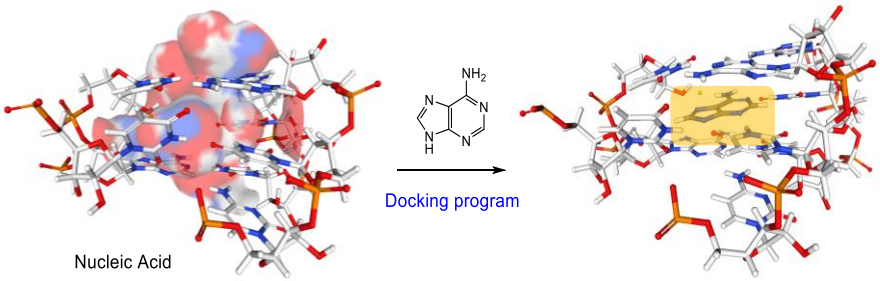
Challenges and Current Status of Computational Methods for Docking Small Molecules to Nucleic Acids.

Jiaying Luo,[a] Wanlei Wei,[a] Jérôme Waldispühl,[b] Nicolas Moitessier[a,]*

[a] *Department of Chemistry and* [b] *School of Computer Science, McGill University, 801 Sherbrooke St. W., Montréal, Québec H3A 0B8, Canada*

Since the development of the first docking program in 1982, their use in *in silico* screening for potentially bioactive molecule discovery has become a common strategy in academia and pharmaceutical industry. Up until recently, their use has largely focused on drugs binding to proteins. However, with the relatively recent discovery of promising drug targets in nucleic acids, including RNA riboswitches, DNA G-quadruplexes, and extended repeats in RNA, there has been a greater interest in developing drugs for nucleic acids. However, due to the major biochemical and physical differences in charges, binding pockets, and solvation, existing docking programs, developed for proteins, face difficulties when adopted directly for nucleic acids. In this review, we cover the current field of *in silico* docking to nucleic acids, available programs, as well as challenges faced in the field.



Contents

* Corresponding author. Tel.: +1-514-398-8543; fax: +1-514-398-3797; e-mail: nicolas.moitessier@mcgill.ca

1. Introduction

1.1. Nucleic acids as drug targets

Nucleic acids (NAs) play important roles in a variety of essential cellular processes, including cellular reproduction and protein biosynthesis (Figure 1). DNA binders could disturb the DNA replication process which affects cell proliferation or regulate the transcription process and may finally inhibit gene expression. Similarly, RNA binders are able to meddle with the transcription and translation processes. Consequently, nucleic acids are potential drug targets for a number of diseases [1]. They are of particular importance in areas of anticancer and antibacterial therapy.

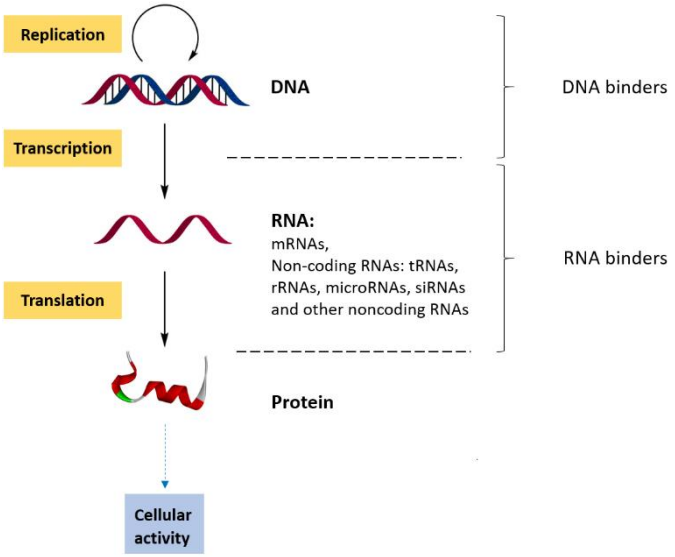


Figure 1: Small molecules that target nucleic acids could ultimately affect nucleic acid functions and its corresponding cellular processes.

1.2. DNA duplexes as drug target.

Small molecules may bind to DNA without causing permanent DNA damage but are, nonetheless, able to induce tumor cell death. As an example, psoralen interacts with double stranded DNA and interferes with DNA replication and transcription (Figure 2). Ultimately, protein biosynthesis is interrupted. Psoralen is used as a drug to treat skin diseases such as vitiligo and psoriasis [1]. Targeting DNA has been the focus of anti-cancer, anti-viral and anti-bacterial drug discovery programs for many years. Quinacrine is recognized as a potential anticancer agent which interacts with DNA without causing DNA damage, and was shown to possess tumor suppression properties [2].

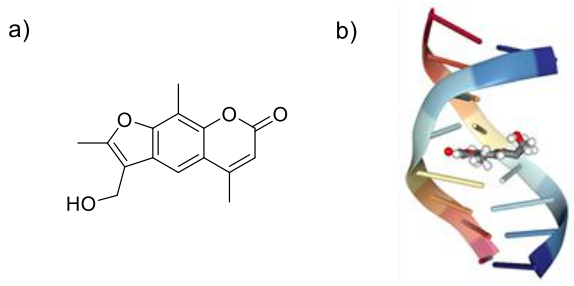


Figure 2: a) Chemical structure of a psoralen molecule (4'-(Hydroxymethyl)-4,5',8-trimethylpsoralen) and b) 3D structure of a DNA duplex bound by psoralen molecule.

1.3. DNA G-Quadruplexes as Targets.

Guanine-rich single strands, observed at the 5' ends of telomeres and promoters, may fold and form stacks of planar G-quartets held by Hoogsteen hydrogen networks and stabilized by cations [3, 4].

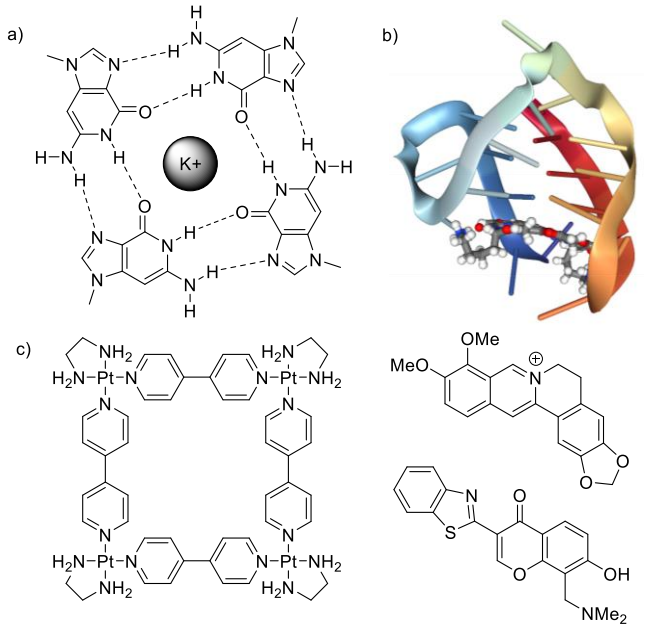


Figure 3: a) G-quadruplex structure; b) Crystal structure of a G-quadruplex bound to a ligand (PDB 2MB3); c) selected ligands.

These unique structures are processed by telomerase, an enzyme active in several cancer cells. Stabilization of the G-quadruplex structures may inhibit telomerase activity and may inhibit the transcription of some oncogenes. While several large polyaromatic compounds [5, 6] binding to these G-quadruplexes were discovered, some more drug-like molecules were also found to target G-quadruplexes [7, 8].

1.4. RNA as drug target.

RNA is upstream of gene-expression pathways, so binding to RNA molecules may also affect the biosynthesis of proteins, which results in disruption of the cell activity. For decades, bacterial proteins have been considered as antibiotic drug target. Non-coding RNAs are central to many cellular processes, making them promising targets for antibiotic drug discovery [9-11]. Riboswitch is one type of non-coding RNAs. Embedded in the 5′-untranslated regions (UTRs) of the mRNAs, it regulates gene expression by interacting with a specific ligand, usually a metabolite. Howe and colleagues discovered that ribocil (Figure 4) can bind to flavin mononucleotide (FMN)-sensing riboswitch, and therefore results in the repression of the corresponding gene expression and in the inhibition of further bacterial infection [12]. This work revealed that riboswitches are a promising class of antibiotic drug targets [12]. For example, TPP-sensing riboswitch regulates the production of TPP (Thiamine pyrophosphate, Figure 4), which is an essential metabolite in all

organisms. The analogue of TPP, pyrithiamine pyrophosphate has been shown to have an antibiotic effect by binding to TPP riboswitch and halting the gene expression of TPP (Figure 4) [13].

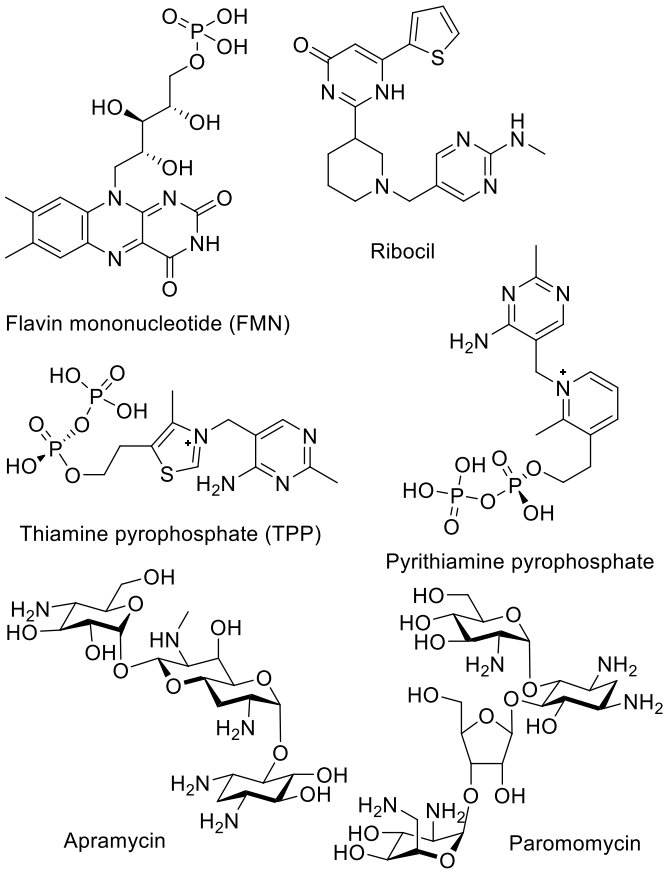


Figure 4: Selected chemical structures of molecules binding to nucleic acids.

Aminoglycosides serve as another class of clinically important antibiotics. They interrupt protein biosynthesis through binding to prokaryotic ribosomal RNA, resulting in bacterial cell death [14]. Aminoglycosides are positively charged, and interact favorably with the negatively charged nucleic acids. Aminoglycosides have been used as antibacterial drugs for the past 50 years. For example, the antibiotics, apramycin [15] and neomycin [16], are often used clinically to treat bacterial infections. More specifically, the former is used in veterinary medicine, being effective against E. Coli, K. pneumoniae, and P. aeruginosa [15] while the latter is used mainly for intestinal infections, especially as preventative medicine against hepatic encephalopathy and hypercholesterolemia [16] (Figure 3e and 3f).

1.5. Extended Repeats as targets.

While repeating subunits of 20-30 nucleotides are normally found in the mRNA transcripts of healthy cells, repeats exceeding the latter limit may suggest diseased phenotypes [17]. One consequence of these extended repeats, when found in mRNA, is that they are thought to sequester essential proteins to the region, including transcription factors and spliceosome complex proteins. This makes these vital proteins unavailable to other regions of the mRNA, and may lead to alternatively spliced isoforms of the protein. Myotonic dystrophy is currently an incurable disease caused by r(CUG)^{exp} extended repeats found on aberrant DM1 protein kinase mRNA transcript. Consequently, this sequesters the vital Muscleblind-like protein 1 away from its intended splice site, leading to defects. It should be noted that extended repeats have been the most successfully targeted mammalian diseases owing to the unique repeating sequences of RNA, which allows for greater specificity in the design of drugs. Usually, multivalent (repeating subunits) drugs are used to target extended repeats due to their high specificity. For example, DCC11 and Bisamidinium 9 are inhibitors of aberrant DM1 protein kinases (Figure 5) [17].

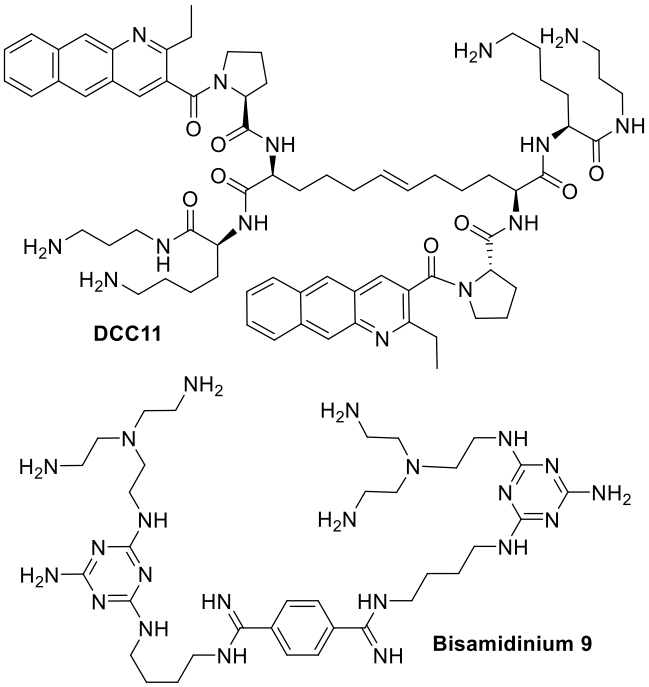


Figure 5. Structure of DCC11 and Bisamidinium **9.**

2. Nucleic acid-ligand complexes

2.1. Interactions

While nucleic acids may bind small molecules, an understanding of the interactions between small molecules and nucleic acids is important for designing and optimizing ligands. The binding between small molecules and nucleic acids are either covalent or non-covalent in nature with more research focusing on the latter. In general, nucleic acid-ligand interactions could be characterized into three different binding modes: 1) nonspecific electrostatic interactions, 2) groove binding, 3) intercalation and stacking (Figure 5) [2, 18].

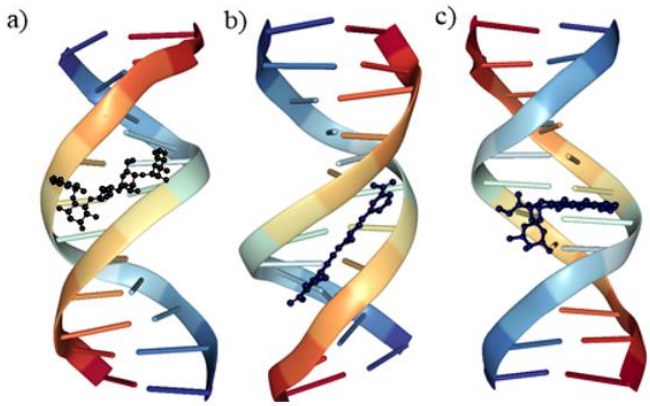


Figure 6: Selected examples of 3D representation of non-covalent ligand binding modes with nucleic acids: a) electrostatic, b) minor groove binding and c) intercalation.

2.2. Polar interactions.

The electrostatic interactions in the structure are often taken into consideration when designing small molecules which strongly bind to nucleic acids. This approach aims to neutralize the negatively charged sugar-phosphate backbone together with positively charged ions (e.g. Mg^{2+}, K^{+} or Ca^{2+}) and polyamines (e.g. spermine) clustered around the backbone. Purely electrostatic interactions complement hydrogen bonding networks, taking advantage of a large number of hydrogen bond donors and acceptors found in nucleic acids. These hydrogen bonding networks allows for the design of sequence-specific binders. For example, highly positively charged aminoglycosides often bind to bacterial RNA through a number of ionic interactions,

hydrogen bonds and interaction with bridging water molecules (water molecules interacting with both the ligand and the NA).

The overall geometry of the Watson-Crick base pairs results in the formation of major grooves and minor grooves within the DNA and RNA helices. Although DNA is known to have three forms of double helical structures: A-form, B-form and Z-form, DNA generally forms B-form helix under physiological condition, while the A-form RNA is the predominant conformation of double-stranded RNA (Figure 6).

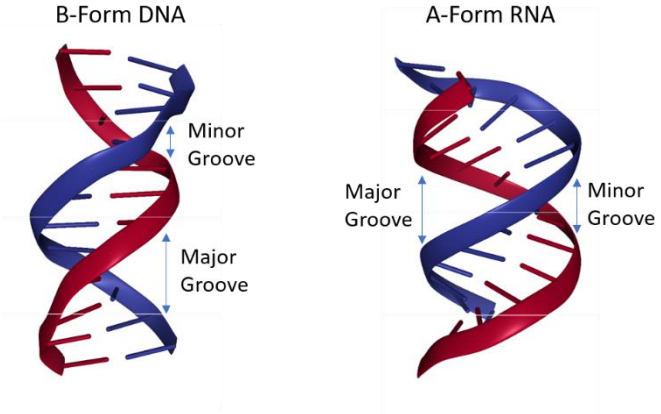


Figure 7: 3D structures of B-form DNA and A-form RNA.

Generally, the major groove of DNA is shallower and wider than that of RNA, which allows a greater number of H-bonding donor and acceptor sites. Additionally, DNA offers a larger pocket for bulky molecules such as proteins and peptides. The methyl group of thymine could also facilitate hydrophobic interactions with the side-chains of alanine, isoleucine and leucine to improve protein binding. In contrast, the minor groove of DNA is spatially shallower and narrower than that of RNA and has less binding sites. However, it is usually unoccupied and therefore available for small drug molecules and ligands to bind upon displacement of water [19-21]. In the fields of anticancer and antibiotic drugs, considerable research effort has been focused on DNA minor groove binders [1, 22]. The wider and shallower minor groove of RNA (compared to DNA), which has a lower electrostatic potential, favors binding to proteins through van der Waals interactions and hydrophobic interactions between the groove and the hydrophobic residues of amino acids. However, compared to DNA, RNA contains uracil which lacks the methyl group found on thymine. This makes RNA less likely to bind proteins via hydrophobic interactions, compared to DNA. Small molecules usually bind to the deeper and narrower major groove of RNA which possesses a larger electrostatic potential [23]. Compared with DNA, RNA shows higher structural diversity. As single stranded RNA is a polymeric molecule with strong polyanionic character surrounded by water molecules and positively charged metal ions, ligands are required to possess high polarity in order to disrupt the hydration shell.

RNA folds upon itself, leading to several recognizable secondary structures such as bulges, internal loops, hairpin loops and multibranched junctions (Figure 7) [24]. These secondary structures finally contribute to a high level of diversity in terms of tertiary structures and create unique binding pockets which are able to selectively bind small molecules [25].

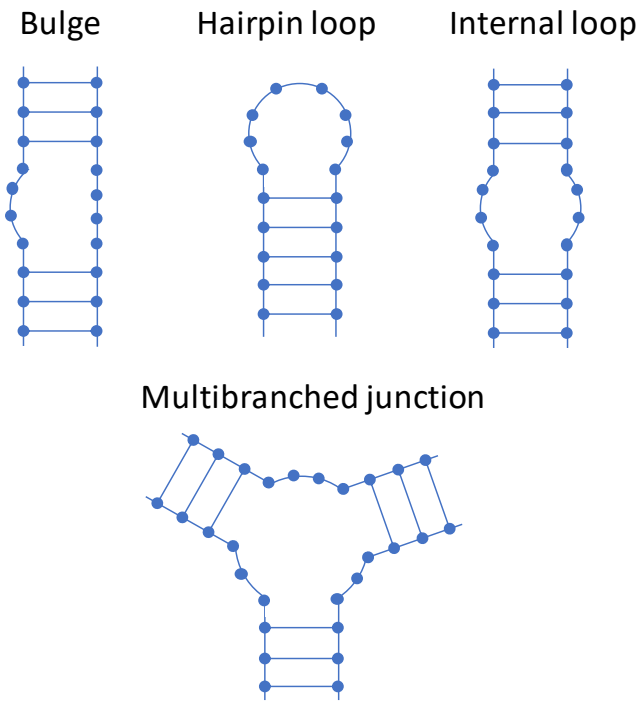


Figure 8: RNA secondary structures mentioned in this review.

2.3. Intercalation and stacking

Intercalation refers to the insertion of polyaromatic molecules, between the stacked base pairs leading to its functional arrest. Intercalators have been used clinically as drugs targeting DNA in cancer therapeutics for decades [26]. Some examples of intercalators are shown in Figure 8.

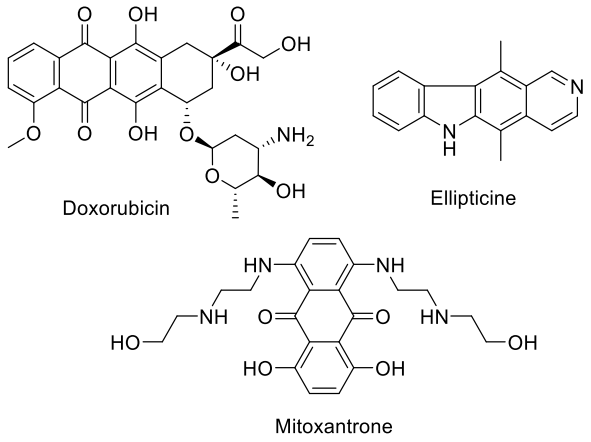


Figure 9: Chemical structures of different intercalators.

Drugs targeting DNA G-quadruplexes are most commonly found to bind to or near the terminal ends the DNA via π-π stacking interactions [27]. In the case of the latter, the small molecule induces minimal conformational changes to the quadruplex during the process of binding. In contrast, in the case of the former, the internal π-π interactions between the nucleic bases must first be perturbed, often leading to large-scale movements. This is also referred to as intercalation. Consequently, the stability conferred by the new conformation must be energetically favorable for this process to occur. In order for stacking to occur with the DNA (and sometimes RNA) bases, quadruplex binders are planar, and therefore contain cyclic, conjugated moieties [28].

2.4. Available structural information

As of June 2018, 465 RNA-ligand co-crystalized structures and 458 DNA-ligand co-crystalized structures between 70 to 1000 Da have been deposited in the Protein Data Bank (PDB) [29], and the number is steadily increasing. These structural data provide opportunities for investigating the molecular interaction between nucleic acids and ligands and enable structure-based computational method for the design of nucleic acid-targeting ligands for specific diseases.

3. Molecular docking for nucleic acids.

3.1. Overview.

Molecular docking is a useful method in structural molecular biology and structure-based drug design [30]. In this review, docking predicts the preferred binding modes of a small molecule (protein-protein and protein-NA docking also exists) in a target macromolecular receptor (e.g., protein or NA). Molecular docking may be used for virtually screening databases containing potential drug candidates, and/or for providing information on the interactions between a ligand and a particular binding site [31]. In general, molecular docking could be divided into two parts: 1) conformational search and 2) scoring. In the first step, several different binding modes of the ligand are generated and in the second step, a scoring function is applied to estimate the binding energy and to select favorable binding modes.

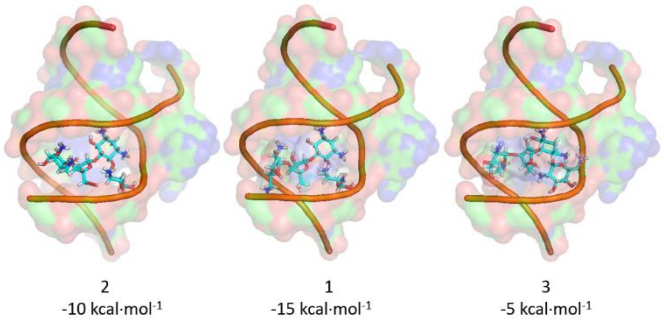


Figure 10: Docking methodology consists of two steps: 1). Binding mode generation (top), and 2). Binding mode scoring and prediction.

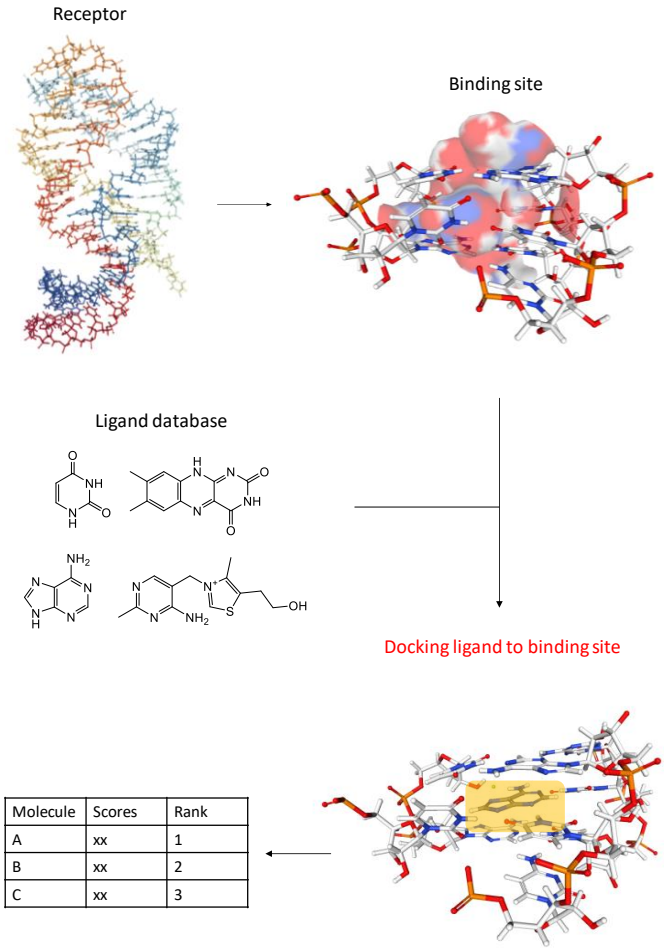


Figure 11. Outline of molecular docking as used to screen potential binders.

In 2008, we reported that there were already over 60 molecular docking methods successfully developed for proteins [30]. Since then, many more have become available. However, docking to nucleic acids is comparatively underdeveloped. Generally, protein-ligand docking programs could be adapted to RNA-ligand docking as RNA-ligand and protein-ligand macromolecular complexes follow similar physicochemical binding principles [32, 33]. However, many proteins contain a well-defined, generally hydrophobic binding site. In contrast, nucleic acids offer more solvent exposed binding pockets with a high charge density and polarity [30]. These differences require modifications to existing protein-ligand docking programs for application to nucleic acid-ligand docking. Several mature docking programs originally developed for protein-ligand docking have been applied to RNA-ligand docking and researchers have also developed new methods and scoring functions, tailored for nucleic-acid docking. Docking programs and scoring functions discussed in this review are listed in Tables 1 and 2 and will be discussed in the following sections.

Table 1. Overview of docking programs discussed in this chapter

Docking program	Developed for nucleic acid-ligand docking?	Reference
AutoDock	No, but it ,could be used with or without adaptation	[34-36]
DOCK 4	No, used with optimization for RNA	[34]
DOCK 6	Extended version of DOCK 5 specifically optimized for RNA	[37, 38]
AutoDock Vina	Used as provided	[39]
GOLD	Used as provided	[40]
Glide	Used as provided	[39, 40]
ICM	Scoring function optimized for RNA	[39]
MORDOR	Yes	[41]
rDOCK	Yes	[39]
RiboDock	Yes	[42]
FITTED	No, used with optimization for RNA	Unpublished results from our group

Table 2. Reported scoring function developed for RNA-ligand docking

Scoring function	Reference
DrugScoreRNA	[38, 43]
LigandRNA	[38]
KScore	[44]
iMDLScore1	[45]
iMDLScore2	[45]

3.2. The challenges of nucleic acid-ligand docking.

3.2.1. From Proteins to NAs

Docking programs have developed for decades with the first version of DOCK reported in 1982.[46] This pioneering work was followed by a myriad of other programs including the most widely used programs AutoDock [47, 48], Glide [49, 50], GOLD [51, 52], FITTED [53, 54] amongst others. Over the years, these programs have redefined the way medicinal chemists design and/or optimize molecules binding to proteins. A vast majority of drug targets are proteins [10] and these were the focus of the docking program development.

Medicinal chemists interested in targeting NAs first looked at the performances of existing protein-ligand docking programs with NAs and determined whether modifications were required. As discussed in a section below, many docking programs were found accurate enough, although they are expected to be more accurate if re-trained with NA-ligand complexes. For example, while AutoDock was used as was, with little to no modifications (apart from incorporating additional parameters) [34], DOCK was revised and DOCK 6.0 incorporated special features for docking small molecules to NAs. First, an improved conformational

sampling method was implemented which improved accuracy when docking larger and more flexible ligands as compared to previous versions (many molecules binding to RNA such as G-quadruplex binders and aminoglycosides are large and flexible). Interestingly, it should be pointed out that this method, optimized for RNA-ligand complexes, showed reduced accuracy on a ligand-protein sets. Second, a major implementation accounting for metal ions and water was added to DOCK (see below).

In contrast, other programs were specifically designed for docking small molecules to NAs. These include MORDOR [41] which specifically address the problem of RNA flexibility and RiboDock [42] (which evolved into rDock [39]).

3.2.2. Scoring functions

Scoring functions were also designed specifically for predicting RNA-ligand affinities. The high polarity of NAs compared to proteins and some unconventional interactions necessitated special treatments (e.g., electrostatic interaction, and solvation). RiboDock includes an empirical scoring function [42] which is a combination of hydrogen bonds, steric interactions, attractive lipophilicity amongst others. Interestingly, a term accounting for interaction between positively charged carbons (e.g., guanidinium) and negatively charged groups (such as carbonyl) was added to model these uncommon electrostatic interactions which may be perceived as stacking instead as discussed by Morley and Afshar. Another term favoring parallel π-system (stacking) was also been introduced. An alternate scoring function, DrugScore^{RNA} has been reported by Pfeffer and Gohlke [43]. This function is an extension of DrugScore, initially developed for scoring protein-ligand complexes [55]. These knowledge-based scoring functions have been trained from statistical analyses of set of complexes favoring interactions observed more often. LigandRNA is also a knowledge-based function [38]. LigandRNA, in contrast to DrugScore^{RNA} considers the directionality of hydrogen bonds.

Computing the entropic energy cost upon ligand binding in a high throughput manner (amenable to several binding modes while docking) has been a major challenge in the development of docking programs and scoring functions. In 2005, Ruvinsky and Kozintsev proposed to use the binding mode cluster size generated by Autodock to estimate the vibrational entropy of any given binding mode. [47] Earlier this year, this approach was revised and applied by McElfresh and Deligkaris to DNA-ligand complexes [35]. Applied to 24 crystal structures, the addition of this term led to significant improvements in the accuracy (from 13 to 18 out of 24 within 2.5Å of the experimental structure).

While similar strategies were used to develop conformational search algorithms and scoring functions than with proteins, additional challenges have appeared which required specific developments.

3.2.3. RNA and DNA flexibility

One major factor that limited the accuracy of nucleic acid-ligand docking is the intrinsic flexibility of nucleic acids. Most of the current methods ignore flexibility of RNA treated as a rigid receptor. However, RNA could have induced fit movements or conformational changes in response to ligand binding. For example, riboswitches are known to have alternative conformations upon binding to ligands. The crystal structures of the purine riboswitch imply that ligand binding induces a local conformational change in the binding pocket (Figure 11). Rotational flip motions of nucleotide allow ligands to form stable hydrogen bonding network with the RNA. With these movements, the ligand is encapsulated inside the binding pocket and becomes nearly solvent inaccessible.

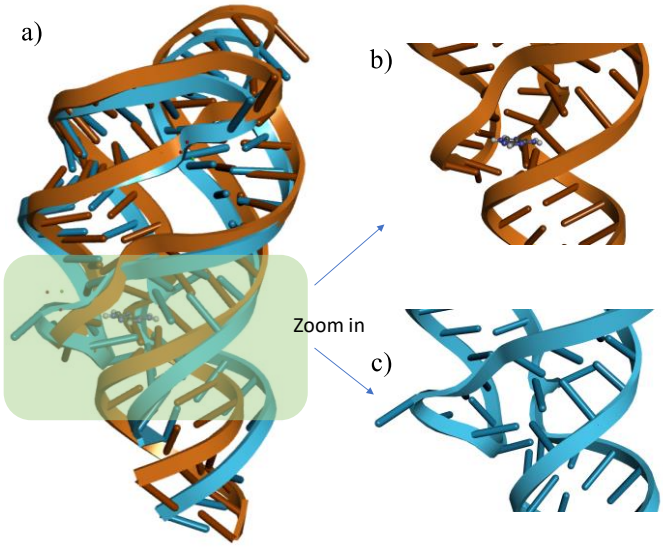


Figure 12: 3D structures of a purine riboswitch. a) overlap of apo (blue, PDB 5E54) and ligand bound (orange, PDB 5SWE) structures. Zoom in view of the binding pocket, b) ligand-bound state and c) apo state.

For optimal docking accuracy, flexibility of the nucleic acid should be considered. Molecular Dynamics has been successfully used in a number of studies on ligand-G-quadruplexes [28, 56] and ligand-RNA, although the throughput of such techniques does not allow for screening of libraries of potential ligands. Over the past decade or so, protein flexibility has been implemented into a few docking programs including our program FITTED [53, 54]. However, none of the techniques were found to be universally applicable [57, 58].

To handle flexibility of RNA, MORDOR allows an induced fit of the ligand and the receptor using molecular mechanics minimization techniques constraining the RNA molecules to the experimental structure (either crystal structure or NMR structure). The ligand is then allowed to explore the surface of receptor after being placed at a certain "hot spot" on the receptor. Depending on the size of receptor and ligand, MORDOR takes 0.5 to 3h per docking run, although this timeframe is highly dependent on the computer used. The energy minimization steps during docking is time-consuming and renders this method hardly applicable to large libraries [41]. This method was found to be highly accurate with an accuracy above 70% within 2.5 Å of the experimental binding mode.

In another study aimed at discovering compounds which target TAR RNA, ICM was used after DOCK. After DOCK quickly matches compounds to the target site and filters out a large number of poses, ICM allows rapid exploration conformational space via Monte Carlo searches using internal coordinates in order to explore the ligand binding in more detail. ICM also account for flexible NAs using internal coordinates and includes terms for hydrogen bonding and solvation [59].

Another option to deal with RNA flexibility is to use ensemble docking, using a set of predetermined RNA conformations. Our group used an ensemble of RNA structures to include the flexibility of RNA in the docking of aminoglycosides into flexible RNA [60]. This was achieved through the combination of a set of AutoDock grids generated to model distinct RNA conformations in grids modeling the conformational ensemble.

In the study carried out by Stelzer et. al, RNA dynamic ensembles were first constructed using a combination of NMR spectroscopy and computational molecular dynamics simulations and then docked small molecules onto the ensembles to solve the problem of taking into account RNA conformational adaptation [61]. Elastic potentials grids for modeling target conformational changes in proteins was introduced by Gohlke and co-workers in 2009 for protein [62] and extended to RNA [63]. In this method, a 3D grid of potential fields pre-calculated using the DrugScoreRNA scoring function is adapted to another conformation by moving grid intersection points in space, keeping the potential field values unchanged. This approach provides accurate and efficient description of RNA-ligand interactions and is twice as successful

as docking into an apo RNA structure. However, the approach is limited to local (i.e., minor) conformational changes. Base movements predominantly governed by rotational flip motions or nucleotide movements leading to an exchange of interaction types, which are often observed in riboswitches, cannot be considered using this approach. Besides, it fails to model the creation of highly electronegative pockets for binding of positively charged groups, a phenomenon which is often observed in RNA [63].

A novel Monte Carlo (MC) algorithm for *ab initio* drug docking to DNA that combines full flexibility of both molecules was proposed by Rohs et al [64]. The flexibility of the entire DNA-ligand complex is defined by MC variables which are combined into collective variables (rigid-body translations and rotations of the ligand relative to the DNA) and internal variables (internal flexibility of the specific ligand). This method can sample the docking geometry without any prior binding site selection [64].

3.2.4. Water Molecules and Metal ions

Another challenge in docking small molecules into nucleic acids is the presence of water molecules and metal ions often necessary for optimal ligand binding. The highly charged nature of nucleic acids leads to strong interactions with water molecules and metal ions. In the case of NA-ligand complexes, bridging water molecules play a key role in the stabilization of the complex, providing additional hydrogen bonding interactions.

Metal ions binding sites were first investigated by Hermann and Westhof using MD simulations in the hope to identify interaction sites for aminoglycosides (i.e., with multiple ammonium groups) [65, 66].

Li. *et al.* later investigated the docking ability of Glide with 25 hydrated RNA-ligand structures. The significant enhancement in success rate (from 44 to 84%, RMSD < 2.5 Å) demonstrated one should include water molecules in docking ligands to nucleic acids [40]. However, the presence of water molecules in docking narrows the conformational search space and can sterically exclude some incorrect modes with higher ranking. As a result, this is expected to improve self-docking (water properly positioned for the co-crystalized ligand) but not cross-docking which may require a different set of water molecules. To address this issue, our group described a novel approach to model key interacting water molecules by including virtually "displaceable" molecules [60]. Alternatively, a GB/SA model (Generalized Born model with solvent accessible surface area (SA) term) had been implemented in the solvation module of DOCK to reproduce the electrostatic solute-water interactions [37, 67]. Metal ions such as Mg^{2+} and Mn^{2+} are often found at the binding site of nucleic acid, serving as a metal bridge coordinating ligand and nucleic acid residues. When Mg^{2+} was taken into account as part of the RNA target, it resulted in the improvement in pose prediction of TPP docking in TPP riboswitch [45]. However, the issue of bridging water molecules and metal ions are not yet addressed in most of the available docking method for nucleic acids.

3.2.5. Highly charged macromolecules

Another challenge is the high number of charges on NAs. In reality, all the phosphate negative formal charges are neutralized by counterions in solutions. However, these are often mobile and could hardly be considered in docking as their location is not fixed. A strategy that our group and others [34, 43] used is to neutralize the phosphate groups themselves by changing their partial charges (e.g., adding +1 to the phosphorus atom) to implicitly consider counter-ions [60].

Another critical aspect is the modulation of their polarization of water molecules when hydrogen bonded to various groups such as phosphates. Each water is expected to have a different polarization (and therefore to interact more or less with ligands) while all the waters are often considered equivalent by docking programs. We have recently found that the binding energy can be drastically different when a ligand is bound to this type of water (unpublished).

3.3. Performance of nucleic acid-ligand docking.

3.3.1. Prediction of binding modes.

The performance of a docking programs may be evaluated based on its ability to accurately predict the binding mode when docking the native (co-crystallized) ligand back to the corresponding macromolecule (i.e., self-docking) or to other crystal structures of the same macromolecule (i.e., same sequence) but co-

crystallized with a different ligand (i.e., cross-docking). A scheme illustrating self-docking and cross-docking is present in Figure 12. The root mean square deviation (RMSD) between the docked ligand structure and experimentally co-crystalized ligand structure is used as a metric to determine the accuracy of the docking program to predict binding modes. In some cases, the docking program is also assessed for its ability to predict binding affinities or rank ligand by affinity (e.g., virtual screening) [32].

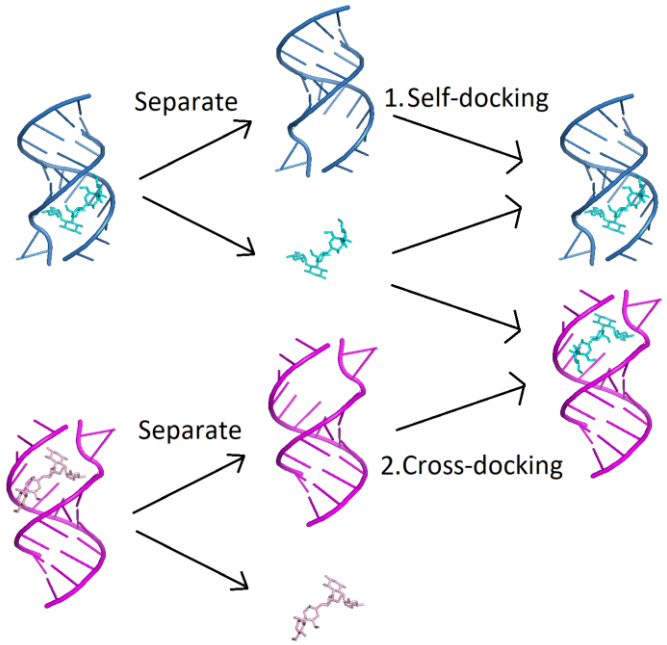


Figure 13: Self-docking (1) and cross-docking (2).

While there is a large body of work in the field of small molecule-protein docking, literature data is scarce with NA-ligand docking and it is a lot more difficult to compare different docking programs as different (and often quite small) test data sets or assessment criteria were used in the reported evaluation studies.

Most validation studies reported the success rates of predicting the correct binding modes rather than predicting the binding affinity. As the success rates of different programs may vary with different data sets, the conclusions drawn from such small test datasets have to be taken with great care. As of today, little has been done with truly statistically relevant-sized test sets. In 2004, Detering and Varani[34] showed that DOCK and AutoDock could reproduce the experimental binding modes within 2.5 Å with success rates of 60% out of 16 RNA-ligand complexes. In a test set of 25 DNA-ligand complexes, AutoDock yielded a success rate of 68% [35]. Considering the same RMSD cutoff as successful docking, AutoDock Vina was used without any specific optimization for NAs and achieved a success rate of 29% in a set of 56 RNA-ligand complexes, while Glide and rDock obtained success rates of 18 and 54%, respectively [39]. Li et al. carried out comparative experiments between Glide and GOLD programs with a test set of 60 RNA-ligand complexes and obtained success rates of 60% and 62%, respectively (RMSD < 2.5 Å) [40]. Our group extended AutoDock by taking account of the role of water molecules in the binding of ligand to RNA and the flexibility of these macromolecules. We were able to successfully dock a small set of 11 aminoglycosides with an average RMSD of 1.41 Å back into their receptors, while the average RMSD was 3.25 Å by the original program [60]. DOCK 6 was reported with an extremely high success rate of 80% (within 2 Å). However, the training set consisted of only 10 RNA-ligand complexes with less than seven rotatable bonds [37]. MORDOR, a program specially designed for RNA-ligand docking, reproduced experimental binding modes within 2.5 Å with a success rate of 74% out of 57 cases [41]. RiboDock was tested with 10 RNA-ligand complexes and achieved 50% success rate within 2 Å [68]. In self-docking experiments of 31 RNA-ligand complexes, DrugScoreRNA identified the first scoring ranked ligand pose within 2 Å with a success rate of 42% [69]. LigandRNA was able to find the ligand poses within 2 Å in 36% out of 42 RNA-ligand complexes, while DOCK6 obtained 36% and DrugScoreRNA 31% [38].

A comprehensive study was carried out by Chen et al., comparing 5 docking programs and 11 scoring functions [45]. Although high success rates (73%) were obtained from GOLD and rDOCK, the authors considered at least one pose with RMSD < 3 Å out the five top-ranked pose as successful docking, a very loose criterion compared to other studies.

With the number of available structures, evaluation studies are nowadays carried out on larger sets. In 2011, ICM was able to find 53% of 96 RNA-ligand structures were successfully reproduced within 2.5 Å [61]. Our group recently carried out self-docking experiments with 229 RNA-ligand complexes using FITTED [53, 54]. The success rate increased from 77 to 83% (RMSD<2.5 Å) after considering all possible hydrogen donor and acceptor from nucleic acid and Mg^{2+} and Mn^{2+} as part of the receptor (unpublished results).

Drug –DNA intercalation is often employed in cancer therapies. AutoDock was able to successfully distinguish between the intercalation site and the minor-groove site and reproduce the binding mode within 2 Å with a success rate of 80% (a datasete of 67 DNA-intercalator complexes) [70, 71] By using DNA sequence and intercalation site information, a novel intercalation methodology called 'Intercalate' was able to create the DNA 3D structure with the intercalation site and perform docking at the binding site with associate binding free energies [72].

3.3.2. Distinguishing active from inactive compounds

The ability to predict the binding affinities was assessed in some studies. ICM obtained a linear correlation of R=0.71 between the experimental and estimated binding energies in a set of 48 RNA-ligand complexes [61]. AutoDock obtained an R^2=0.84 in a dataset of RNA-aminoglycoside complexes [36]. New scoring functions developed for RNA-ligand docking such as KScore, iMDLScore1 and iMDLScore2 were investigated for their ability to predict binding affinities. The linear correlation coefficient between the computed and experimental binding energies for RNA-ligand and DNA-complexes achieved by KScore were as high as 0.81 and 0.68 although on very small sets (data sets of 15 and 9 complexes, respectively) [44]. Two empirical scoring functions, iMDLScore1 yielded a good result of scoring in test set of 45 RNA-ligand complexes, with 0.70 of the linear correlation coefficient between the computed and experimental binding affinity while iMDLScore2 yield 0.79 [45].

DOCK and AutoDock were also evaluated in retrospective studies against DNA minor groove [73]. AutoDock was shown to lead to improved enrichment of a library (with an enrichment value SE(f =1%) =86%.) in active compounds over DOCK.

Another study examined GOLD, GLIDE, CDOCKER and AutoDock for DNA minor groove binding with 57 DNA-ligand complexes. In GOLD and GLIDE, the orientation of best score pose and the lowest RMSD pose is close to each other and the deviation of various conformation poses is also smaller than other methods [74].

The studies discussed above can provide us insight into the different programs and scoring functions, including their accuracy, speed and applicable targets, but they still have some limitations. First, the accuracy (success rate) is greatly dependent on the size of the dataset, the selection of the NA-ligand complexes and the criterion of success. In this scenario, the reported success rates are less convincing. Second, although self-docking is a good indication of the performance of a program, it provides limited evidence about the reliability in drug discovery. In terms of discovering new ligands, the program is expected to have the ability to filter out non-binding molecules and properly dock compounds with good affinity to the NA even when the starting NA structure is not exactly adjusted to the ligand (induced-fit binding). However, self-docking only indicates the program's ability of reproducing the experimental binding mode of ligands when the macromolecule structure is already adjusted to them (Figure 12). Thirdly, few studies have reported the predicted binding affinity or how these correlate with experimental binding affinitites. Fourth, limited studies (see next section) have looked at the accuracy of virtual screening (identification of binders within large libraries of small molecules). This, the general applicability and predictive ability of these methods remains elusive. In order to obtain a better understanding on the strengths and weaknesses of NA-ligand docking, more extensive evaluation studies are needed.

4. Application to NA Binder Discovery

The real validation for docking programs is the discovery of novel binders through virtual screening (VS). As early as 1997, Kuntz and co-workers carried out a docking-based VS on RNA major groove. While predicting potency was a major objective, they also computed their specificity by docking to B DNA [75]. 11 compounds were selected, and 3 aminoglycosides were confirmed to stabilize RNA duplexes without affecting DNA duplexes.

In 2010, G-quadruplexes were targeted, and 20,000 compounds were docked using ICM. 5 compounds were selected for testing and fonsecin B was found as a G-quadruplex stabilizer [76]. A known G-quadruplex binder, methylene blue, has also been improved through the screening of analogues and testing of a promising compound uncover a stronger G-quadruplex binder [77]. In fact, G-quadruplex has been the target for a number of successful prospective screening studies including docking as part of the workflow [78-82]. Triplex DNA structures have also been targeted with similar approaches [83].

In 2011, Daldrop et al. screened a library against an adenine-riboswitch (AR). Crystallography revealed the presence of a conserved water molecule, a large array of hydrogen bonds between ligands and AR as well as π-stacking.[84] For this VS campaign, they used DOCK 3.5.54 with minor modifications which was validated on retrospective studies. A selected set of over 2,500 compounds was docked and a selection of 5 top-scoring compounds were experimentally evaluated. 4 out of these 5 compounds exhibited binding affinity to AR in the micromolar range [84]. As another validation, three of these 4 active compounds were co-crystallized with AR and their experimental binding modes were found to be close to those predicted by DOCK during the VS campaign.

5. Conclusion and perspectives.

With ever increasing structural information and growing interest in NAs as drug targets, docking programs were evaluated, or even developed to design novel NA binders as potential drugs. While it was found possible to apply protein-ligand docking methods to the docking of ligands to NAs, the binding mode prediction (rarely exceeding 60%) was comparatively less accurate than for protein-ligand docking (can be as high as 80% for some programs, although this is highly protein-dependent) [32]. One reason of lower success rate is that the flexibility of NAs, a key structural factor in NA-ligand binding, was ignored. Handling NA flexibility is one concerning problem when developing novel docking methods or transferring methods developed specifically for proteins (generally less flexible) to NAs. Water molecules and metal ions are also crucial in ligand binding and can hardly be ignored by docking programs.

So far, most validation studies have been focused on RNA. Compared to RNA, although DNA being less flexible may be less problematic. Although both protein-ligand docking programs and novel NA-ligand docking programs have shown promising performances on NA-ligand docking, there is space for improvement.

In addition, the increasing number of NA-ligand co-crystalized structures provide opportunities for more extensive evaluation studies with larger data sets in order to obtain a better picture on the strengths and weaknesses of NA-ligand docking programs.

Depsite the several limitations, docking-based screens have already helped medicinal chemists identifying novel NA binders. These prospective investigations revealed the potential of docking programs in NA binder discovery campaigns.

Acknowledgements

NM and JW thank CIHR (BOP-149429) and Genome Canada/Genome Québec (B/CB 2015) for financial support. WW thanks the Fonds de recherche du Québec - Nature et technologies for a scholarship (*Bourses de doctorat en recherche*).

References and notes

[1] M.L. Wang, Y.Y. Yu, C. Liang, A.P. Lu, G. Zhang, Recent Advances in Developing Small Molecules Targeting Nucleic Acid, Int. J. Mol. Sci., 17 (2016) 24.

[2] K. Gurova, New hopes from old drugs: revisiting DNA-binding small molecules as anticancer agents, Future Oncol., 5 (2009) 1685.
[3] R. Hänsel-Hertsch, M. Di Antonio, S. Balasubramanian, DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential, Nature Reviews Molecular Cell Biology, 18 (2017) 279.
[4] J.T. Davis, G-Quartets 40 Years Later: From 5'-GMP to Molecular Biology and Supramolecular Chemistry, Angew. Chem., Int. Ed., 43 (2004) 668-698.
[5] M. Franceschin, L. Rossetti, A. D'Ambrosio, S. Schirripa, A. Bianco, G. Ortaggi, M. Savino, C. Schultes, S. Neidle, Natural and synthetic G-quadruplex interactive berberine derivatives, Bioorg. Med. Chem. Lett., 16 (2006) 1707-1711.
[6] R. Kieltyka, J. Fakhoury, N. Moitessier, H.F. Sleiman, Platinum phenanthroimidazole complexes as G-quadruplex DNA selective binders, Chem. Eur. J., 14 (2008) 1145-1154.
[7] S. Cosconati, A. Rizzo, R. Trotta, B. Pagano, S. Iachettini, S. De Tito, I. Lauri, I. Fotticchia, M. Giustiniano, L. Marinelli, C. Giancola, E. Novellino, A. Biroccio, A. Randazzo, Shooting for Selective Druglike G-Quadruplex Binders: Evidence for Telomeric DNA Damage and Tumor Cell Death, J. Med. Chem., 55 (2012) 9785-9792.
[8] Y.-H. Lin, S.-M. Chuang, P.-C. Wu, C.-L. Chen, S. Jeyachandran, S.-C. Lo, H.-S. Huang, M.-H. Hou, Selective recognition and stabilization of new ligands targeting the potassium form of the human telomeric G-quadruplex DNA, Sci. Rep., 6 (2016) 31019.
[9] S. Colameco, M.A. Elliot, Non-coding RNAs as antibiotic targets, Biochem. Pharmacol., 133 (2017) 29-42.
[10] C.M. Connelly, Michelle H. Moon, John S. Schneekloth, The Emerging Role of RNA as a Therapeutic Target for Small Molecules, Cell Chem. Biol., 23 (2016) 1077-1090.
[11] N.F. Rizvi, G.F. Smith, RNA as a small molecule druggable target, Bioorg. Med. Chem. Lett., 27 (2017) 5083-5088.
[12] J.A. Howe, H. Wang, T.O. Fischmann, C.J. Balibar, L. Xiao, A.M. Galgoci, J.C. Malinverni, T. Mayhood, A. Villafania, A. Nahvi, N. Murgolo, C.M. Barbieri, P.A. Mann, D. Carr, E. Xia, P. Zuck, D. Riley, R.E. Painter, S.S. Walker, B. Sherborne, R. de Jesus, W. Pan, M.A. Plotkin, J. Wu, D. Rindgen, J. Cummings, C.G. Garlisi, R. Zhang, P.R. Sheth, C.J. Gill, H. Tang, T. Roemer, Selective small-molecule inhibition of an RNA structural element, Nature, 526 (2015) 672-677.
[13] S. Thore, C. Frick, N. Ban, Structural Basis of Thiamine Pyrophosphate Analogues Binding to the Eukaryotic Riboswitch, J. Am. Chem. Soc., 130 (2008) 8116-8117.
[14] M.-P. Mingeot-Leclercq, Y. Glupczynski, P.M. Tulkens, Aminoglycosides: Activity and Resistance, Antimicrob. Agents and Chemother., 43 (1999) 727-737.
[15] K.P. Smith, J.E. Kirby, Evaluation of apramycin activity against carbapenem-resistant and -susceptible strains of Enterobacteriaceae, Diagn. Microbiol. Infect. Dis., 86 (2016) 439-441.
[16] S.A. Waksman, H.A. Lechevalier, Neomycin, a New Antibiotic Active against Streptomycin-Resistant Bacteria, Including Tuberculosis Organisms, Science, 109 (1949) 305-307.
[17] A. Donlic, A.E. Hargrove, Targeting RNA in mammalian systems with small molecules, Wiley Interdiscip. Rev. RNA, (2018) e1477.
[18] L. Strekowski, B. Wilson, Noncovalent interactions with DNA: An overview, Mut. Res., 623 (2007) 3-13.
[19] G.S. Khan, A. Shah, R. Zia ur, D. Barker, Chemistry of DNA minor groove binding agents, J. Photochem. Photobiol. B, 115 (2012) 105-118.
[20] M.J. Hannon, Supramolecular DNA recognition, Chem. Soc. Rev., 36 (2007) 280-295.
[21] P. Guo, A.A. Farahat, A. Paul, N.K. Harika, D.W. Boykin, W.D. Wilson, Compound Shape Effects in Minor Groove Binding Affinity and Specificity for Mixed Sequence DNA, J. Am. Chem. Soc., (2018).
[22] M.P. Barrett, C.G. Gemmell, C.J. Suckling, Minor groove binders as anti-infective agents, Pharmacol. Therap., 139 (2013) 12-23.
[23] Y. Xie, K. Tam Victor, Y. Tor, The Interactions of Small Molecules with DNA and RNA, 2010.
[24] N.B. Leontis, E. Westhof, Geometric nomenclature and classification of RNA base pairs, RNA, 7 (2001) 499-512.
[25] S. Fulle, H. Gohlke, Molecular recognition of RNA: challenges for modelling interactions and plasticity, J. Mol. Recognit., 23 (2010) 220-231.

[26] A. Mukherjee, W.D. Sasikala, Chapter One - Drug–DNA Intercalation: From Discovery to the Molecular Mechanism, in: T. Karabencheva-Christova (Ed.) Adv. Protein Chem. Struct. Biol., Academic Press, 2013, pp. 1-62.
[27] K.J. Castor, Z.M. Liu, J. Fakhoury, M.A. Hancock, A. Mittermaier, N. Moitessier, H.F. Sleiman, A Platinum(II) Phenylphenanthroimidazole with an Extended Side-Chain Exhibits Slow Dissociation from a c-Kit G-Quadruplex Motif, Chem. Eur. J., 19 (2013) 17836-17845.
[28] F. Moraca, J. Amato, F. Ortuso, A. Artese, B. Pagano, E. Novellino, S. Alcaro, M. Parrinello, V. Limongelli, Ligand binding to telomeric G-quadruplex DNA investigated by funnel-metadynamics simulations, Proc. Natl. Acad. Sci. USA, 114 (2017) E2136.
[29] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, The Protein Data Bank, Eur. J. Biochem., 80 (1977) 319-324.
[30] N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, C.R. Corbeil, Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go, Br. J. Pharmacol., 153 (2008) S7-S26.
[31] G.M. Morris, M. Lim-Wilby, Molecular Docking, in: A. Kukol (Ed.) Molecular Modeling of Proteins, Humana Press, Totowa, NJ, 2008, pp. 365-382.
[32] T. Wehler, R. Brenk, Structure-Based Discovery of Small Molecules Binding to RNA, in: A.L. Garner (Ed.) RNA Therap., Springer International Publishing, Cham, 2018, pp. 47-77.
[33] F. Stefaniak, E.I. Chudyk, M. Bodkin, W.K. Dawson, J.M. Bujnicki, Modeling of ribonucleic acid–ligand interactions, WIREs Comput. Mol. Sci., 5 (2015) 425-439.
[34] C. Detering, G. Varani, Validation of Automated Docking Programs for Docking and Database Screening against RNA Drug Targets, J. Med. Chem., 47 (2004) 4188-4201.
[35] G.W. McElfresh, C. Deligkaris, A vibrational entropy term for DNA docking with autodock, Comp. Biol. Chem., 74 (2018) 286-293.
[36] F. Barbault, B. Ren, J. Rebehmed, C. Teixeira, Y. Luo, O. Smila-Castro, F. Maurel, B. Fan, L. Zhang, L. Zhang, Flexible computational docking studies of new aminoglycosides targeting RNA 16S bacterial ribosome site, Eur. J. Med. Chem., 43 (2008) 1648-1656.
[37] P.T. Lang, S.R. Brozell, S. Mukherjee, E.F. Pettersen, E.C. Meng, V. Thomas, R.C. Rizzo, D.A. Case, T.L. James, I.D. Kuntz, DOCK 6: Combining techniques to model RNA–small molecule complexes, RNA, 15 (2009) 1219-1230.
[38] A. Philips, K. Milanowska, G. Łach, J.M. Bujnicki, LigandRNA: computational predictor of RNA–ligand interactions, RNA, 19 (2013) 1605-1616.
[39] S. Ruiz-Carmona, D. Alvarez-Garcia, N. Foloppe, A.B. Garmendia-Doval, S. Juhos, P. Schmidtke, X. Barril, R.E. Hubbard, S.D. Morley, PLoS Comp. Biol., 10 (2014) e1003571.
[40] Y. Li, J. Shen, X. Sun, W. Li, G. Liu, Y. Tang, Accuracy Assessment of Protein-Based Docking Programs against RNA Targets, J. Chem. Inf. Model., 50 (2010) 1134-1146.
[41] C. Guilbert, T.L. James, Docking to RNA via Root-Mean-Square-Deviation-Driven Energy Minimization with Flexible Ligands and Flexible Targets, J. Chem. Inf. Model., 48 (2008) 1257-1268
[42] S.D. Morley, M. Afshar, Validation of an empirical RNA-ligand scoring function for fast flexible docking using RiboDock®, J. Comput.-Aided Mol. Des., 18 (2004) 189-208.
[43] P. Pfeffer, H. Gohlke, DrugScoreRNA Knowledge-Based Scoring Function To Predict RNA−Ligand Interactions, J. Chem. Inf. Model., 47 (2007) 1868-1876.
[44] X. Zhao, X. Liu, Y. Wang, Z. Chen, L. Kang, H. Zhang, X. Luo, W. Zhu, K. Chen, H. Li, X. Wang, H. Jiang, An Improved PMF Scoring Function for Universally Predicting the Interactions of a Ligand with Protein, DNA, and RNA, Journal of Chemical Information and Modeling, 48 (2008) 1438-1447.
[45] L. Chen, G.A. Calin, S. Zhang, Novel Insights of Structure-Based Modeling for RNA-Targeted Drug Discovery, J. Chem. Inf. Model., 52 (2012) 2741-2753.
[46] I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge, T.E. Ferrin, A geometric approach to macromolecule-ligand interactions, J. Mol. Biol., 161 (1982) 269-288
[47] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, J. Comp. Chem., 19 (1998) 1639-1662.

[48] S.F. Sousa, P.A. Fernandes, M.J. Ramos, Protein–ligand docking: Current status and future challenges, Prot. Struct. Funct. Genet., 65 (2006) 15-26.
[49] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy, J. Med. Chem., 47 (2004) 1739-1749
[50] T.A. Halgren, R.B. Murphy, R.A. Friesner, H.S. Beard, L.L. Frye, W.T. Pollard, J.L. Banks, Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening, J. Med. Chem., 47 (2004) 1750-1759
[51] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol., 267 (1997) 727-748.
[52] M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray, R.D. Taylor, Improved protein-ligand docking using GOLD, Prot. Struct. Funct. Genet., 52 (2003) 609-623.
[53] C.R. Corbeil, P. Englebienne, N. Moitessier, Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0, J. Chem. Inf. Model., 47 (2007) 435-449
[54] N. Moitessier, J. Pottel, E. Therrien, P. Englebienne, Z. Liu, A. Tomberg, C.R. Corbeil, Medicinal Chemistry Projects Requiring Imaginative Structure-Based Drug Design Methods, Acc. Chem. Res., 49 (2016) 1646-1657.
[55] H. Gohlke, M. Hendlich, G. Klebe, Knowledge-based scoring function to predict protein-ligand interactions, J. Mol. Biol., 295 (2000) 337-356.
[56] F.S. Di Leva, E. Novellino, A. Cavalli, M. Parrinello, V. Limongelli, Mechanistic insight into ligand binding to G-quadruplex DNA, Nucl. Acids Res., 42 (2014) 5447-5455.
[57] O. Korb, T.S.G. Olsson, S.J. Bowden, R.J. Hall, M.L. Verdonk, J.W. Liebeschuetz, J.C. Cole, Potential and Limitations of Ensemble Docking, J. Chem. Inf. Model., 52 (2012) 1262-1274.
[58] E. Therrien, N. Weill, A. Tomberg, C.R. Corbeil, D. Lee, N. Moitessier, Docking Ligands into Flexible and Solvated Macromolecules. 7. Impact of Protein Flexibility and Water Molecules on Docking-Based Virtual Screening Accuracy, J. Chem. Inf. Model., 54 (2014) 3198-3210.
[59] K.E. Lind, Z. Du, K. Fujinaga, B.M. Peterlin, T.L. James, Structure-Based Computational Database Screening, In Vitro Assay, and NMR Assessment of Compounds that Target TAR RNA, Chem. Biol., 9 (2002) 185-193.
[60] N. Moitessier, E. Westhof, S. Hanessian, Docking of aminoglycosides to hydrated and flexible RNA, J. Med. Chem., 49 (2006) 1023-1033
[61] A.C. Stelzer, A.T. Frank, J.D. Kratz, M.D. Swanson, M.J. Gonzalez-Hernandez, J. Lee, I. Andricioaei, D.M. Markovitz, H.M. Al-Hashimi, Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble, Nat. Chem. Biol., 7 (2011) 553.
[62] S. Kazemi, D.M. Krüger, S. Finton, H. Gohlke, Elastic Potential Grids: Accurate and Efficient Representation of Intermolecular Interactions for Fully Flexible Docking, ChemMedChem, 4 1264-1268.
[63] D.M. Krüger, J. Bergs, S. Kazemi, H. Gohlke, Target Flexibility in RNA−Ligand Docking Modeled by Elastic Potential Grids, ACS Med. Chem. Lett., 2 (2011) 489-493.
[64] R. Rohs, I. Bloch, H. Sklenar, Z. Shakked, Molecular flexibility in ab initio drug docking to DNA: binding-site and binding-mode transitions in all-atom Monte Carlo simulations, Nucl. Acids Res., 33 (2005) 7048-7057.
[65] T. Hermann, E. Westhof, Exploration of metal ion binding sites in RNA folds by Brownian dynamics simulations, Structure, 10 (1998) 1303-1314.
[66] T. Hermann, E. Westhof, Docking of Cationic Antibiotics to Negatively Charged Pockets in RNA Folds, J. Med. Chem., 42 (1999) 1250-1261
[67] K. Xinshan, S.R. H., K.I. D., Calculation of ligand ‐ nucleic acid binding free energies with the generalized ‐ born model in DOCK, Biopolymers, 73 (2004) 192-204.
[68] S.D. Morley, M. Afshar, Validation of an empirical RNA-ligand scoring function for fast flexible docking using RiboDock®, Journal of Computer-Aided Molecular Design, 18 (2004) 189-208.
[69] P. Pfeffer, H. Gohlke, DrugScoreRNAKnowledge-Based Scoring Function To Predict RNA−Ligand Interactions, Journal of Chemical Information and Modeling, 47 (2007) 1868-1876.

[70] P.A. Holt, J.B. Chaires, J.O. Trent, Molecular docking of intercalators and groove-binders to nucleic adds with autodock and surflex, J. Chem. Inf. Model., 48 (2008) 1602-1615
[71] Y. Gilad, H. Senderowitz, Docking Studies on DNA Intercalators, J. Chem. Inf. Model., 54 (2014) 96-107.
[72] A. Soni, P. Khurana, T. Singh, B. Jayaram, A DNA intercalation methodology for an efficient prediction of ligand binding pose and energetics, Bioinf., 33 (2017) 1488-1496.
[73] D.A. Evans, S. Neidle, Virtual Screening of DNA Minor Groove Binders, J. Med. Chem., 49 (2006) 4232-4238
[74] H.K. Srivastava, M. Chourasia, D. Kumar, G.N. Sastry, Comparison of Computational Methods to Model DNA Minor Groove Binders, J. Chem. Inf. Model., 51 (2011) 558-571.
[75] Z. Yan, S. Sikri, D.L. Beveridge, A.M. Baranger, Identification of an Aminoacridine Derivative That Binds to RNA Tetraloops, J. Med. Chem., 50 (2007) 4096-4104.
[76] H.-M. Lee, D.S.-H. Chan, F. Yang, H.-Y. Lam, S.-C. Yan, C.-M. Che, D.-L. Ma, C.-H. Leung, Identification of natural product Fonsecin B as a stabilizing ligand of c-myc G-quadruplex DNA by high-throughput virtual screening, Chem. Commun., 46 (2010) 4680-4682.
[77] D.S.-H. Chan, H. Yang, M.H.-T. Kwan, Z. Cheng, P. Lee, L.-P. Bai, Z.-H. Jiang, C.-Y. Wong, W.-F. Fong, C.-H. Leung, D.-L. Ma, Structure-based optimization of FDA-approved drug methylene blue as a c-myc G-quadruplex DNA stabilizer, Biochimie, 93 (2011) 1055-1064.
[78] T. Kaserer, R. Rigo, P. Schuster, S. Alcaro, C. Sissi, D. Schuster, Optimized Virtual Screening Workflow for the Identification of Novel G-Quadruplex Ligands, J. Chem. Inf. Model., 56 (2016) 484-500.
[79] S. Alcaro, C. Musetti, S. Distinto, M. Casatti, G. Zagotto, A. Artese, L. Parrotta, F. Moraca, G. Costa, F. Ortuso, E. Maccioni, C. Sissi, Identification and Characterization of New DNA G-Quadruplex Binders Selected by a Combination of Ligand and Structure-Based Virtual Screening Approaches, J. Med. Chem., 56 (2013) 843-855.
[80] P.A. Holt, R. Buscaglia, J.O. Trent, J.B. Chaires, A Discovery Funnel for Nucleic Acid Binding Drug Candidates, Drug Dev. Res., 72 (2011) 178-186.
[81] J.-Q. Hou, S.-B. Chen, L.-P. Zan, T.-M. Ou, J.-H. Tan, L.G. Luyt, Z.-S. Huang, Identification of a selective G-quadruplex DNA binder using a multistep virtual screening approach, Chem. Commun., 51 (2015) 198-201.
[82] D.-L. Ma, D.S.-H. Chan, P. Lee, M.H.-T. Kwan, C.-H. Leung, Molecular modeling of drug–DNA interactions: Virtual screening to structure-based design, Biochimie, 93 (2011) 1252-1266.
[83] P.A. Holt, P. Ragazzon, L. Strekowski, J.B. Chaires, J.O. Trent, Discovery of novel triple helical DNA intercalators by an integrated virtual and actual screening platform, Nucl. Acids Res., 37 (2009) 1280-1287.
[84] P. Daldrop, F.E. Reyes, D.A. Robinson, C.M. Hammond, D.M. Lilley, R.T. Batey, R. Brenk, Novel Ligands for a Purine Riboswitch Discovered by RNA-Ligand Docking, Chem. Biol., 18 (2011) 324-335.