

# Molecular Modeling Applied to Nucleic Acid-Based Molecule Development

Arne Krüger <sup>1</sup>, Flávia M. Zimbres <sup>2</sup>, Thales Kronenberger <sup>3</sup> and Carsten Wrenger <sup>1,\*</sup>

<sup>1</sup> Unit for Drug Discovery, Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, SP 05508-000, Brazil; krueger.arne@icb.usp.br

<sup>2</sup> Department of Biochemistry and Molecular Biology and Center for Tropical and Emerging Global Diseases, University of Georgia, Athens, GA 30602, USA; flavia.zimbres@uga.edu

<sup>3</sup> Department of Internal Medicine VIII, University Hospital of Tübingen, 72076 Tübingen, Germany; kronenberger7@gmail.com

\* Correspondence: cwrenger@icb.usp.br; Tel.: +55-(11)-2648-8127; Fax: +55-(11)-3091-7417

Received: 31 July 2018; Accepted: 16 August 2018; Published: 27 August 2018

**Abstract:** Molecular modeling by means of docking and molecular dynamics (MD) has become an integral part of early drug discovery projects, enabling the screening and enrichment of large libraries of small molecules. In the past decades, special emphasis was drawn to nucleic acid (NA)-based molecules in the fields of therapy, diagnosis, and drug delivery. Research has increased dramatically with the advent of the SELEX (systematic evolution of ligands by exponential enrichment) technique, which results in single-stranded DNA or RNA sequences that bind with high affinity and specificity to their targets. Herein, we discuss the role and contribution of docking and MD to the development and optimization of new nucleic acid-based molecules. This review focuses on the different approaches currently available for molecular modeling applied to NA interaction with proteins. We discuss topics ranging from structure prediction to docking and MD, highlighting their main advantages and limitations and the influence of flexibility on their calculations.

**Keywords:** aptamers; docking; molecular dynamics; nucleic acids

## 1. Introduction—Nucleic Acids as a Class of Drugs and Biomarkers

In drug discovery and delivery, DNA and RNA play an important role as therapeutics, biomarkers, and ligands for targeted delivery [1]. One technique that was especially developed over the past decades and gained increasing importance is SELEX (systematic evolution of ligands by exponential enrichment). The SELEX technique was developed and described for the first time by two independent laboratories in 1990 [2,3]. Starting with a library of randomly generated oligonucleotides composed of a central variable region of random nucleotides flanked by constant sequence regions (priming regions), the entire sequences can easily be amplified via conventional PCR. The product can then be readily selected against a specific target using positive and negative selection cycles. DNA or RNA oligonucleotides, which are a result of several selection cycles, are called aptamers and can bind with high specificity and affinity to their targets. Aptamers represent interesting alternatives to antibodies and are used in a variety of applications, as already mentioned above [1]. Although antibodies and aptamers bear resemblance to each other as dissociation constants at the nano- to picomolar level, aptamers show superior advantages that should be considered. Aptamers (i) can distinguish between isomeric and conformational forms of the same protein [4,5]; (ii) are amplified in vitro without the use of animals and therefore show little batch-to-batch variation; (iii) can be produced within a short time period by low-cost high throughput systems [1]; (iv) offer a broad range of targets, including nonimmunogenic agents like toxins, and do not cause a

immunogenic response themselves [6]; (v) are smaller, which allows faster uptake and targeting of otherwise inaccessible structures; (vi) are very stable and resistant to fluctuations in temperature and pH [1]; and (vii) can be conveniently and easily modified in various ways without affecting the binding affinity [7].

Originally used in *in vitro* experiments, it has been shown that SELEX can also be applied onto more complex targets, e.g., whole cells such as erythrocytes and cancer cells (Cell-SELEX). Therefore, potential targets can be targeted at their native conformation [1,8]. Aptamers are used in a wide variety of applications, including Aptasensors, ribozymes, and DNazymes, and the interest in new oligonucleotide products has fuelled research in the field of production and process optimization [1]. Thus, the research in the field of functional nucleic acids (NAs) has become of high interest over the last decades [9]. Former examples of nucleic acid (NA)-binding proteins include glutamine tRNA synthetase (GlnRS) and T4 bacteriophage DNA polymerase (gp43), which are also targets of aptamer interaction [9]. However, in recent years, it has been shown that aptamers can also bind to proteins that are normally not involved in NA interaction. In DNA–protein interaction, binding predominantly occurs in the major groove of DNA, although minor groove interactions have also been reported [10–12]. Furthermore, interactions of aptamer and protein mainly occur due to hydrogen bonding or electropositive charge–charge interactions and, to a less extent, rely on van der Waals forces and hydrophobic groove contributions [9]. Therefore, a wide range of possible modifications, such as alterations of the base, phosphate group, or sugar ring with functional groups (such as fluoro-(F), amino-(NH<sub>2</sub>) or *O*-methyl-(OCH<sub>3</sub>) groups), could not only increase nuclease resistance [13] but also allow hydrophobic epitopes to be targeted on the surfaces of the proteins, therefore increasing the variety of aptamer targets [14]. For instance, modified NA-ligands have been constructed harboring special 5'-position-modified pyrimidine bases in DNA by attaching functional groups [15,16]. Other approaches have utilized backbone modifications of NA, such as a 1,5-anhydrohexitol nucleic acid, cyclohexenyl nucleic acid, 20-O,40-C, methylene-b-D-ribonucleic acid, arabino nucleic acid, and 20-fluoro-arabino-nucleic acid [17] (also reviewed by Meek et al. [18]). Available modified nucleotides, for instance, are altered at the 2'-position of ribose, which has been shown to influence RNA structure flexibility [19]. However, functionalized NA molecules are not limited to improve binding properties but could also influence serum stability and pharmacokinetic properties [20]. It is expected that the incorporation of click chemistry-based approaches for the development of new modified nucleotides and the use of newly developed polymerases that accept them will improve the accessibility of functionalized NAs in the market [21,22].

## 2. In Silico Approaches for Structure Prediction and Docking

### 2.1. Systematic Evolution of Ligands by Exponential Enrichment—Alternative Algorithms

Despite the successful cases, SELEX can still be time- and resource-consuming and the success varies, depending—among other factors—on the quality and composition of the applied oligonucleotide library [23]. Therefore, several different approaches have been developed to simplify the whole selection process. Most notable of these are *in silico* techniques that allow for aptamer design and selection completely from scratch [24,25]. These include the computational design of an aptamer library [26,27] as well as a selection of aptamers against a given target [27,28]. The general problems with these approaches are the lack of random libraries and the necessary calculation power, which is disproportionately high [23,27,28]. Another interesting approach was published in 2016 by Ahirwar and colleagues who tried to design an aptamer to target estrogen receptor alpha (ER $\alpha$ ) *in silico*. Instead of generating a pool of random sequences, they deployed a “bottom-up” method utilizing a natural binding ligand of ER $\alpha$ —estrogen response element—to create aptamers and optimized the binding affinities [29].

An interesting alternative approach was recently presented by a group from Heidelberg trying to circumvent the main flaws of aptamer design and development using experimental and *in silico* approaches. In 2015, the group participated in the International Genetically Engineered Machine (iGEM) competition and developed software named M.A.W.S., short for Making Aptamers Without

SELEX. The algorithm loads the target molecule and creates a bounding box. For each nucleotide, the best conformation is calculated and the position with the lowest entropy is selected as the starting point for the next calculation cycle. Thereby, the algorithm constructs aptamers from scratch without using a given aptamer library against any given target structure [23].

## 2.2. Aptamer Secondary and Tertiary Structure Prediction

The binding affinity and specificity of aptamers derive from their specific secondary and tertiary structures, which allow for the recognition of different target structures [30]. The viability of aptamer application is based on the high flexibility of the (deoxy)ribose-phosphodiester backbone, which possesses a total of six torsion angles/flexible bonds and therefore allows for a wide variety of secondary and tertiary structures [9]. On the other hand, structures are narrowed due to the limitation of only four bases. Therefore, the already mentioned modifications were introduced to broaden the range of structures and NA–protein interactions [16,31].

Nucleic acid modeling must consider the flexibility of the phosphodiester backbone and all possible base pairings, including noncanonical base pairing as well as the influence of hydrophobic interactions and best free energy conformations [32]. These factors pose a great challenge in terms of algorithm approach and, later, computing power. Several algorithms exist to calculate the secondary structures and recognize motifs of short to medium ranged DNA and RNA molecules that apply different approaches, such as MEME Suite [33], MEMERIS [34], GLAM2 [35], mFold [36], and Aptamotif [37], etc. Secondary structures occur as a result of intramolecular nucleotide pairing and are the reason for target–ligand interactions [36]. Among pseudoknots and G-quadruplex, the most common structures are stem-loops, which comprise four different substructures: (i) hairpin loop, (ii) bulge loop, (iii) interior loop, and (iv) multibranch loop that again form more complex structures such as kissing hairpins [1,37]. While some algorithms are physic-based and employ thermodynamics (such as the Turner’s thermodynamics table) or the nearest-neighbor model, e.g., mFold, and only consider canonical base-pairings [36,38,39], others compare free-energy values to databases and also include noncanonical base-pairing and are therefore considered knowledge-based algorithms (such as MC-Fold|MC-Sym) [40]. Noncanonical base pairs can be formed by hydrogen bonds or stabilized by polar hydrogen bonds and even through interactions between C-H and O or N groups [41] and contribute interfering energies to the secondary structure prediction [40].

While secondary structures can be predicted based on the nucleotide sequence, the tertiary structures are far more complex. The aptamer shape is directly responsible for the binding affinity and specificity, and its 3D structure is collectively coded by information from the secondary structure [1]. Several algorithms for tertiary structure prediction of short NA molecules have been reported, but only a few of them are still accessible. In general, two main approaches exist for the prediction of tertiary structures: (i) prediction based on sequence analysis and (ii) prediction based on homologous sequence structures from databases. However, both approaches have disadvantages, which include the bad prediction of base pairing and pseudoknot formations and limited availability of known tertiary structures in databases, respectively. Furthermore, prediction of DNA and RNA is slightly different due to the additional oxygen atom in the RNA sugar, which allows for more hydrogen bond formation as well as the tendency of DNA to form double helix structures [42].

Nevertheless, it has to be noted that a group of researchers recently provided evidence that structural conversion between DNA and RNA molecules *in silico* resulted in almost identical hairpin formations and small fragment structures compared to experimentally gathered data [43]. Since research has far more interest in RNA molecules, significantly more prediction algorithms exist for RNA calculation. These include web-based algorithms, such as RNAComposer [44,45], BARNACLE [46], FARNAL [47], MC-Fold|MC-Sym pipeline [40], and many others.

## 3. Historical Overview of Docking Algorithm Development

Interactions between aptamer and target are primarily based on polar and ionic interactions (extensively revised in references [9,14,48], and discussed in detail below), in addition to shape complementarity that results in binding properties comparable to monoclonal antibodies [6].

Predictions of these interactions are highly complex due to the flexibility and size of NAs as well as the already-mentioned chemical modifications, which lack experimentally determined structures as references in the databases. As an alternative, several different programs exist to tackle the problems as described in the following sections (Table 1).

While research focused on protein–protein interaction early on, computational approaches for protein–NA prediction and calculation have lagged behind [49,50]. With the discovery of ribozymes—small RNA molecules that exhibit enzymatic function—and post-transcriptionally regulated gene expression, interest in the field of RNA biology has fomented and grown exponentially [51–53]. Docking programs and algorithms have been designed to calculate protein–protein interactions and have been modified to also allow RNA sequences as ligands by implementing different parameters for the scoring functions without changing the core of their algorithms [54]. Docking algorithms fall into two main categories: (i) the machine learning algorithms, which predict molecule interactions based on sequence-based and/or structure-based information; and (ii) the template-based algorithms, which calculate interactions with information from known crystal structures [54].

First reports on the topic came from Katchalski-Katzir and colleagues, which employed rigid docking mode in combination with Fourier transformation to evaluate possible interaction sites in a six-dimensional-shape complementarity approach, termed GRAMM [55]. Rigid docking involves the putative binding of the three-dimensional NA on the receptor protein by considering the NA as a single immobile entity and only changing the overall coordinates by rotational and translational transformations. While this method was solely based on shape, newer algorithms, such as FTDock, have already implemented electrostatics and biochemical information, although they are still limited to rigid docking [56,57]. While the FTDock was designed for protein–protein interaction, nucleic acids can also be submitted. Moreover, the program has been further developed using a scoring mechanism and an algorithm for energy calculations, side chain optimization, and backbone refinement. Together with FTDock, they are described as 3D-Dock, an optimized docking program [58].

In parallel to these early approaches, the use of spherical polar Fourier correlation method, implemented in Hex docking, instead of using the Fast Fourier Transformation (FFT), has been shown to drastically accelerate the calculation speed [59]. Two additional convolution algorithms—DOT and DOT2—have also implemented Poisson–Boltzmann methods to better evaluate highly polar intermolecular interactions [60,61].

**Table 1.** A chronological overview of docking algorithms according to the mode of action.

Algorithm	Milestone	Reference
GRAMM	Rigid docking, six-dimensional shape complementarity; fast Fourier transformation	[55]
FTDock	Implementation of electrostatics and biochemical information	[56,57]
3D-Dock	Additionally, energy calculations, side chain optimization, and backbone refinement	[58]
Hex	Spherical polar Fourier correlation method	[59]
Dot/Dot2	Implementation of Poisson–Boltzmann methods	[60,61]
HADDOCK	Flexibility of amino acid side chains	[62]
PatchDock	Local feature matching instead of six-dimensional transformation fitting	[63]
ParaDock	Shape complementarity but flexible NA structure prediction	[64]
NPDock	Rigid body docking while considering the specific features of NA	[53]
HDOCK	Docking between two big molecules; template-based and template-free rigid docking mode	[65,66]
Gold	Full flexibility or rotamer-based search for both ligand and selected amino acids residues; docking in a determined binding pocket. Presents a range of different scoring functions, from machine-learning-based to physicochemical-based ones	[67]
Autodock Autodock Vina	Full flexibility or rotamer-based search for both ligand and selected amino acids residues; docking in a determined binding pocket. Energy-based scoring function and ability to handle surface pockets	[68]

The program HADDOCK represented the first attempt to include not just the interaction information, but also allow flexible amino acid side chains, thereby increasing the quality of the docking models. Initially conceptualized for protein–protein interaction, HADDOCK nowadays also allows NA as input molecules [62]. HADDOCK’s algorithm has been refined to encompass the main challenges of protein–DNA docking, namely, the tendency of DNA to form double helix structures, together with the identification of DNA interaction sites [42]. While DNA often interacts with proteins via its major groove, further interaction with the minor groove must also be considered as the bigger DNA molecules represent one helical turn, which increases the possible binding sites [10]. PatchDock was developed with the aim of overcoming some of the calculation hurdles of existing shape complementarity docking algorithms. Although also geometry-based, the program calculates interaction sites with a higher efficiency using local feature matching instead of the usually applied six-dimensional transformation fitting [63].

While protein docking programs were optimized to also fit NA as input structures, first attempts for algorithms specially designed for NA–protein docking were reported around 2011. ParaDock, developed by Banitt and Wolfson, is an *ab initio* approach to calculate interactions solely based on the sequence of DNA and a rigid protein structure. While ParaDock still utilizes shape complementarity, the DNA structure is calculated from scratch in a flexible mode [64]. This is in contrast to most of the aforementioned software, which need the interface distant constraints as input [57,63,64].

The Bujnicki lab provided knowledge-based potentials for download to allow for protein–RNA docking in the same year. The potentials—named DARS-RNP and QUASI-RNP—use Decoys and quasi-chemical methods to describe the reference state and allow RNA and DNA docking, respectively [69]. Shortly afterwards, the same group presented a program called NPDock (short for nucleic acid–protein docking) which is based on the GRAMM algorithm implementing the DARS- and QUASI-RNP potentials and only allows rigid body docking, although considering the specific features of nucleic acids [53]. HDOCK, which was only recently released as a new algorithm focusing on “big molecule–big molecule” docking, operates on both template-based or template-free rigid docking mode, which also allows docking of molecules with unknown structures [65,66].

In contrast to rigid docking, flexible docking programs allow various conformations of the target molecules to find the state closest to the native form. Programs such as Gold [67], Autodock and Autodock Vina [68] had great success in small molecule drug discovery and have been employed to small and large NA docking. As a common characteristic, they all allow either full flexibility or rotamer-based search for both ligand and selected amino acids residues for docking in a determined binding pocket [67,68].

Although one might initially consider flexible docking as the best option, it should be kept in mind that calculation time exponentially grows with the increasing size of a molecule. Since aptamers are often initially composed of up to 70 nucleotides or even more, tertiary structures hold too much variability for flexible docking algorithms based on genetic (GA) or Lamarckian approaches to deal without high-performance computation (HPC) infrastructure. In this sense, binding site information from known structures is normally included in the docking procedure [65]. In cases where no crystal structure information on the protein complexes are available, NA binding site predictions can be carried out externally using other programs [70]. However, single strand DNA (ssDNA)-binding prediction algorithms are rarely available, which leads to the use of RNA-based counterparts as the most used option and results in another layer of uncertainty. Still, it must be mentioned that the atomic composition of RNA and DNA is highly similar and hence the usage of an RNA prediction algorithm for DNA binding sites have been shown [36,71].

#### 4. Benchmarking and Quality Tests

Several studies report the use of the previously discussed algorithms in practical approaches to either predict aptamer binding to target proteins or to evaluate the quality of the docking calculations. In the following paragraphs, recent applications are presented and benchmark studies are emphasized.

In 2016, Torabi and colleagues utilized the HADDOCK program to predict the binding mode of retinol binding protein 4 (RBP4) and RBP4 binding aptamer (RBA) for further MD simulations. RBP4 is a biomarker used for type two diabetes pre-diagnosis, and the binding mode prediction with subsequent MD simulations was intended to elucidate the main aptamer–target interaction mechanics to support further aptamer design [72]. The group could discover the main forces driving the interaction between RBP4–RBA, shedding light on the mode of action, which highlighted this molecule’s inhibitory potential.

In a different study, HADDOCK, Autodock Vina, and Patchdock were applied as *in silico* controls to evaluate the binding affinity and specificity of aptamers against ER $\alpha$ . The aptamers were previously *ab initio* and *in silico*, designed using EREs as a template (see above, [29]). After validating the docking settings on a set of known protein–NA complexes against randomly generated RNA structures, the best docking protocol was used to predict aptamer–target interaction and select the best aptamer to bind ER $\alpha$ . The group could not only show that all docking algorithms concluded similar results but could also prove that the *in silico* predicted specificity and affinity *in vitro* [29]. Additional studies were designed especially to test existing docking programs, including the community-wide Critical Assessment of Prediction of Interactions (CAPRI) initiative [73,74] and a wide variety of benchmark databases.

For instance, Roberts and colleagues put their own DOT and DOT2 algorithm to the test using a set of four different protein–NA complexes [50]. They evaluated the prediction accuracy without any experimental knowledge compared to the structural data gathered from experimental procedures [50]. Similar studies were carried out for the HADDOCK algorithm, including several benchmark databases for protein–NA complexes [42]. Even more extensive efforts were made to benchmark the performance of the HDock algorithm on several known benchmark databases [65]. Additionally, many databases were created to offer opportunities for easy benchmarking of established docking programs and algorithms [49,75–77].

## 5. Evaluating the Quality of Docking

When evaluating the quality of the docking, some main ideas should be taken into consideration. The driving forces for NA–protein interaction are a milieu of van der Waals forces, hydrophobic interactions, hydrogen bonds, base stacking forces, and ionic interactions between amino acid side chains and either the phosphate groups or bases of NA [10,24]. Additionally, the importance of the shape complementarity provided by secondary and tertiary structures of NA and proteins, which contributes to the binding mode, should be stressed [24].

Ionic interactions commonly formed between positive amino acid side chains and the negatively charged DNA have been repeatedly proven important for NA–protein interaction [24,78–81]. Base stacking forces contribute to the stability of dsDNA but in particular support the binding of ssDNA to proteins involving stacking of bases and aromatic protein side chains [82]. At the same time, they are highly influenced by electrostatic interactions and van der Waals forces [83]. Hydrogen bonds, driven by dipole–dipole interactions, are especially formed more easily, and their energy can range between 4 and 40 kJ/mol [24,84,85]. Although their energy depends on pressure, angle, the distance between donor and acceptor (of at least 2.5 Å), and environment, the formation is not influenced much by surface structure [84,86]. The number of hydrogen bonds can be counted for each docking complex to measure the quality of the docking pose and the interaction itself, as already mentioned by Jones and colleagues [67] and discussed by Ahirwar et al. in 2016 [29]. Formation of hydrogen bonds has been demonstrated to be essential for biomolecular function and, hence, structure represents a key parameter of complex stability [87,88]. Despite the importance of  $\pi$ – $\pi$  interaction, cation– $\pi$ , and the overall electrostatic contribution, comparatively few studies have used other interactions as criteria for selection of NA docking poses [28,89]. As an example, Rabal employed clustering of docking results followed by evaluation of electrostatic and polar interactions between protein and RNA aptamers, similarly to ligand interaction fingerprints already employed as docking postprocessing of protein–ligand complexes [89].

Alternatives, such as consensus scoring, have been proposed to reduce the bias of single scoring functions [90], and the evaluation of true-positives retrieval rate from different programs can help. However, the notion that the highest dock score directly correlates with real ligand binding—and therefore with a biological effect—can be erroneous. Especially when applied to small molecules, docking analyses alone can create an inaccurate picture of ligand binding (an extensive discussion on the docking limitations was addressed by Chen [91]).

The knowledge of critical residues and, in this sense, the presence of respective interactions can bridge the virtual inferences and experimental results. In this sense, molecular modeling can give the structural perspective of the mutation effects while also benefitting from the experimental information.

Optimal prediction of native macromolecule conformations remains a challenge, especially since most of the docking approaches currently in use are based on a single rigid conformation. Despite the prolonged effort on the development of new docking algorithms, the ability to capture a full motion of the interaction between NAs and proteins is out of the scope of the technique, and more accurate predictions can be drawn from an integrated pipeline.

## 6. Molecular Dynamics Simulations Applied to Nucleic Acids (NA) and Protein–NA Interactions

Computational techniques can help to interpret protein–NA interactions and complement experimental results. To understand any protein–ligand interaction, the time dimension must be added to the snapshots of proteins frozen in crystal structures and docking poses. Molecular dynamics simulations can describe protein dynamics in detail, including the precise position of each atom at any instant in the simulation time—along with the corresponding energies—provided that at least one structure is known as a starting point. Briefly, molecular dynamics starts from the static structures experimentally determined, which represents the atom coordinates of macromolecules. These molecules are immersed in the solvent and have their positions updated along the simulation according to classical mechanic calculations of their interactions among themselves and with the solvent. The classical mechanic facet is represented by empirical force fields with optimized parameters for biological molecules. Furthermore, quantitative analysis of the conformational ensembles of the molecules during the long-enough simulations can reveal the thermodynamic properties of the biological system [92].

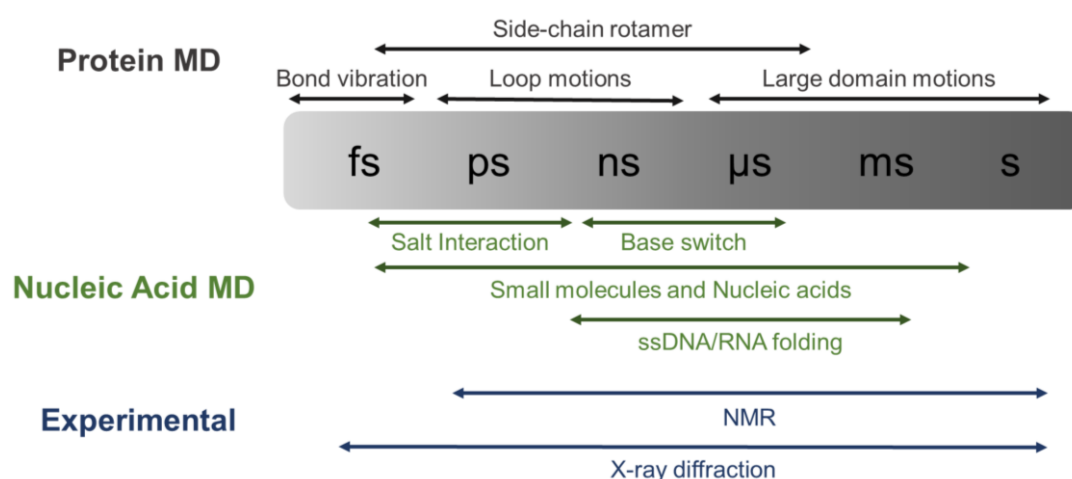
Simulation of nucleic acids and its interactions is an ever-growing field, which has lagged in comparison with the globular protein simulations. While the first protein simulations started in the early 1960s, the first simulation with nucleic acids dates from 1983 [93,94]. The reasons for this, as previously discussed by Mackerell [95], are the lack of NA–protein experimental complexes for validation of the studies, which also relied on inadequate models to treat electrostatic interactions and solvent.

Initial molecular dynamics simulation relied on the classical single point charge Coulombic model to describe electrostatic interactions, where the number of interaction partners was determined by a maximum distance cutoff. The exchange in the paradigm for electrostatic treatment—from the low cutoff Coulombic interaction model towards the Particle mesh Ewald method (PME)—was a milestone to NA simulations, since they are highly charged and nonglobular molecules. PME enabled the fast calculation of long-range interactions and its grid approach, instead of the classical static spherical cutoff, and improved the accuracy of NA simulation [96]. Additionally, modern HPC facilities enabled the use of explicit solvent representation as a standard option for NA's simulations despite the higher computational cost compared to the implicit counterpart. HPC facilities, which initially relied on the algorithms Central Processing Unit CPU-parallelization, are now focused on the use of GPU-distributed calculations.

In terms of available force fields, specific torsions and bond's parameters for protein–NA simulation were originally implemented on AMBER [97]. These parameters have been refined along the years through updates (AMBER ff94 up to ff99, and more recently the bsc1 or OL15 corrections [98]), besides classical force fields such as CHARMM, which have improved DNA representation parameters in the most recent version (CHARMM36) [99,100]. Most of the parameter corrections

along the years focused on better representing NA-specific chemical features, which were poorly represented by other force fields, namely, anionic sugar–phosphate backbone [101] and the glycosidic dihedral chi angle for RNA and DNA [102]. However, a detailed discussion of the modified parameters is out of the scope of this review. While dsDNA counts on the stacking for the characteristic structural stability, as shown by long timescale simulations [99], the single-stranded NAs have a higher flexibility assuming multiple conformations. Molecular dynamics simulations can investigate the ssDNA distortion upon the protein binding considering specific NA interactions that arise from the free nucleotide bases [82]. Fast timescale simulations can define fluctuations around a defined state, representing properties such as interactions stability, amino acid flipping, or conformational change of small loops (Figure 1). Accordingly, Protein–double stranded (ds)NA complexes remain stable through an intricate network of hydrogen bonds and ionic interactions [103,104].

Molecular dynamic simulation of transcription factors on a microsecond scale showed that charged amino acids (side chains with  $\text{NH}_3^+$ ) were responsible for this interaction. On the one hand, intermittent arginine–phosphate salt bridges with lifetimes of the order of hundreds of picoseconds [105,106], which is in line with NMR studies [107], were relevant for general dsDNA binding (Figure 1). The importance of explicit solvents to represent those interactions cannot be understated, since Arg–Phosphate indirect interactions, water-mediated, are as relevant as the direct bridge [108]. On the other hand, complementarily, interactions between arginine and NA bases helped to elucidate the differential binding to the consensus sequences.



**Figure 1.** Timescale defines what can be observed by molecular dynamics simulation. The timescale of dynamic processes in proteins (Black), protein–nucleic acid complexes (Green), and the experimental methods (Blue) that can be observed from the different methods.

Short simulations of the *Corynebacterium pseudotuberculosis* cold shock protein A bound to an aptamer molecule were performed by Caruso [109]. They have shown that ssDNA can also rely on intermittent nonspecific hydrogen bond interactions, in addition to the salt-bridges, to confer complex stability [109]. Interactions between aromatic amino acids and the free nucleic acid bases contributed to an overall enthalpy energy gain. Protein–aptamer complexes simulation on a short scale of few nanoseconds were able to suggest the stability of the aptamer secondary structure; however, induced-fit effects would require longer simulations and extensive sampling [72,110,111].

It should be noted, however, that protein dynamics are characterized not only by the timescale of the atomic fluctuations (short simulations of few nanoseconds) but also by the amplitude and the directionality of the fluctuations [112]. Long range conformational transitions, despite rarely occurring in short MDs, are relevant because many biological processes—including protein–protein and protein–NA interactions—occur on the timescale of microseconds. Although it is possible to routinely perform microsecond-length simulations of fully solvated atomistic nucleic acids with



reliable convergence [113–115], unfortunately, protein–NA dynamics on the microsecond-to-millisecond timescale cannot routinely be performed in most laboratories.

New approaches that either simplify force fields or the conformational sampling have been developed along the years to overcome these limitations including, for instance, normal mode analysis (NMA) [116,117] and Markov state models (MSM) [118]. NMA can describe slow large-amplitude motions of the proteins, sometimes restrained to the C $\alpha$  atoms only, using calculated low frequencies of the vibrational normal modes. Due to the simplified nature of NMA calculations relying on the intrinsic harmonic oscillations of protein, this method is limited to investigating fluctuations around a specific conformation. In this sense, the transcription factor Catabolite Activator Protein (CAP) has been showcased as an example of protein–NA complex treated by NMA [119]. Despite the clear influence of the DNA in the different simulated systems, by comparing NMA calculated from structures cocrystallized with and without the nucleic acid, the DNA was not explicitly considered in the NMA calculations.

On the other hand, MSM relies on comprehensive sampling of simulated trajectories to generate a network model. This network can allocate the conformational space into discrete states and suggest the kinetic aspect from the transition probabilities between the states. MSM has the advantage of statistically approaching the diversity of conformations and—even when the amount of collected data is poor—MSM can guide data collection, e.g., by selecting new MD simulation's starting points.

For instance, thrombin and the thrombin-binding aptamer (TBA), with the sequence 5'-GGTTGGTGTGGTTGG-3'), which is under investigation as an anticoagulant drug, have been employed as models for better understanding the effect of long simulations on both the aptamer folding/unfolding process and interaction with the protein [120]. Crystal structures of this complex reveal TBA-binding in the fibrinogen binding site (also called exosite-I), preventing thrombin cleavage by fibrinogen. Long simulations of thrombin, complexed with TBA, showed a reduced number of protein conformation states towards a single population with reduced flexibility in the surface loops. Complementarily, simulations of thrombin without aptamer-binding had a larger number of unique conformations, which became inaccessible after aptamer-binding [120]. Also, thrombin has a second RNA-addressable site—the exosite II—which can be exploited to inhibit the thrombin-dependent platelet activation [121]. Inhibition of both sites can decrease procoagulant activity synergistically, which suggests an allosteric mechanism.

In conclusion, both NMA and Markov models have had great success in reproducing the conformational diversity of protein and NA states individually; however, they must still achieve the popularity and precision that MD currently have, especially concerning Protein–ssNA applications. MD of free TBA followed by the inference of MSM revealed an unfolding process with interconnected multistate intermediates, which can provide important insights into the diverse conformational set of this important aptamer [122].

In terms of force field simplification, simulations employing coarse-grained models were employed for studying large-scale conformational changes in free RNA molecules and riboswitches [123–125]. Coarse-grained models simplify groups of individual atoms as a single entity, encompassing the main properties of the parts. By decreasing the number of simulated elements and therefore the degrees of freedom, there is a boost in simulation performance in detriment to the detailed system description. Currently, there is a range of tools involved in the single stranded RNA (ssRNA) structure prediction and validation [126]; however, a valid prediction pipeline for ssDNA has recently been developed [43]. Jeddi and Saiz employed an approach to integrating the secondary structure information of ssDNA into the 3D ssRNA modeling software available, followed by the conversion of the final 3D structure into its ssDNA equivalent [43]. Additionally, they showed that atomistic MD simulations could be used to improve the correlation between the predicted and native structures, mainly on hairpin-like structural motifs. Liu and collaborators investigated the stability of small molecules ligand binding pose by MD [127,128]. The proposed binding mode encompassed real ligands and decoys both self-docked in the original structure and cross-docked into homologs and homology models.

In general, simulations should reproduce known experimental data or complement its interpretation; however, on one hand, there is no experimental equivalent for the MD ability to capture protein motions in so many different scales and on the other hand, the reproducibility is a recurrent issue (these issues have been extensively discussed by [129]). NMR spectroscopy is a technique capable of exploring some of the protein motions. Specifically, NMR experiments such as nuclear spin relaxation and relaxation dispersion enable the evaluation of motions at pico–nanoseconds and micro–milliseconds timescale, respectively [130]. Accordingly, given that enough details are provided, MD simulation, as well as any experimental procedure, should be reproducible; however, documentation on important details, such as force field version and parameters adjustment, are often lacking. Additionally, there are several technical limitations to the use of MDs which will be described in the following section. Molecular dynamics are stochastic in nature and highly dependent on the initial state since the simulation time often is insufficient to overcome potential energy barriers and sample diverse conformation changes. In this case, simulation represents the atomic changes within the microenvironment around the initial state and its respective thermodynamic ensemble. Solvent representation in simulations is important not just for the overall protein folding but also, as previously mentioned, to model the electrostatic interactions since most of the salt bridge interactions are mediated by water networks. However, often solvent and ion numbers are reduced to save computational power, which decreases the ability to accurately represent the water shell around the complexes [131].

Finally, although experiments can determine what is moving and how fast, molecular dynamics simulations can answer why things move because the underlying forces and corresponding energies are included in the simulation. The resulting predictions can inspire new experiments trying to answer how the nucleic acids interact with the proteins and therefore, despite the aforementioned limitations, MD can still be employed as an important tool as long as timescale and parameter are considered.

## 7. Conclusions and Future Perspectives

Cheatham highlighted the importance of sampling for understanding nucleic acids properties, particularly when the binding energy estimation is concerned [132]. Herein, we presented the importance of nucleic acids in recent biological and biomedical research and gave an introduction to discovery and the development of functional NAs, including SELEX and alternative *in silico* approaches. We further gave an overview of *in silico* modeling approaches of NAs and proteins and especially discussed the importance of appropriated scale simulations to evaluate the stability between protein and large nucleic acids (mostly aptamers) in the complex. We also highlighted the importance of post-processing, such as evaluation of electrostatic and hydrogen bond interactions after docking. However, even after MD simulations, a molecular docking result can just show that the nucleic acid binds well to the target protein. Therefore, it is not advisable to overinterpret the docking results before other experimental validations have been performed. In terms of docking itself, new alternative algorithms are likely to be developed in the near future, together with machine learning techniques, which can improve the tridimensional structure prediction for the aptamers, therefore filling the current structural gap.

The current state of the art shows that long simulations with dsDNA and even protein–dsDNA are possible. However, simulations of ssNA remain a challenge, since the importance of base-stacking interactions versus exposure to explicit solvent is still unbalanced, which could be solved with new parameters for the current force fields. Additionally, several factors that can influence the kinetics of protein–nucleic acid interactions, such as viscosity and pH of a solution, and changes on the protonation state of amino acids are out of the scope of classical molecular dynamics. Recently, the use of polarizable force fields on simulations have been discussed as the next generation step for molecular dynamics simulations, which would not only address the mentioned limitations but also better represent base stacking [133]. Finally, the ever-growing computational power, which initially benefited from parallelization, has nowadays culminated in the advanced GPU high-performance

centers. We expect that large-scale simulations for protein–NA complexes will shed a light on the interactions and can become an integral part of this field of research.

**Author Contributions:** Conceptualization, A.K., F.M.Z., C.W., and T.K.; Writing—Original Draft Preparation, A.K., F.M.Z., C.W., and T.K.; Writing—Review & Editing, A.K., F.M.Z., C.W., and T.K.; Supervision, C.W. and T.K.; Project Administration, C.W. and T.K.; Funding Acquisition, C.W.

**Funding:** This research was funded by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), grant numbers 2018/08820-0, 2017/03966-4 and 2015/26722-8.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Zhou, J.; Rossi, J. Aptamers as targeted therapeutics: Current potential and challenges. *Nat. Rev. Drug Discov.* **2017**, *16*, 181–202, doi:10.1038/nrd.2016.199.
2. Ellington, A.D.; Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990**, *346*, doi:10.1038/346818a0.
3. Tuerk, C.; Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990**, *249*, 505–510, doi:10.1126/science.2200121.
4. Hess, G.P.; Ulrich, H.; Breiting, H.-G.; Niu, L.; Gameiro, A.M.; Grever, C.; Srivastava, S.; Ippolito, J.E.; Lee, S.M.; Jayaraman, V.; et al. Mechanism-based discovery of ligands that counteract inhibition of the nicotinic acetylcholine receptor by cocaine and MK-801. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 13895–13900, doi:10.1073/pnas.240459497.
5. Faria, M.; Ulrich, H. The use of synthetic oligonucleotides as protein inhibitors and anticancer drugs in cancer therapy: Accomplishments and limitations. *Curr. Cancer Drug Targets* **2002**, *2*, 355–368.
6. Ulrich, H.; Wrenger, C. Disease-specific biomarker discovery by aptamers. *Cytom. Part A* **2009**, *75A*, 727–733, doi:10.1002/cyto.a.20766.
7. Sefah, K.; Shangguan, D.; Xiong, X.; O'Donoghue, M.B.; Tan, W. Development of DNA aptamers using Cell-SELEX. *Nat. Protoc.* **2010**, *5*, 1169–1185, doi:10.1038/nprot.2010.66.
8. Morris, K.N.; Jensen, K.B.; Julin, C.M.; Weil, M.; Gold, L. High affinity ligands from in vitro selection: Complex targets. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 2902–2907.
9. Gelinas, A.D.; Davies, D.R.; Janjic, N. Embracing proteins: Structural themes in aptamer-protein complexes. *Curr. Opin. Struct. Biol.* **2016**, *36*, 122–132, doi:10.1016/j.sbi.2016.01.009.
10. Pabo, C.O.; Sauer, R.T. Protein-DNA recognition. *Annu. Rev. Biochem.* **1984**, *53*, 293–321, doi:10.1146/annurev.bi.53.070184.001453.
11. Harris, L.-A.; Williams, L.D.; Koudelka, G.B. Specific minor groove solvation is a crucial determinant of DNA binding site recognition. *Nucleic Acids Res.* **2014**, *42*, 14053–14059, doi:10.1093/nar/gku1259.
12. Mendieta, J.; Pérez-Lago, L.; Salas, M.; Camacho, A. Functional specificity of a protein-DNA complex mediated by two arginines bound to the minor groove. *J. Bacteriol.* **2012**, *194*, 4727–4735, doi:10.1128/JB.00677-12.
13. Ulrich, H.; Martins, A.H.B.; Pesquero, J.B. RNA and DNA aptamers in cytomics analysis. *Cytom. Part A* **2004**, *59*, 220–231, doi:10.1002/cyto.a.20056.
14. Davies, D.R.; Gelinas, A.D.; Zhang, C.; Rohloff, J.C.; Carter, J.D.; O'Connell, D.; Waugh, S.M.; Wolk, S.K.; Mayfield, W.S.; Burgin, A.B.; et al. Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 19971–19976, doi:10.1073/pnas.1213933109.
15. Vaught, J.D.; Bock, C.; Carter, J.; Fitzwater, T.; Otis, M.; Schneider, D.; Rolando, J.; Waugh, S.; Wilcox, S.K.; Eaton, B.E. Expanding the chemistry of DNA for in vitro selection. *J. Am. Chem. Soc.* **2010**, *132*, 4141–4151, doi:10.1021/ja908035g.
16. Rohloff, J.C.; Gelinas, A.D.; Jarvis, T.C.; Ochsner, U.A.; Schneider, D.J.; Gold, L.; Janjic, N. Nucleic acid ligands with protein-like side chains: Modified aptamers and their use as diagnostic and therapeutic agents. *Mol. Ther. Nucleic Acids* **2014**, *3*, doi:10.1038/mtna.2014.49.
17. Pinheiro, V.B.; Taylor, A.I.; Cozens, C.; Abramov, M.; Renders, M.; Zhang, S.; Chaput, J.C.; Wengel, J.; Peak-Chew, S.-Y.; McLaughlin, S.H.; et al. Synthetic genetic polymers capable of heredity and evolution. *Science* **2012**, *336*, 341–344, doi:10.1126/science.1217622.

18. Meek, K.N.; Rangel, A.E.; Heemstra, J.M. Enhancing aptamer function and stability via in vitro selection using modified nucleic acids. *Methods* **2016**, *106*, 29–36, doi:10.1016/j.ymeth.2016.03.008.
19. Masaki, Y.; Miyasaka, R.; Ohkubo, A.; Seio, K.; Sekine, M. Linear relationship between deformability and thermal stability of 2'-O-Modified RNA hetero duplexes. *J. Phys. Chem. B* **2010**, *114*, 2517–2524, doi:10.1021/jp909851j.
20. Yoon, S.; Rossi, J.J. Future strategies for the discovery of therapeutic aptamers. *Expert Opin. Drug Discov.* **2017**, *12*, 317–319, doi:10.1080/17460441.2017.1290077.
21. Tolle, F.; Brändle, G.M.; Matzner, D.; Mayer, G. A versatile approach towards nucleobase-modified aptamers. *Angew. Chem. Int. Ed.* **2015**, *54*, 10971–10974, doi:10.1002/anie.201503652.
22. Chen, T.; Hongdilokkul, N.; Liu, Z.; Adhikary, R.; Tsuen, S.S.; Romesberg, F.E. Evolution of thermophilic DNA polymerases for the recognition and amplification of C2'-modified DNA. *Nat. Chem.* **2016**, *8*, 556–562, doi:10.1038/nchem.2493.
23. Team:Heidelberg/Software/Maws-2015.igem.org. Available online: <http://2015.igem.org/Team:Heidelberg/software/maws> (accessed on 23 May 2018).
24. Hu, W.-P.; Lin, H.-T.; Tsai, J.J.P.; Chen, W.-Y. Investigating interactions between proteins and nucleic acids by computational approaches. In *Computational Methods with Applications in Bioinformatics Analysis*; Advanced Series in Electrical and Computer Engineering; World Scientific: Singapore, 2017; Volume 20, pp. 98–117. ISBN 978-981-320-797-4.
25. Gong, S.; Wang, Y.; Wang, Z.; Zhang, W. Computational methods for modeling aptamers and designing riboswitches. *Int. J. Mol. Sci.* **2017**, *18*, doi:10.3390/ijms18112442.
26. Luo, X.; McKeague, M.; Pitre, S.; Dumontier, M.; Green, J.; Golshani, A.; Derosa, M.C.; Dehne, F. Computational approaches toward the design of pools for the in vitro selection of complex aptamers. *RNA* **2010**, *16*, 2252–2262, doi:10.1261/rna.2102210.
27. Chushak, Y.; Stone, M.O. In silico selection of RNA aptamers. *Nucleic Acids Res.* **2009**, *37*, doi:10.1093/nar/gkp408.
28. Hu, W.-P.; Kumar, J.V.; Huang, C.-J.; Chen, W.-Y. Computational selection of RNA aptamer against angiopoietin-2 and experimental evaluation. *BioMed Res. Int.* **2015**, doi:10.1155/2015/658712.
29. Ahirwar, R.; Nahar, S.; Aggarwal, S.; Ramachandran, S.; Maiti, S.; Nahar, P. In silico selection of an aptamer to estrogen receptor alpha using computational docking employing estrogen response elements as aptamer-alike molecules. *Sci. Rep.* **2016**, *6*, 21285, doi:10.1038/srep21285.
30. Cataldo, R.; Ciriaco, F.; Alfinito, E. A validation strategy for in silico generated aptamers. *arXiv* **2017**, arXiv:1711.07397.
31. Jarvis, T.C.; Davies, D.R.; Hisaminato, A.; Resnicow, D.I.; Gupta, S.; Waugh, S.M.; Nagabukuro, A.; Wadatsu, T.; Hishigaki, H.; Gawande, B.; et al. Non-helical DNA triplex forms a unique aptamer scaffold for high affinity recognition of nerve growth factor. *Structure* **2015**, *23*, 1293–1304, doi:10.1016/j.str.2015.03.027.
32. Sim, A.Y.L.; Minary, P.; Levitt, M. Modeling nucleic acids. *Curr. Opin. Struct. Biol.* **2012**, *22*, 273–278, doi:10.1016/j.sbi.2012.03.012.
33. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208, doi:10.1093/nar/gkp335.
34. Hiller, M.; Pudimat, R.; Busch, A.; Backofen, R. Using RNA secondary structures to guide sequence motif finding towards single-stranded regions. *Nucleic Acids Res.* **2006**, *34*, doi:10.1093/nar/gkl544.
35. Frith, M.C.; Saunders, N.F.W.; Kobe, B.; Bailey, T.L. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput. Biol.* **2008**, *4*, e1000071, doi:10.1371/journal.pcbi.1000071.
36. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, *31*, 3406–3415.
37. Hoinka, J.; Zotenko, E.; Friedman, A.; Sauna, Z.E.; Przytycka, T.M. Identification of sequence-structure RNA binding motifs for SELEX-derived aptamers. *Bioinformatics* **2012**, *28*, i215–i223, doi:10.1093/bioinformatics/bts210.
38. Xia, T.; SantaLucia, J.; Burkard, M.E.; Kierzek, R.; Schroeder, S.J.; Jiao, X.; Cox, C.; Turner, D.H. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* **1998**, *37*, 14719–14735, doi:10.1021/bi9809425.

39. Mathews, D.H.; Sabina, J.; Zuker, M.; Turner, D.H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **1999**, *288*, 911–940, doi:10.1006/jmbi.1999.2700.
40. Parisien, M.; Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **2008**, *452*, 51–55, doi:10.1038/nature06684.
41. Hermann, T.; Westhof, E. Non-Watson-Crick base pairs in RNA-protein recognition. *Chem. Biol.* **1999**, *6*, R335–R343.
42. van Dijk, M.; Bonvin, A.M.J.J. Pushing the limits of what is achievable in protein-DNA docking: Benchmarking HADDOCK's performance. *Nucleic Acids Res.* **2010**, *38*, 5634–5647, doi:10.1093/nar/gkq222.
43. Jeddi, I.; Saiz, L. Three-dimensional modeling of single stranded DNA hairpins for aptamer-based biosensors. *Sci. Rep.* **2017**, *7*, 1178, doi:10.1038/s41598-017-01348-5.
44. Popena, M.; Szachniuk, M.; Antczak, M.; Purzycka, K.J.; Lukasiak, P.; Bartol, N.; Blazewicz, J.; Adamiak, R.W. Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* **2012**, *40*, doi:10.1093/nar/gks339.
45. Biesiada, M.; Purzycka, K.J.; Szachniuk, M.; Blazewicz, J.; Adamiak, R.W. Automated RNA 3D Structure Prediction with RNA Composer. *Methods Mol. Biol.* **2016**, *1490*, 199–215, doi:10.1007/978-1-4939-6433-8\_13.
46. Frelsen, J.; Moltke, I.; Thiim, M.; Mardia, K.V.; Ferkinghoff-Borg, J.; Hamelryck, T. A probabilistic model of RNA conformational space. *PLoS Comput. Biol.* **2009**, *5*, e1000406, doi:10.1371/journal.pcbi.1000406.
47. Das, R.; Baker, D. Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 14664–14669, doi:10.1073/pnas.0703836104.
48. Kinghorn, A.B.; Fraser, L.A.; Lang, S.; Shiu, S.C.-C.; Tanner, J.A. Aptamer Bioinformatics. *Int. J. Mol. Sci.* **2017**, *18*, doi:10.3390/ijms18122516.
49. van Dijk, M.; Bonvin, A.M.J.J. A protein-DNA docking benchmark. *Nucleic Acids Res.* **2008**, *36*, doi:10.1093/nar/gkn386.
50. Roberts, V.A.; Pique, M.E.; Ten Eyck, L.F.; Li, S. Predicting protein-DNA interactions by full search computational docking. *Proteins* **2013**, *81*, 2106–2118, doi:10.1002/prot.24395.
51. Ruvkun, G. Molecular biology. Glimpses of a tiny RNA world. *Science* **2001**, *294*, 797–799, doi:10.1126/science.1066315.
52. Kruger, K.; Grabowski, P.J.; Zaug, A.J.; Sands, J.; Gottschling, D.E.; Cech, T.R. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **1982**, *31*, 147–157.
53. Tuszynska, I.; Magnus, M.; Jonak, K.; Dawson, W.; Bujnicki, J.M. NPDock: A web server for protein-nucleic acid docking. *Nucleic Acids Res.* **2015**, *43*, W425–W430, doi:10.1093/nar/gkv493.
54. Si, J.; Cui, J.; Cheng, J.; Wu, R. Computational Prediction of RNA-Binding Proteins and Binding Sites. *Int. J. Mol. Sci.* **2015**, *16*, 26303–26317, doi:10.3390/ijms161125952.
55. Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A.A.; Aflalo, C.; Vakser, I.A. Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 2195–2199.
56. Gabb, H.A.; Jackson, R.M.; Sternberg, M.J. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* **1997**, *272*, 106–120, doi:10.1006/jmbi.1997.1203.
57. Sternberg, M.J.; Aloy, P.; Gabb, H.A.; Jackson, R.M.; Moont, G.; Querol, E.; Aviles, F.X. A computational system for modelling flexible protein-protein and protein-DNA docking. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1998**, *6*, 183–192.
58. Carter, P.; Lesk, V.I.; Islam, S.A.; Sternberg, M.J.E. Protein-protein docking using 3D-Dock in rounds 3, 4, and 5 of CAPRI. *Proteins* **2005**, *60*, 281–288, doi:10.1002/prot.20571.
59. Ritchie, D.W.; Kemp, G.J. Protein docking using spherical polar Fourier correlations. *Proteins* **2000**, *39*, 178–194.
60. Mandell, J.G.; Roberts, V.A.; Pique, M.E.; Kotlovyy, V.; Mitchell, J.C.; Nelson, E.; Tsigelny, I.; Ten Eyck, L.F. Protein docking using continuum electrostatics and geometric fit. *Protein Eng.* **2001**, *14*, 105–113.
61. Roberts, V.A.; Thompson, E.E.; Pique, M.E.; Perez, M.S.; Ten Eyck, L.F. DOT2: Macromolecular docking with improved biophysical models. *J. Comput. Chem.* **2013**, *34*, 1743–1758, doi:10.1002/jcc.23304.
62. Dominguez, C.; Boelens, R.; Bonvin, A.M.J.J. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737, doi:10.1021/ja026939x.

63. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res.* **2005**, *33*, W363–W367, doi:10.1093/nar/gki481.
64. Banitt, I.; Wolfson, H.J. ParaDock: A flexible non-specific DNA–Rigid protein docking algorithm. *Nucleic Acids Res.* **2011**, *39*, doi:10.1093/nar/gkr620.
65. Yan, Y.; Zhang, D.; Zhou, P.; Li, B.; Huang, S.-Y. HDock: A web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* **2017**, *45*, W365–W373, doi:10.1093/nar/gkx407.
66. Yan, Y.; Huang, S. A New pairwise shape-based scoring function to consider long-range interactions for protein-protein docking. *Biophys. J.* **2017**, *112*, doi:10.1016/j.bpj.2016.11.2521.
67. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748, doi:10.1006/jmbi.1996.0897.
68. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461, doi:10.1002/jcc.21334.
69. Tuszynska, I.; Bujnicki, J.M. DARS-RNP and QUASI-RNP: New statistical potentials for protein-RNA docking. *BMC Bioinform.* **2011**, *12*, 348, doi:10.1186/1471-2105-12-348.
70. Miao, Z.; Westhof, E. A Large-Scale Assessment of Nucleic Acids Binding Site Prediction Programs. *PLoS Comput. Biol.* **2015**, *11*, e1004639, doi:10.1371/journal.pcbi.1004639.
71. Wang, L.; Brown, S.J. BindN: A web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. *Nucleic Acids Res.* **2006**, *34*, W243–W248, doi:10.1093/nar/gkl298.
72. Torabi, R.; Bagherzadeh, K.; Ghourchian, H.; Amanlou, M. An investigation on the interaction modes of a single-strand DNA aptamer and RBP4 protein: A molecular dynamic simulations approach. *Org. Biomol. Chem.* **2016**, *14*, 8141–8153, doi:10.1039/C6OB01094F.
73. Lensink, M.F.; Wodak, S.J. Docking, scoring, and affinity prediction in CAPRI. *Proteins* **2013**, *81*, 2082–2095, doi:10.1002/prot.24428.
74. Lensink, M.F.; Velankar, S.; Wodak, S.J. Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. *Proteins* **2017**, *85*, 359–377, doi:10.1002/prot.25215.
75. Pérez-Cano, L.; Jiménez-García, B.; Fernández-Recio, J. A protein-RNA docking benchmark (II): Extended set from experimental and homology modeling data. *Proteins* **2012**, *80*, 1872–1882, doi:10.1002/prot.24075.
76. Nithin, C.; Mukherjee, S.; Bahadur, R.P. A non-redundant protein-RNA docking benchmark version 2.0. *Proteins* **2017**, *85*, 256–267, doi:10.1002/prot.25211.
77. Huang, S.-Y.; Zou, X. A nonredundant structure dataset for benchmarking protein-RNA computational docking. *J. Comput. Chem.* **2013**, *34*, 311–318, doi:10.1002/jcc.23149.
78. Chen, Y.C.; Lim, C. Predicting RNA-binding sites from the protein structure based on electrostatics, evolution and geometry. *Nucleic Acids Res.* **2008**, *36*, doi:10.1093/nar/gkn008.
79. Jones, S.; Shanahan, H.P.; Berman, H.M.; Thornton, J.M. Using electrostatic potentials to predict DNA-binding sites on DNA-binding proteins. *Nucleic Acids Res.* **2003**, *31*, 7189–7198.
80. Mandel-Gutfreund, Y.; Schueler, O.; Margalit, H. Comprehensive analysis of hydrogen bonds in regulatory protein DNA-complexes: In search of common principles. *J. Mol. Biol.* **1995**, *253*, 370–382, doi:10.1006/jmbi.1995.0559.
81. Xu, D.; Lin, S.L.; Nussinov, R. Protein binding versus protein folding: The role of hydrophilic bridges in protein associations. *J. Mol. Biol.* **1997**, *265*, 68–84, doi:10.1006/jmbi.1996.0712.
82. Theobald, D.L.; Schultz, S.C. Nucleotide shuffling and ssDNA recognition in *Oxytricha nova* telomere end-binding protein complexes. *EMBO J.* **2003**, *22*, 4314–4324, doi:10.1093/emboj/cdg415.
83. Neidle, S. *Principles of Nucleic Acid Structure*; Elsevier: New York, NY, USA, 2010; ISBN 978-0-08-055352-8.
84. Steiner, T. The hydrogen bond in the solid state. *Angew. Chem. Int. Ed.* **2002**, *41*, 48–76, doi:10.1002/1521-3773(20021014)41:1<48::AID-ANIE48>3.0.CO;2-U.
85. Greenwood, N.N.; Earnshaw, A. *Chemistry of the Elements*; Elsevier: New York, NY, USA, 2012; ISBN 978-0-08-050109-3.
86. Dougherty, R.C. Temperature and pressure dependence of hydrogen bond strength: A perturbation molecular orbital approach. *J. Chem. Phys.* **1998**, *109*, 7372–7378, doi:10.1063/1.477343.
87. Wu, M.-Y.; Dai, D.-Q.; Yan, H. PRL-Dock: Protein-ligand docking based on hydrogen bond matching and probabilistic relaxation labeling. *Proteins* **2012**, *80*, 2137–2153, doi:10.1002/prot.24104.

88. Meyer, M.; Wilson, P.; Schomburg, D. Hydrogen bonding and molecular surface shape complementarity as a basis for protein docking. *J. Mol. Biol.* **1996**, *264*, 199–210, doi:10.1006/jmbi.1996.0634.
89. In Silico Aptamer Docking Studies: From a Retrospective Validation to a Prospective Case Study/TIM3 Aptamers Binding. *Mol. Ther. Nucleic Acids* **2016**, doi:10.1038/mtna.2016.84. Available online: [https://www.cell.com/molecular-therapy-family/nucleic-acids/fulltext/S2162-2531\(17\)30103-8](https://www.cell.com/molecular-therapy-family/nucleic-acids/fulltext/S2162-2531(17)30103-8) (accessed on 24 July 2018).
90. Charifson, P.S.; Corkery, J.J.; Murcko, M.A.; Walters, W.P. Consensus Scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100–5109, doi:10.1021/jm990352k.
91. Chen, Y.-C. Beware of docking! *Trends Pharmacol. Sci.* **2015**, *36*, 78–95, doi:10.1016/j.tips.2014.12.001.
92. Salsbury, F.R. Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. *Curr. Opin. Pharmacol.* **2010**, *10*, 738–744, doi:10.1016/j.coph.2010.09.016.
93. Dynamics of DNA Oligomers. *J. Biomol. Struct. Dyn.* **1983**, doi:10.1080/07391102.1983.10507437. Available online: <https://www.tandfonline.com/doi/abs/10.1080/07391102.1983.10507437> (accessed on 27 April 2018).
94. Levitt, M. Computer simulation of DNA double-helix dynamics. *Cold Spring Harb. Symp. Quant. Biol.* **1983**, *47 Pt 1*, 251–262.
95. Mackerell, A.D.; Nilsson, L. Molecular dynamics simulations of nucleic acid-protein complexes. *Curr. Opin. Struct. Biol.* **2008**, *18*, 194–199, doi:10.1016/j.sbi.2007.12.012.
96. Cheatham, T.E.I.; Miller, J.L.; Fox, T.; Darden, T.A.; Kollman, P.A. Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *J. Am. Chem. Soc.* **1995**, *117*, 4193–4194, doi:10.1021/ja00119a045.
97. Weiner, P.K.; Kollman, P.A. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *J. Comput. Chem.* **1981**, *2*, 287–303, doi:10.1002/jcc.540020311.
98. Galindo-Murillo, R.; Robertson, J.C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T.E. Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12*, 4114–4127, doi:10.1021/acs.jctc.6b00186.
99. Drew, H.R.; Wing, R.M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R.E. Structure of a B-DNA dodecamer: Conformation and dynamics. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 2179–2183.
100. Huang, J.; MacKerell, A.D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34*, 2135–2145, doi:10.1002/jcc.23354.
101. Svozil, D.; Šponer, J.E.; Marchan, I.; Pérez, A.; Cheatham, T.E., III; Forti, F.; Luque, F.J.; Orozco, M.; Šponer, J. Geometrical and Electronic Structure Variability of the Sugar–Phosphate Backbone in Nucleic Acids. Available online: <https://pubs.acs.org/doi/abs/10.1021/jp801245h> (accessed on 10 August 2018).
102. Zgarbová, M.; Šponer, J.; Otyepka, M.; Cheatham, T.E., III; Galindo-Murillo, R.; Jurečka, P. Refinement of the Sugar–Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. Available online: <https://pubs.acs.org/doi/abs/10.1021/acs.jctc.5b00716> (accessed on 10 August 2018).
103. Richard Sinden DNA Structure and Function. In *DNA Structure and Function*; Academic Press: San Diego, CA, USA, 2012; pp. 1–57. ISBN 978-0-08-057173-7.
104. Ahmad, S.; Keskin, O.; Sarai, A.; Nussinov, R. Protein–DNA interactions: Structural, thermodynamic and clustering patterns of conserved residues in DNA-binding proteins. *Nucleic Acids Res.* **2008**, *36*, 5922–5932, doi:10.1093/nar/gkn573.
105. Etheve, L.; Martin, J.; Lavery, R. Dynamics and recognition within a protein–DNA complex: A molecular dynamics study of the SKN-1/DNA interaction. *Nucleic Acids Res.* **2016**, *44*, doi:10.1093/nar/gkv1511.
106. Etheve, L.; Martin, J.; Lavery, R. Protein–DNA interfaces: A molecular dynamics analysis of time-dependent recognition processes for three transcription factors. *Nucleic Acids Res.* **2016**, *44*, 9990–10002, doi:10.1093/nar/gkw841.
107. Zandarashvili, L.; Esadze, A.; Iwahara, J. NMR studies on the dynamics of hydrogen bonds and ion pairs involving lysine side chains of proteins. *Adv. Protein Chem. Struct. Biol.* **2013**, *93*, 37–80, doi:10.1016/B978-0-12-416596-0.00002-6.
108. Chen, C.; Esadze, A.; Zandarashvili, L.; Nguyen, D.; Pettitt, B.M.; Iwahara, J. Dynamic equilibria of short-range electrostatic interactions at molecular interfaces of protein–DNA complexes. *J. Phys. Chem. Lett.* **2015**, *6*, 2733–2737, doi:10.1021/acs.jpclett.5b01134.

109. Caruso, I.P.; Panwalkar, V.; Coronado, M.A.; Dingley, A.J.; Cornélio, M.L.; Willbold, D.; Arni, R.K.; Eberle, R.J. Structure and interaction of *Corynebacterium pseudotuberculosis* cold shock protein A with Y-box single-stranded DNA fragment. *FEBS J.* **2017**, *285*, 372–390, doi:10.1111/febs.14350.
110. La Penna, G.; Chelli, R. Structural insights into the osteopontin-aptamer complex by molecular dynamics simulations. *Front. Chem.* **2018**, *6*, doi:10.3389/fchem.2018.00002.
111. Lin, P.-H.; Tsai, C.-W.; Wu, J.W.; Ruaan, R.-C.; Chen, W.-Y. Molecular dynamics simulation of the induced-fit binding process of DNA aptamer and L-argininamide. *Biotechnol. J.* **2012**, *7*, 1367–1375, doi:10.1002/biot.201200003.
112. Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007**, *450*, 964–972, doi:10.1038/nature06522.
113. Galindo-Murillo, R.; Roe, D.R.; Cheatham, T.E. Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAACGAACGAACGC). *Biochim. Biophys. Acta BBA Gen. Subj.* **2015**, *1850*, 1041–1058, doi:10.1016/j.bbagen.2014.09.007.
114. Galindo-Murillo, R.; Roe, D.R.; Iii, T.E.C. On the absence of intrahelical DNA dynamics on the  $\mu$ s to ms timescale. *Nat. Commun.* **2014**, *5*, doi:10.1038/ncomms6152.
115. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, doi:10.1021/ct200909j. Available online: <https://pubs.acs.org/doi/abs/10.1021/ct200909j> (accessed on 27 April 2018).
116. Matsumoto, A.; Olson, W.K. Sequence-Dependent Motions of DNA: A normal mode analysis at the base-pair level. *Biophys. J.* **2002**, *83*, 22–41, doi:10.1016/S0006-3495(02)75147-3.
117. Alexandrov, V.; Lehnert, U.; Echols, N.; Milburn, D.; Engelman, D.; Gerstein, M. Normal modes for predicting protein motions: A comprehensive database assessment and associated Web tool. *Protein Sci.* **2005**, *14*, 633–643, doi:10.1110/ps.04882105.
118. Da, L.-T.; Sheong, F.K.; Silva, D.-A.; Huang, X. Application of Markov state models to simulate long timescale dynamics of biological macromolecules. In *Protein Conformational Dynamics*; Han, K., Zhang, X., Yang, M., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 29–66. ISBN 978-3-319-02970-2.
119. Wako, H.; Endo, S. Normal mode analysis as a method to derive protein dynamics information from the Protein Data Bank. *Biophys. Rev.* **2017**, *9*, 877–893, doi:10.1007/s12551-017-0330-2.
120. Xiao, J.; Salsbury, F.R. Molecular dynamics simulations of aptamer-binding reveal generalized allostery in thrombin. *J. Biomol. Struct. Dyn.* **2017**, *35*, 3354–3369, doi:10.1080/07391102.2016.1254682.
121. Nimjee, S.M.; Oney, S.; Volovyk, Z.; Bompiani, K.M.; Long, S.B.; Hoffman, M.; Sullenger, B.A. Synergistic effect of aptamers that inhibit exosites 1 and 2 on thrombin. *RNA* **2009**, *15*, 2105–2111, doi:10.1261/rna.1240109.
122. Zeng, X.; Zhang, L.; Xiao, X.; Jiang, Y.; Guo, Y.; Yu, X.; Pu, X.; Li, M. Unfolding mechanism of thrombin-binding aptamer revealed by molecular dynamics simulation and Markov State Model. *Sci. Rep.* **2016**, *6*, doi:10.1038/srep24065.
123. Lin, J.-C.; Hyeon, C.; Thirumalai, D. Sequence-dependent folding landscapes of adenine riboswitch aptamers. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6376–6382, doi:10.1039/C3CP53932F.
124. Mechanical Unfolding of RNA: From hairpins to structures with internal multiloops. *Biophys. J.* **2007**, doi:10.1529/biophysj.106.093062. Available online: [https://www.cell.com/biophysj/abstract/S0006-3495\(07\)70884-6](https://www.cell.com/biophysj/abstract/S0006-3495(07)70884-6) (accessed on 2 May 2018).
125. Boniecki, M.J.; Lach, G.; Dawson, W.K.; Tomala, K.; Lukasz, P.; Soltysinski, T.; Rother, K.M.; Bujnicki, J.M. SimRNA: A coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.* **2016**, *44*, doi:10.1093/nar/gkv1479.
126. Dufour, D.; Marti-Renom, M.A. Software for predicting the 3D structure of RNA molecules. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *5*, 56–61, doi:10.1002/wcms.1198.
127. Liu, K.; Watanabe, E.; Kokubo, H. Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations. *J. Comput. Aided Mol. Des.* **2017**, *31*, 201–211, doi:10.1007/s10822-016-0005-2.
128. Liu, K.; Kokubo, H. Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations: A cross-docking study. *J. Chem. Inf. Model.* **2017**, *57*, 2514–2522, doi:10.1021/acs.jcim.7b00412.
129. Šponer, J.; Banáš, P.; Jurečka, P.; Zgarbová, M.; Kührová, P.; Havrila, M.; Krepl, M.; Stadlbauer, P.; Otyepka, M. Molecular dynamics simulations of nucleic acids. From tetranucleotides to the ribosome. *J. Phys. Chem. Lett.* **2014**, *5*, 1771–1782, doi:10.1021/jz500557y.



130. Case, D.A. Molecular dynamics and NMR spin relaxation in proteins. *Acc. Chem. Res.* **2002**, *35*, 325–331.
131. Chen, A.A.; Draper, D.E.; Pappu, R.V. Molecular simulation studies of monovalent counterion-mediated interactions in a model RNA kissing loop. *J. Mol. Biol.* **2009**, *390*, 805–819, doi:10.1016/j.jmb.2009.05.071.
132. Cheatham, T.E.; Case, D.A. Twenty-five years of nucleic acid simulations. *Biopolymers* **2013**, *99*, 969–977, doi:10.1002/bip.22331.
133. Lemkul, J.A.; Alexander, D.; MacKerell, J. Polarizable Force Field for DNA Based on the Classical Drude Oscillator: II. Microsecond Molecular Dynamics Simulations of Duplex DNA. Available online: <https://pubs.acs.org/doi/abs/10.1021/acs.jctc.7b00068> (accessed on 10 August 2018).



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).