. 2013;**19**(12):1605-1616 [Internet]. [cited 2022 Jul 22]. Available from: http://rnajournal.cshlp.org/content/19/12/1605.full

[90] Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, et al. DOCK 6: Combining techniques to model RNA-small molecule complexes. RNA. 2009;**15**(6):1219-1230 [Internet]. [cited 2022 Jul 22]. Available from: https://pubmed.ncbi.nlm.nih.gov/19369428/

[91] Aminpour M, Montemagno C, Tuszynski JA. An Overview of Molecular Modeling for Drug Discovery with Specific Illustrative Examples of Applications. Molecules. 2019;**24**(9):

1693. DOI: 10.3390/molecules24091693. PMID: 31052253; PMCID: PMC6539951

[92] Bentham Science Publisher BSP. Scoring functions for protein-ligand docking. Current Protein & Peptide Science. 2006;**7**(5):407-420

[93] Li J, Fu A, Zhang L. An Overview of Scoring Functions Used for Protein-Ligand Interactions in Molecular Docking. Interdisciplinary Sciences. 2019;**11**(2):320-328. DOI: 10.1007/s12539-019-00327-w. Epub 2019 Mar 15. PMID: 30877639

[94] Meli R, Morris GM, Biggin PC. Scoring functions for protein-ligand binding affinity prediction using structure-based deep learning: A review. Frontiers in Bioinformatics. 2022;**2**

[95] Adcock SA, McCammon JA. Molecular dynamics: Survey of methods for simulating the activity of proteins. Chemical Reviews. 2006;**106**(5): 1589-1615

[96] Ain QU, Aleksandrova A, Roessler FD, Ballester PJ. Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2015;**5**(6):405-424. DOI: 10.1002/wcms.1225. Epub 2015 Aug 28. PMID: 27110292; PMCID: PMC4832270

[97] Guedes IA, Pereira FSS, Dardenne LE. Empirical ScoringFunctions for Structure-Based VirtualScreening: Applications, CriticalAspects, and Challenges. Frontiers in Pharmacology. 2018;**9**:1089

[98] Fujimoto KJ, Minami S, Yanai T. Machine-learning- and knowledge-based scoring functions incorporating ligand and protein fingerprints. ACS Omega. 2022;**7**(22):19030-19039

[99] Stefaniak F, Bujnicki JM. AnnapuRNA: A scoring function for predicting RNA-small molecule binding poses. PLoS Computational Biology. 2021;**17**(2):e1008309. DOI: 10.1371/journal.pcbi.1008309. PMID: 33524009; PMCID: PMC7877745

[100] Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, Schmidtke P, et al. rDock: A fast, versatile and open source program for docking ligands to proteins and nucleic acids. PLoS Computational Biology. 2014;**10**(4):1-7

[101] Morley SD, Afshar M. Validation of an empirical RNA-ligand scoring function for fast flexible docking using Ribodock. Journal of Computer-Aided Molecular Design. 2004;**18**(3):189-208

[102] Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. Accounts of Chemical Research. 2000;**33**(12):889-897

[103] Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opinion on Drug Discovery. 2015;**10**(5):449-461

[104] Aamir M, Singh VK, Dubey MK, Meena M, Kashyap SP, Katari SK, et al. In silico prediction, characterization, molecular docking, and dynamic studies on fungal SDRs as novel targets for searching potential fungicides against fusarium wilt in tomato. Frontiers in Pharmacology. 2018;**9**(OCT):1038

[105] UCSF. Flexible Docking with MORDOR [Internet]. [cited 2022 Jul 31].

Available from: http://mondale.ucsf.edu/index_mordor.html

[106] Jiang Y, Chen SJ. RLDOCK method for predicting RNA-small molecule binding modes. Methods. 2022;**197** (January 2021):97-105

[107] Zhou Y, Jiang Y, Chen SJ. RNA–ligand molecular docking: Advances and challenges. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2022;**12**(3):1-32

[108] Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: Current status and future challenges. Proteins. 2006; **65**(1):15-26

[109] Ballester PJ, Mitchell JBO. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. Bioinformatics. 2010;**26**(9):1169-1175 [Internet]. [cited 2022 Jul 28]. Available from: https://academic.oup.com/bioinformatics/article/26/9/1169/199938

[110] Nguyen DD, Wei GW. AGL-score: Algebraic graph learning score for protein-ligand binding scoring, ranking, docking, and screening. Journal of Chemical Information and Modeling. 2019;**59**(7):3291-3304 [Internet]. [cited 2022 Jul 28]. Available from: https://pubs.acs.org/doi/full/10.1021/acs.jcim.9b00334

[111] Li L, Wang B, Meroueh SO. Support vector regression scoring of receptor-ligand complexes for rank-ordering and virtual screening of chemical libraries. Journal of Chemical Information and Modeling. 2011;**51**(9):2132-2138 [Internet]. [cited 2022 Jul 28]. Available from: https://pubs.acs.org/doi/full/10.1021/ci200078f

[112] Durrant JD, Friedman AJ, Rogers KE, McCammon JA. Comparing neural-network scoring functions and the state of the art: Applications to common library screening. Journal of Chemical Information and Modeling. 2013;**53**(7):1726-1735 [Internet]. [cited 2022 Jul 28]. Available from: https://pubs.acs.org/doi/full/10.1021/ci400042y

[113] Jiménez J, Doerr S, Martínez-Rosell G, Rose AS, de Fabritiis G. DeepSite: Protein-binding site predictor using 3D-convolutional neural networks. Bioinformatics. 2017;**33**(19):3036-3042 [Internet]. [cited 2022 Jul 28]. Available from: https://academic.oup.com/bioinformatics/article/33/19/3036/3859178

[114] Lim J, Ryu S, Park K, Choe YJ, Ham J, Kim WY. Predicting drug-target interaction using a novel graph neural network with 3D structure-embedded graph representation. Journal of Chemical Information and Modeling. 2019;**59**(9):3981-3988 [Internet]. [cited 2022 Jul 28]. Available from: https://pubs.acs.org/doi/full/10.1021/acs.jcim.9b00387

[115] Chhabra S, Xie J, Frank AT. RNAPosers: Machine learning classifiers for ribonucleic acid-ligand poses. Journal of Physical Chemistry B. 2020;**124**(22): 4436-4445 [Internet]. [cited 2022 Jul 27]. Available from: https://pubs.acs.org/doi/full/10.1021/acs.jpcb.0c02322

[116] Stefaniak F, Bujnicki JM. AnnapuRNA: A scoring function for predicting RNA-small molecule binding poses. PLoS Computational Biology. 2021;**17**(2):e1008309 [Internet]. [cited 2022 Jul 29]. Available from: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008309