Scuola Internazionale Superiore di Studi Avanzati - Trieste

Dissecting Mg^{2+}-RNA interactions using atomistic molecular dynamics

Author:
Richard ANDRÉ CUNHA

Supervisor:
Dr. Giovanni BUSSI

A thesis submitted in fulfillment of the requirements for the degree of Philosophiae Doctor in Physics and Chemistry of Biological Systems in the

Molecular and Statistical Biophysics Sector

SISSA - Via Bonomea 265 - 34136 TRIESTE - ITALY

SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI

DOCTORAL THESIS

Dissecting Mg^{2+}-RNA interactions using atomistic molecular dynamics

---

Author:
Richard ANDRÉ CUNHA

Supervisor:
Dr. Giovanni BUSSI

A thesis submitted in fulfillment of the requirements for the degree of Philosophiae Doctor in Physics and Chemistry of Biological Systems in the

Molecular and Statistical Biophysics Sector

October, 2017

Declaration of Authorship

I, Richard ANDRÉ CUNHA, declare that this thesis titled, "Dissecting Mg^{2+}-RNA interactions using atomistic molecular dynamics" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this Institute.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: 

Date: September 27th, 2017.

Contents

Dedicated to my family.

Chapter 1

Preface

The central dogma of molecular biology [1] summarizes one of the most important mechanisms for the functioning of living organisms, stating that deoxyribonucleic acid (DNA) is transcribed into ribonucleic acid (RNA), which is then translated into proteins. However, it is still not sufficient to capture how important RNA are for cellular life. Nucleic acids are at the core of any living cell on this planet and thus deserve indisputably deserve scientific attention. In particular, RNA molecules are proposed as the key chemical species that ignited the beginning of life on pre-biotic earth [2–4]. Independently of this hypothesis, studying RNA molecules today is essential for numerous applications in life sciences, spanning from drug development to cancer treatment [5–7]. That being said, in the last half century there have been unprecedented efforts into understanding RNAs and their role in the cell to the utmost detail. RNA is transcribed from DNA and translated into proteins, which then perform an abundance of functions in the cell. On top of that, it can catalyze chemical reactions, regulate gene expression and even carry genetic information which is retrotranscribed into DNA [8–11]. The outstanding versatility of RNA molecules is due to their unique chemical features, resulting in a very flexible backbone combined with strong interactions between the nucleobases [12]. The balance between canonical base pairs and a multitude of backbone conformations is the main factor for RNA being well structured yet dynamical [13]. On the other hand, RNA folding can only occur in the presence of positively charged particles that compensate the electrostatic repulsion arising from the negatively charged sugar-phosphate backbone, inevitably tying nucleic acids and ions together [14].

Metal ions are instrumental for proper RNA folding and dynamics, while also being crucial cofactors for ribozyme catalysis [15]. Monovalent cations (Na^+, K^+) are the workhorses compensating the overall negatively charge nucleic acids [16], while divalent cations are frequently the protagonists of relevant folding events and catalysis [17–19]. Mg^{2+} ions, which are the most

freely available divalent cations in cells, commonly perform as structural pillars in RNA tertiary structures [20, 21]. Despite the ubiquitous presence of Mg^{2+} around RNA, the experimental characterization of their interaction is challenging, because Mg^{2+} do not offer a direct spectroscopic handle for detection and requires high-resolution X-ray crystallography [22]. On top of that, their assignment through X-ray diffraction is difficult, since the Mg^{2+} is isoelectronic with water and Na^{+} ions [23, 24]. Therefore, the use of theoretical and computational tools can clearly help reinforce the experimental characterization of Mg^{2+}-RNA interaction and contribute to the most needed dynamical view of these molecules [25].

The results presented in this thesis aim to provide a meaningful description of the interaction between Mg^{2+} ions and RNA through atomistic molecular dynamics coupled with enhanced sampling techniques. The simulations done in this work were designed to tackle the two most fundamental issues in describing divalent ions interaction with RNA using molecular dynamics. First, the quality and fidelity of the models used, and second the proper sampling of rare events. Through the employment of modified state-of-the-art simulations techniques, I was able to predict Mg^{2+} binding sites and their correspondent affinities on an RNA duplex. The affinities qualitatively agree with the interaction frequency trends observed in the structural databases (PDB [1] or NDB [2]). Furthermore, I evaluated relevant aspects of RNA simulation concerning force field choices for Mg^{2+} ions, RNA backbone non-bridging oxygens, and water. Lastly, I developed a robust methodological framework that allows for future molecular dynamics simulations aimed to study multiple concurrent binding events associated with high free-energy barriers. Since RNA folding is intrinsically dependent on ionic conditions, I hope that this work will facilitate future research on this important subject.

The work presented in this thesis is organized as follows: Chapter 2 presents a general introduction to ribonucleic acids, their interaction with ions, and a few considerations on the experimental characterization of the Mg^{2+}-RNA interaction. Chapter 3 is dedicated to a brief review of the underlying theory supporting the simulation techniques used in this work, namely molecular dynamics and enhanced sampling. Chapter 4 presents a comparison of Mg^{2+} binding affinities obtained from different force fields against experimental titration affinities, and also from different methodological schemes. At last, Chapter 5 is devoted to a detailed discussion on how I combined

[1] https://www.rcsb.org/
[2] http://ndbserver.rutgers.edu/

well-tempered metadynamics, bias exchange, and replica-specific biases to assess Mg^{2+} binding in a flexible duplex. In the same Chapter, I discuss how flexibility, monovalent ion competition, and hybridization affect Mg^{2+} affinity to RNA.

The results presented in Chapter 4 are part of a manuscript in preparation. In addition, the data presented on 5 are largely based on the following publication:

- Richard A. Cunha and Giovanni Bussi. "Unraveling Mg^{2+}-RNA binding with atomistic molecular dynamics". *RNA* 23.5 (2017), pp. 628–638.

Parts of this thesis were inspired by the ensuing coauthored paper.

- Jiří Šponer, Giovanni Bussi, Miroslav Krepl, Pavel Banáš, Sandro Bottaro, Richard A. Cunha, Alejandro Gil-Ley, Giovanni Pinamonti, Simón Poblete, Petr Jurečka, Nils G. Walter, and Michal Otyepka. "RNA Structural Dynamics as Captured by Molecular Simulations: A Comprehensive Overview". *Chemical Reviews* (under revision).

Chapter 2

Introduction to ion-RNA interactions

This Chapter discusses fundamental concepts related to ribonucleic acids and their interaction with ions. Section 2.1 presents details about the structure of RNA, with emphasis on its implications for RNA function. Additionally, this part offers a brief insight on the biological impact of the RNA functions in the cell. Section 2.2 is entirely focused on the particularities of ions-RNA interactions and aims to portray the characteristics of the RNA ionic cloud, while Section 2.3 focuses on the unique role of Mg^{2+} ions to RNA. In the last part, Section 2.4, a few considerations on the relevant experimental techniques that support and validate a significant parcel of the work in this thesis are presented. Part of this Section focuses on a few approximations related to X-ray crystallography since the analysis of the crystallographic presence of Mg^{2+} is an integral part of future chapters.

2.1 Ribonucleic acids

There are five naturally occurring nucleobases that are primarily classified in purines and pyrimidines (see Figure 2.1). The purine nucleobases are adenine (A) and guanine (G), and the pyrimidine ones are cytosine (C), uracil (U) and thymine (T). Nucleobases determine the identity of nucleosides which contains a furanone-ring bound to the nucleobase through a glycosidic bond. The hydroxyl group bound to the 2'-carbon of the pentose classifies it as a ribose, as it can be seen in Figure 2.1. Nucleotides are the basic monomers of nucleic acids molecules and are defined as a nucleoside with a phosphate moiety bound to the pentose. Successive nucleotides are linked through phosphodiester bonds, in which the 3'-carbon atom of the ribose is connected through the phosphate to the 5'-carbon of the next one (see Figure 2.1). Therefore, nucleic acid molecules have unlinked nucleotides at the end

of the polyanionic chains with hydroxyl groups attached to the 5′ or the 3′-carbon. Consequently, RNA chains are asymmetric, such that the same sequence of monomers constitute different molecules if the first nucleotide starts at the 5′ or the 3′-end. Due to the direction of the synthesis of the RNA polymerase, the 5′-end of the chain is considered the beginning of the RNA strand. The sugar ring moiety is an integral part of both RNA and DNA backbones. The difference between ribose and deoxyribose is the key factor for the exceptional disparity in chemical and functional behaviors between the two groups of nucleic acids. The hydroxyl group in the 2′-carbon position in riboses significantly increases the reactivity of RNAs. This might be linked to the fact that nature chose the more stable and structured DNA molecule to carry genetic information, while RNA also performs other functions, including catalyzing reactions [18].

RNA and DNA bases forms the so-called Watson-Crick (WC) or canonical base pairs. In RNA molecules G pairs with C and A with U, while in DNA the A-U pair is substitute by A-T. The canonical base pair interactions contribute to the stability of RNA double helices by about 4-12 $\mathrm{kJ\,mol^{-1}}$ per base pair [27]. Many non-canonical interactions in RNA have been discussed. An example is the G-U wobble pair, which is present in a significant fraction of A-form RNA helices. There are a number of ways that nucleobases might interact with each other, which varies with the edges of the bases that are interacting. This leads to a repertoire of complex non-helical structures like the commonly found hairpin loops, or the complex pseudoknots [12, 28]. Apart from base-edges pairings, which are hydrogen-bond based, there is also stacking interactions between neighboring nucleobases. RNA and DNA duplexes are heavily stabilized by both base pairing and stacking. Among the helical structures of RNA the most common is the A-form duplex, which is also the most recurrent motif found in RNA structures deposited on the PDB [29, 30].

The depiction of RNA molecules is essentially done in three modes of incremental complexity: primary, secondary and tertiary structures. The primary structure of RNA molecules refers to the sequence in which its nucleotides are sequentially distributed in the chain. Sequencing techniques have emerged as revolutionary tools in genomics. The increased throughput and massive decreasing in cost, compared to older sequencing technologies, made an immense amount of data on primary structures of RNAs available [31, 32]. The secondary structure of RNAs displays all the WC base pairs in

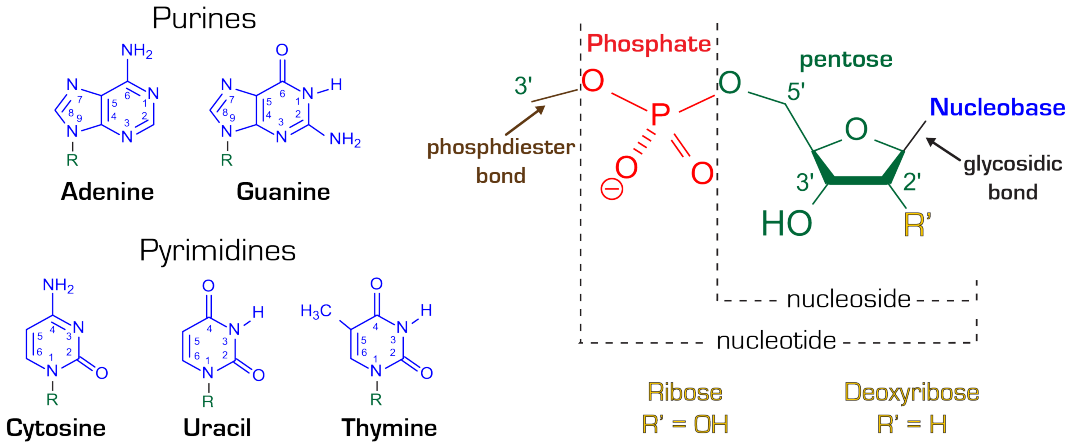


FIGURE 2.1: The figure shows the sketched chemical structures of all the five canonical, or primary, bases and their respective names. In the figure it is also evidenced the structural elements of a nucleoside and a nucleotide. The pentose ring is colored in green, the nucleobases in blue, and the phosphate moiety in red. The pentose and phosphates are part of the so-called RNA backbone, since they are common to any RNA molecule.

a molecule. As a result, the canonical contacts are then translated into a two-dimensional map. Although RNA molecules are mostly single-stranded, sequences of nonidentical nucleobases often fold towards themselves resulting in WC-complementary sequences thus producing a series of short antiparallel canonical double helices. Secondary structure only contains information about canonical pairs, but it completely lacks a description of non canonical interactions. Therefore, the information about the position with respect to other bases for some nucleotides is missing from secondary structure maps. The analysis of the tertiary structure of RNA molecules offers a complete three-dimensional molecular description. The polymeric and chemical nature of RNA allows for an organized yet dynamical tridimensional arrangement of the bases, which means that their tertiary structure does not arise from a single type of interaction. Even though canonical interactions are crucial for the stability of duplexes by locking the relative position of the bases, the intrinsic flexibility of the backbone brings into play numerous other degrees of freedom. There are approximately 50 rotamers which arise from combinations of seven consecutive dihedral angles along the backbone [33], see Figure 2.2. Therefore, the multitude of conformations resulting from the backbone flexibility promotes the formation of non-canonical interactions emerging from base-phosphate and base-sugar contacts making such interactions fundamental to RNA folding.

As discussed above, RNA structures are the product of a multitude of interactions within its moieties. Differently from DNA, the function an RNA

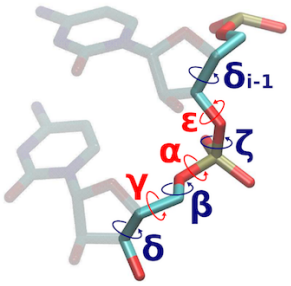


FIGURE 2.2: Figure showing the structure of a piece of RNA taken from a A-form duplex with the definition of the torsional angles indicated as suggested by the notation introduced by reference [33]. Figure adapted from [25]

molecule exerts in the cell depends on its tertiary structure, secondary structure (helices based on WC pairs), and on its primary structure. The relative importance of the RNA structural level depends on the precise function it will perform. As an example, ribozymes commonly catalyze reactions by spatially arranging reactants and products in a favorable way for the reaction to occur. Consequently, changes in the secondary structure that do not alter the catalytic scaffold are irrelevant for its functioning. Evolutionary analysis of the ribosome, one the most complex mixed RNA-protein machinery in the cell, concluded that only 72% of the primary structure is conserved while tertiary contacts are sustained up to 90% [34]. Such particularities make RNA an extremely versatile molecule, such that it participates in all essential molecular mechanisms of cellular life. The central dogma of molecular biology states that the genetic information carried by the messenger RNA (mRNA) is translated into a sequence of amino acids utilizing the transfer RNA (tRNA) as a molecular adaptor [1, 35, 36]. The versatility of RNA allows it to exert functions that extend further the central dogma. The discover of ribozymes paved the way for the idea that RNA could do more than transport genetic information [8, 37]. A significant amount of transcribed RNA is not coding for proteins [38]. Although there is much to know about the non-coding RNAs, a large fraction of it has already been assigned functions [39]. Non-coding regulatory RNAs such as riboswitches further confirm the array of genetic functions that RNA can perform [10, 40]. In conclusion, nowadays it is known that RNA can carry and store genetic information (retroviruses) and on top of that it can also perform functions once thought to be perfomed only by proteins, like catalyzing chemical reactions or regulate gene expression. The versatility of RNA is related to its enormous folding space, which is comparable, if not bigger, than the one of proteins while still carrying the

currency for genetic information.

2.2 The Ion-RNA Interaction

As discussed in the previous Sections, the three-dimensional arrangement of RNA molecules relies on specific molecular interactions that may occur within combinations of its constituent moieties. Since RNA is a polyanionic molecule, an obvious energetic barrier to the formation of any structural organization comes from the negative charge repulsion arising from the backbone phosphates. Thus, the effect of ions is surely crucial to RNA folding and function. In fact, ions stabilize all degrees of RNA structures [14]. At first, the overall charge screening provided by cations mediate the long-range repulsion that would not let RNA chains assemble by interacting with themselves [41]. Concerning the secondary structure, the presence of cations in the grooves of the A-form duplexes directly contributes to their stability [42]. Furthermore, certain ionic conditions can even enforce particular modes of helicity [43, 44]. At last, the formation of some tertiary contacts strongly dependent on the presence of ions, and even in some cases on ion-specific direct contacts [15, 16, 45]. In addition, displacement and interchanges of ions balance the free-energy of binding of RNA with other biomolecules such as proteins [46].

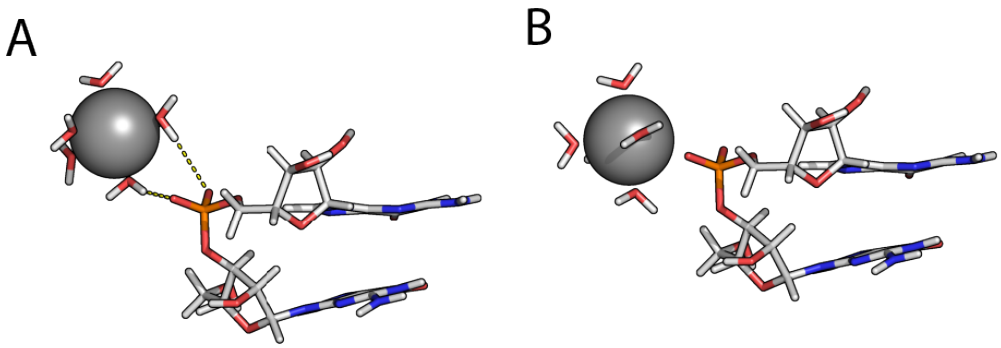


FIGURE 2.3: The figure shows the interaction modes of a Mg^{2+} ion (grey) with the phosphate moiety of a G dinucleotide. On the left panel (A) it is show a fully hydrated Mg^{2+} in which two waters of its fist coordination shell are forming hydrogen bonds with the phosphate oxygens. The hydrogen bonds are represented by the yellow dotted lines. On the right (B) it is show a partially dehydrated ion, chelated with the oxygen of the phosphate. This interaction is often called direct, or inner sphere contact. In this figure carbon atoms are represented in white, oxygens in red, nitrogens in blue and the phosphate atom in orange.

The ion-RNA interactions have long been appreciated, especially with

cations. In fact, the responsiveness of RNA folding to salt conditions was evident even in early studies of tRNAs [17, 47]. There are two main ways that cations interact with RNA. The first and most typical mode of interaction is through indirect contacts mediated by water molecules, often called outer-sphere or indirect contacts (Figure 2.3A). In this mode, the RNA atoms will be occupying the second coordination shell of the ion, thus the ion-RNA interactions is mediated by a water molecule. The other mode, called inner-sphere binding, occurs through direct contact between the ion and the RNA molecule (Figure 2.3B). Hence, an RNA atom is placed directly in the first coordination shell of the ion. This mode of interaction is less frequent, yet not less important. Some of the ions bound to RNA might be identified through X-ray crystallography and NMR spectroscopy [22, 48]. There are difficulties in the determination of the precise ion-binding positions through the interpretation of their electron density, yet most of the inner-sphere bound cations could, in principle, be characterized by X-ray crystallography. Notwithstanding, the amount of bound ions identified in all the known RNA structures does not compensate their total molecular charge, thus indicating that the majority of the ions interacting with RNA are diffuse and virtually invisible to traditional structural biology [49].

Concerning NMR spectroscopy, although it allows identifying ion-binding scaffolds in solution while accounting for the dynamic aspects of ion-RNA binding, the predictions are often less accurate since NMR rely on equivalent ions, regarding the charge, that can produce a spectroscopic signal [19]. Among the most used ions are transition metals that strongly interact with RNA due to large polarizability and/or free d orbitals. Although the spectroscopic properties of metals such as Mn^{2+}, Cd^{2+}, Pb^{2+}, Eu^{3+}, and Ti^{+} are very purposeful in terms of detectability, one has to assume that they will produce the same effects on RNA as the non-detectable cations commonly found *in vivo* (K^{+} and Mg^{2+}).

Despite the fact that ion-RNA interactions are crucial to its functioning, and that it is possible to characterize them using structural biology techniques, their physical properties and energetics are often misinterpreted. One might expect that a traditional two-state approximation would suffice to describe ion-RNA relations, but often that's not the case [14]. The charged chemical species around RNA form an ionic cloud that has a dynamic and non-trivial behavior [50]. Nonetheless, there are powerful underlying physical-chemical principles that help assessing the effects of a polyanionic molecule. The most important one is the concept of charge neutrality in solution. With

the support of the charge neutrality principle, one might define the RNA ionic cloud as the integration of all charged chemical species in the space around the nucleic acid such that the total charge of the ion atmosphere exactly cancels that of the nucleic acid. Moreover, the inner-sphere contacts might induce conformational changes that interfere in the underlying kinetics and thermodynamics of such interactions. [51].

The concepts behind ion-RNA interactions might seem counter-intuitive at times. The ion atmosphere is responsible for achieving charge neutrality, as emphasize before, however the amount of cations in it does not exactly match the negative charge of the nucleic acid. Figure 2.4, gives a overall idea of how Mg^{2+} ions might be distribute around a RNA duplex, but still a deeper description is needed in order to proper account for all the factors underlying the ion cloud behavior. This is because the amount of positive ions around RNA depends on its negative charge and on the increase of coion activity in the bulk. The increase of the anionic activity in the bulk occurs due to exclusion of coions from the RNA close surroundings. The energetic compensation provided by cations to RNA is massive. Moreover, there is also an entropic interplay arising from the organization of the layers of the ion atmosphere, and entrapment of ions by structural changes in the nucleic acid [52, 53]. Thus, RNA folding and ionic effects are so much interwoven that one is only comprehensible when considering the other as well [54].

The ionic atmosphere compensates the RNA charge and has a dynamic behavior, due to the interchangeability of the charged species in solution. The synergy of the ionic cloud with nucleic acids makes the direct ionic effect difficult to single out. There are a few techniques that can provide a estimation for the number of excess ions in the RNA ionic cloud, and they can help indicating how changes in the structure depend on ions. Measuring the interaction coefficients (Γ) it is possible to determine the number of ions detracted from the solution to serve exclusively to RNA charge screening, thus giving a more quantitative view of the ionic cloud around RNA. Γ_+ represents the number of $(PO_2)^-$ moieties that are neutralized by the excess of cations. Similarly Γ_- measures the charge neutralized by the deficiency of anions, and is negative. Γ_+ and $|\Gamma_-|$ must add to the total negative charge of the RNA. The definition of the interaction coefficient for monovalent cations can be expressed as

$$\Gamma_+ \equiv \left(\frac{\partial c_+}{\partial c_{RNA}}\right)_{\mu_+} \approx \left(\frac{c_+ - c_{bulk}}{c_{RNA}}\right)_{\mu_+}, \tag{2.1}$$

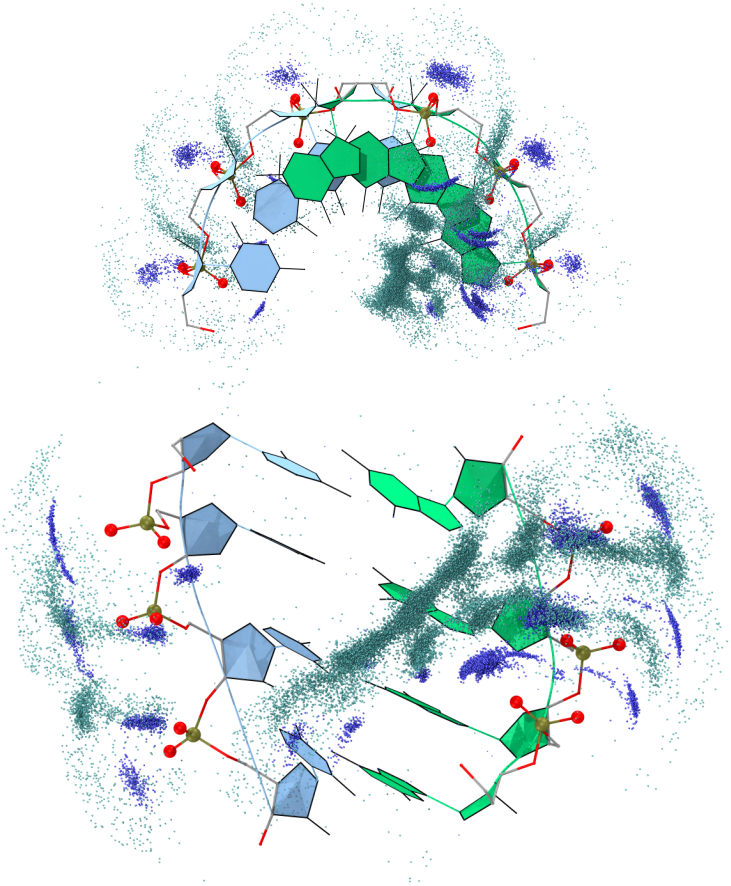


FIGURE 2.4: Each point displayed in this figure represents the position of a Mg^{2+} ion interacting with a $(GC)_4$ A-form duplex obtained on a 9 µs bias-exchanged simulation (details on Chapter 5). The upper and lower panel shows different angles of an overall view of the Mg^{2+} ion distribution. The blue colored dots, show directly bound ions while the green represents the indirectly bound ones.

in which c_+ and c_{RNA} are the molalities of the monovalent cations and RNA, respectively. At constant cation activity (μ_+), Γ_+ is the equivalent of measuring the 'extra' cations constricted to interact with RNA. This interactions could be through inner or outer-sphere constants, but it is important to note that these cations are detracted from the bulk. A way to measure Γ_+ is by performing a equilibrium dialysis experiment, in which variations in the c_+ can be measure in response to c_{RNA} and posed against c_{bulk} in the opposite site of the dialysis membrane [55].

Based on the main molecular forces driving the ion-RNA interaction it is possible to conceptually divide it in three different categories. First, diffusively associated ions which are mobile but still captured by the RNA electrostatic field. Notice that this class of ions are considered to be associated but not strictly bound to RNA, and are often separated from the nucleic acid by

two or more layers of water [56]. Secondly, indirectly bound cations, which might occupy a binding position through a water intermediated interaction, as show in Figure 2.3 A and in green points in Figure 2.4. The molecular forces associated with indirectly bound cations are more than only electrostatic, since the formation of hydrogen bond networks between the ions first-shell depends on a few factors like number of waters coordinated with their geometry complex. Moreover, polarization effects affect the strength of the hydrogen bond of the ion-coordinated waters [25, 57]. This kind of interaction might lead to a higher ion affinity for certain RNA scaffolds like the duplex major groove where the electrostatic potential is high [58]. However, fully hydrated cations, especially monovalents, are still fairly mobile with residence times reaching at most hundreds of ns [59]. Lastly, directly bound cations, in which the ion is coordinating a RNA atom. A special subclass of this mode of interaction are structural ions, which are often bound to more than one RNA atom and inserted deep inside the structures, and have important structural and catalytic roles. The proper description of this kind of interaction takes into account all the molecular forces involved with the previous ones, plus the dehydration free energies and the charge transfer between the ion and the RNA atom bound to it. As pointed out previously these kinds of ions have a very high residence time, which reaches up to ms.

RNA folds hierarchically and the process is environmentally dependent as is the stability of the folded structures. In fact, charge compensation performed by monovalent cations, which are abundant in the cellular environment, leads to the formation of secondary structure but not tertiary contacts. Divalent cations, are much more efficient in this task, given the entropic advantage of neutralizing the same amount of charge with half the number of ions [41]. Among the divalent cations, Mg^{2+} stands out as the most effective in RNA folding and stabilization, as it will be discussed in detail in the next section.

2.3 The importance of Magnesium ions to RNA

It is convenient first to understand the chemical attributes of cations which are relevant to how they contribute to RNA folding. Properties such as ionic radius, hydration free-energy, water coordination number, ligand exchange rate, and reactivity are key to define the ion interaction with RNA. Mg^{2+} and K^{+} are the most abundant cations *in vivo*, and also the most commonly found interacting with RNA. Due to their closed-shell electronic configuration, the metals from the first groups of the alkali and alkaline earth period interact with RNA predominantly through electrostatic forces. Thus, only part of the interaction is attributed to electron transfer effects meaning that they have a small reactivity in terms of formation of covalent bonds with RNA [60].

As a further matter, any alkaline earth metal has a smaller ionic radius than its immediate alkali group neighbor. The difference between the radii of Mg^{2+} and K^{+} is even bigger than of an alkali/alkaline earth group neighbors since the monovalent have an extra electron shell with respect to the divalent. As a consequence, Mg^{2+} ionic radius of 0.72 Å is roughly half of the K^{+} radius of 1.38 Å. The combination of a small radius with a +2 charge makes steric effects milder for Mg^{2+} and results in a clear enthalpic and entropic advantage on the ion-RNA interaction. The high charge density of Mg^{2+} let it accommodate six tightly bound water molecules in an octahedral geometry, while generating a strong enough dipole-dipole interaction to even have a structured second coordination shell [60]. In comparison, K^{+} can have six to eight water molecules in its first coordination shell, and its hydration free energy is around -330 $\mathrm{kJ\,mol^{-1}}$ while the Mg^{2+} one is -1900 $\mathrm{kJ\,mol^{-1}}$ [60–62].

The unique role of the Mg^{2+}-RNA interaction is defined by the way it interacts with RNA. Both direct and indirect contacts have entropic and enthalpic advantages in comparison to the same interaction done by other common cations in the cell. Mg^{2+} ions promote an overall RNA structure stabilization by charge screening. However, this kind of associated ions, interacting only through electrostatics, can often be equally substituted by high concentrations of monovalent cations. Nonetheless, to achieve the same positive charge density around the RNA surface, it is needed twice the number of monovalent ions, which introduces extra repulsion between themselves. In the case of other competing divalent cations, such as Ca^{2+}, their bigger ionic radius result in a lower positive charge density in comparison to Mg^{2+}. For these reasons, Mg^{2+} is a known RNA folding agent [63]. A common way to the ion-dependent dynamics of RNA folding is by plotting the fraction of

folding RNA molecules as a function of Mg^{2+} ion concentration, in an approach known as Mg^{2+} titration [64–66].

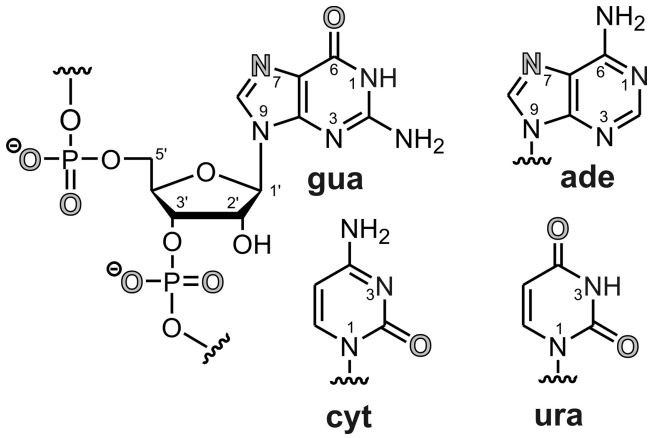


FIGURE 2.5: Putative Mg^{2+} binding sites in RNA. The four natural nucleobases in RNA are displayed together with their numbering scheme. In the case of guanine, the attached phosphate-sugar backbone are also shown. The major metal ion binding positions are shown in bordered bold grey. Additional binding sites could be adenine N1 and cytosine N3, but they are blocked by hydrogen bonds in Watson-Crick base pairs. In addition to the nucleobase sites show, ions seldom bind to adenine N3 [67], and to the ribose 2'-OH [68]. Similarly, steric impediments make the bridging oxygen atoms of the phosphodiester linkage not accessible for direct metal coordination. Figure adapted from reference [69].

In principle, Mg^{2+} could form complexes with any RNA atom with a free electron pair. However, due to either steric impediments or the disposition of the bases with respect to each other in a duplex, some sites do not bind metals. For example, the bridging oxygen atoms of the phosphodiester are steric inaccessible. Similarly, the adenine N1 and cytosine N3 do not bind Mg^{2+} since they are directly involved Watson-Crick hydrogen bonds. The possible metal binding sites are shown in figure 2.5.

Directly bound ions can coordinate multiple RNA atoms forming complex architectures in RNA structures. This kind of interaction is readily observed in the group II intron ribozymes and have an important catalytic role [18–20]. Zheng et. al. in reference [70], classified all directly bound Mg^{2+} ions found in the PDB database with their respective ligands. They proposed a repertoire of binding positions within the analyzed RNA structures, showing that the number of different ligands and their combinations formed by Mg^{2+}-RNA contacts is vast and not straightforward to predict.

2.4 Experimental description of the Mg^{2+}-RNA interaction

Visualizing the Mg^{2+} ions associated to RNA thus present in the ionic atmosphere is not trivial. The ever-changing nature of the ionic cloud surrounding nucleic acids makes it invisible to traditional structural biology techniques, such as X-ray crystallography, NMR, or (cryo-)electron microscopy. However, so-called ion counting experiments can estimate interaction coefficients (Γ) and give quantitative information about the constitution of the ions associated to RNA. These methods use atomic emission spectroscopy-based measurements to quantify the elemental composition of the solution around RNA [71–73]. As an example, buffer equilibration-atomic emission spectroscopy compares the ion concentration in the nucleic acid containing sample with the flow-through, buffer-only one [74]. Anomalous small-angle X-ray scattering is also employed in ion counting, and have the advantage of providing information about the spatial distribution of ions [75].

NMR and X-ray crystallography can characterize Mg^{2+} ions bound to RNA (Figure 2.3 A and B) . Positions determined by X-ray crystallography provide in-depth detail of the Mg^{2+}-RNA ligands and its binding positions but lack the description of dynamics provided by solution experiments. It is worth noting that many structures deposited in the PDB contain wrongly assigned Mg^{2+} ions [76, 77]. These errors have roots beyond the fact that Mg^{2+} is isoelectronic with Na^{+} [24]. In some instances, other divalent cations, such as Zn^{2+}, Mn^{2+}, and Cd^{2+}, which are often used in the crystallization process, bind to sites that are then misidentified as Mg^{2+} binding positions [78].

It is useful to have solution experiments as a complement to X-ray crystallography, and often NMR spectroscopy does the job. However, NMR has to rely on detectable ions such as $Co(NH_3)_6^{3+}$ and Mn^{2+} to mimic the indirect and direct bound Mg^{2+} ions [79]. If the precise site that the $Co(NH_3)_6^{3+}$ complex or the Mn^{2+} binds has a high enough occupancy, then NMR data will most likely agree with X-ray crystallography. Nonetheless, is not always true that the cations used in NMR experiments to substitute Mg^{2+}, always behave in the same way as Mg^{2+} [80].

There are experimental approaches that capture the thermodynamics of Mg^{2+}-RNA binding. As an example, Mg^{2+} titrations quantify the contribution of Mg^{2+} to folding using the Hill equation [64–66]. Other techniques

such as hydrolytic cleavage experiments, equilibrium dialysis, and spectroscopic techniques like EPR and NMR together with titration provides Mg^{2+}-RNA general affinity. However, a single RNA molecule often has several binding sites that are equivalent and have similar affinities. The difference between sites is mostly neglected since a direct characterization of their specific affinities is hard to get for complex RNA molecules. [19] More precisely, NMR chemical shift maps does not yield enough information to define if the changes with respect to Mg^{2+} concentration are due to direct metal ion binding or by structural changes due to Mg^{2+} coordination nearby [81].

Chapter 3

Mg^{2+}-RNA simulations

3.1 Overview

In this Chapter the basic theory behind the methodologies utilized in this work will be presented. The first section will be dedicated to describing the fundaments of a molecular dynamics simulation. Moreover, force field approximations relevant to a proper description of the interaction of Mg^{2+} with RNA (Sub-Sections 3.2.1 and 3.2.2) will be discussed. For a similar purpose, in Section 3.3 we will tackle difficulties associated to the Mg^{2+}-RNA binding sampling problem. In the same Sections it will also be discussed the rare event techniques which were applied to solve the aforementioned problem. Such techniques could be classified in two ways, which depend on the basic way they promote sampling, namely *importance sampling* and *annealing-based methods*. The objective of both approaches is the same, that is sample metastable states separated by high energy barriers. The difference is that the former does it by introducing bias on a specific relevant order parameter that can be rigorously removed later, while the latter does it by increasing the rate of barrier crossing events sampling from multiple replicas with different temperatures. The Section 3.3 will also focus on well-tempered metadynamics, which works in accordance with the *importance sampling* approach. Later on, it will be presented a description of an approach based on Hamiltonian replica exchange methods in with extra bias potential added designed to sample Mg^{2+}-RNA binding.

3.2 Molecular Dynamics

The purpose of molecular dynamics simulations (MD) of biologic systems is to obtain biochemical insights at atomistic resolution. To achieve this goal MD simulations utilize classical mechanics to let the positions and velocities

of the particles of the system of interest evolve through time while under a potential energy function designed to reproduce the relevant atomic interactions. In this way, when combining a starting model, usually obtained from experiments, with massive computational resources, MD simulations will produce trustable trajectories containing an ensemble of structures sampled from a canonical distribution. Furthermore, this kind of equilibrium computer simulation provide insights on how different molecular configurations are and quantify their correspondent populations within its sampling capabilities. Ensembles generated by MD might still linger around a metastable state, such as a folded structure. Recent progress in computational power and very expensive simulations described the folding of small proteins, with calculations that reach up to a few ms [82–84]. This kind of computational power is not readily accessible, and the majority of simulations of biomolecules are limited to a few dozens of microseconds at most and thus fated to sample transitions between states separated by a few K_BT. Nevertheless, MD can provide information on the molecular behavior response to a range of conditions such as different temperatures or solution composition, and even describe fast ligand binding events [85].

The fundamental recipe for MD has not changed much in the last few decades, meaning that the underlying algorithm and most of the approximations are still useful to an array of applications which aim to study the dynamics of systems in equilibrium [86, 87]. The next paragraphs will discuss the basic MD algorithm and the most common approximations utilized in simulations of biomolecules.

It is important to understand that MD simulations are performed within the domain of classical mechanics. This means that all quantum related effects should be in principle incorporated in the basic MD model, such that the interatomic interactions should be tuned to reproduce what in reality is a product of both quantum and classical effects. The atomic model used in MD is basically a charged Lennard-Jones particle, such that non-bonded interactions have three control parameters (σ, ϵ and q) and the bonded ones are harmonic potentials tailored to different interactions. As an example, the interaction of a Mg^{2+} ion with water which, previously discussed in Section 2.3, is dominated by electrostatics but still has a degree of covalent contribution and charge transfer. A percentage of the cation charge is redistributed to its six bounded water molecules resulting in partial charge smaller than $+2$, which is not directly taken into account on standard MD simulations [88].

Nonetheless, Mg^{2+} models are developed to reproduce certain experimentally observed properties, and this can be achieved by tuning its interactions so that their sum ends up cancelling out polarization effects and covalent effects [89]. Details on the parameters and development of models for simulating RNA and Mg^{2+} ions will be presented later.

MD simulations generate real time trajectories starting from a given initial three dimensional structure, whose set of vectors $r^N = \{r_1, r_2, r_3, ..., r_N\}$ contains the coordinates r_i of its N composing atoms, under a certain temperature and pressure. MD simulations assume that the N point particles interact through a continuous potential energy function, V_{r^N}, known as the force field (FF). The FF describes bonded interactions in terms of bond equilibrium length, angles between bonds and dihedral angles. Non-bonded interactions are characterized by Lennard-Jones and Coulomb pairwise potentials. The sum of the gradient of the energy function $V_{r_{ij}}$, where $r_{ij} = \|r_i - r_j\|$ calculated for every ij atom pair, results in the total force F_{ij}, or $-F_{ji}$, according to the equation

$$F_{ij} = \sum_{ij}^{N} -\nabla V_{r_{ij}}, (j \neq i). \tag{3.1}$$

Knowing the total force acting on each particle, by simply using Newton's equation one can calculate the acceleration of each particle, sub-sequentially updating its position for a δt time increment. A few algorithms [90] performs this operations in a optimized and reliable way, such as the leap frog [91], Verlet [92] and velocity Verlet [93]. Starting from a set of coordinates r^N and velocities (v^N), the velocity Verlet algorithm evolves them in time by Δt increments, which are typically of $2\,\mathrm{fs}$. The following repeating sequence of steps summarizes the way the algorithm works, given a starting set of r^N and v^N at time t,

1. compute velocities for half time step ($v^N_{(t+\Delta t/2)}$),
2. compute new positions at full time step ($r^N_{(t+\Delta t)}$),
3. compute forces using the new positions $V_{r^N_{(t+\Delta t)}}$,
4. compute new velocities at full time step ($v^N_{(t+\Delta t)}$),
5. advance to the next step and repeat.

Considering solvent molecules explicitly helps to mimic the molecular behavior of RNA in water properly contributing to the realism of MD simulations. The current computational power available allows for systems with

hundreds of thousands of particles to be simulated, making standard the use of explicit solvent. To minimize unphysical finite-size and boundary effects, periodic boundary conditions (PBC) are applied thus allowing to obtain solution properties [94]. The minimum size of a simulation box has to be considered with care, since the interaction of a particle with itself through its multiple periodic images may lead to artifacts, especially when simulating highly charged species such as RNA and its ionic cloud [95, 96]. In other words, the simulation box containing RNA and ions has to be large enough for the ionic atmosphere to equilibrate without interacting with a periodic image of itself [73, 97].

Despite the tremendous importance of ion binding to RNA structural dynamics, it is still a challenge to characterize it experimentally (see 2.2). Thus, efforts from MD-based simulations could complement the available experimental data and provide insights otherwise unreachable to each approach alone. Combining the trustability of experimental results with the real time description of RNA-ion dynamics arising from MD simulations, one might have enough information to interpret or even to plan new experiments. Thus, improving the quality of FFs and the reachable timescale for simulations might lead to a synergy of experiments and simulation which certainly would help to deal with the challenge of characterizing ion-RNA binding and dynamics. In cases in which the molecular phenomena to be studied can not be assessed through classical MD, the simulations might still support other approaches, for example complementing quantum mechanics (QM) calculations which are limited to systems with fewer particles [98].

3.2.1 RNA force fields

The development of models capable of accurately describing the dynamic properties of nucleic acids by mimicking their chemical interactions is a extremely difficult task. Years of FF developing and research result on very accurate proteins models [84] while for DNA [99] and RNA there is much to be done. The RNA tertiary structure depends on combinations of a very diverse array of conformations, while the tertiary structure of proteins relies on more stable and less numerous motifs [100]. RNA structure and dynamics is a result of a balance of forces arising from canonical and non-canonical base-pairing, stacking, sugar-base and sugar-phosphate contacts, solvent and ions interactions which results in a enormous available conformational space [101, 102]. The interatomic interactions in MD result from a potential energy function (V_{r^N}), which is parametrized to reproduce experimental observables. The most common FFs used for modeling RNA are CHARMM [103] and AMBER [104]. Although, they are parametrized differently, the functional form (V_r) for both FFs is similar. The AMBER FF uses this energy function

$$\begin{aligned} V_{r^N} = & \sum_{bonds} k_b(l - l_{(eq)})^2 \\ & + \sum_{angles} k_a(\theta - \theta_{(eq)})^2 \\ & + \sum_{dihedrals} \frac{k_\phi}{2}\left[1 + \cos(n\phi - \phi_{(eq)})\right]^2 \\ & + \sum_{non-bondedpairs} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \\ & + \sum_{non-bondedpairs} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}. \end{aligned} \tag{3.2}$$

The equilibrium values for bonds, angles, and dihedrals can be taken from experiments or high-level QM calculations. As show by the Coulomb non-bonded term of the FF, every atom has its own pre-set partial charge q_i which is kept constant throughout the whole simulation [105]. Even though they may be chosen to reproduce a few given ensemble properties such as the molecule electro-static potential (ESP), the charge distribution in real molecules is dynamic and structure-dependent [106]. An obvious consequence of this approximation is that the partial charge of a hydrogen bond donor would be the same independently of the nature of the acceptor. Since the differences in charge between the possible donor/acceptor pairs in RNA are not too drastic, the strength and equilibrium distance of the hydrogen bonds are captured

relatively well. Most importantly, QM calculations suggest that RNA properties which depend on particular hydrogen bonds or electrostatic interactions, such as base-pairing stacking stability, are relatively well described with the fixed charge model [107, 108].

A significant contribution to the success of the AMBER FF comes from a charge fitting that works for base pairing and stacking interactions [104, 105, 109]. Concerning the RNA backbone, the fixed-charges approach cannot concurrently describe the ESPs for all the multiple relevant backbone conformers (combination of dihedral angles) [33, 110]. A possible solution would be a framework in which the charge settings change as the RNA conformation demand. Still, one could opt for using polarizable FF developed for RNA. However, this approach not only is in its early stages of development but is also very computationally demanding [111–113].

The AMBER FF constantly improves, in particular with respect to the description of backbone dihedrals angles. Its latest version, referred from here onwards as AMBER-ff12, includes the parmbsc0 changes to the α/γ backbone torsions [115] and the OL corrections to the χ backbone torsion [116]. Another correction has been proposed, namely a change in the van der Waals parameters of the non-bridging phosphate oxygen, which will be referred as a "vdWbb" [117]. Another interesting aspect that should be taken into account is that the RNA molecules described by the AMBER-FF might present different physical behavior depending on the water model that they are used with. The most common water models are the fixed three-point charge ones, SPC [118] and TIP3P [119], and the fixed four-point charges TIP4P [120], TIP4P-ew [121] and the most recent OPC [122].

A recent study by Bergonzo and Cheatham III compared the relative conformational populations of r(GACC) and r(CCCC) tetranucleotides against different water models. The results presented in this paper were obtained using multidimensional replica exchange molecular dynamics since plain MD would not be computationally viable to sample the tetranucleotide conformations extensively. As shown in Figure 3.1, certain combinations of water models with AMBER-ff12 + vdWbb can drastically influence the final conformational ensemble. In addition, the population percentages do not qualitatively agree with the experimental ones. The r(CCCC) tetranucleotide most populated conformations are not even detected in the NMR data. In light of these results, when performing MD one has to check structures and results against experiments carefully [25].

In summary, FFs employed for RNA simulations have been developed to

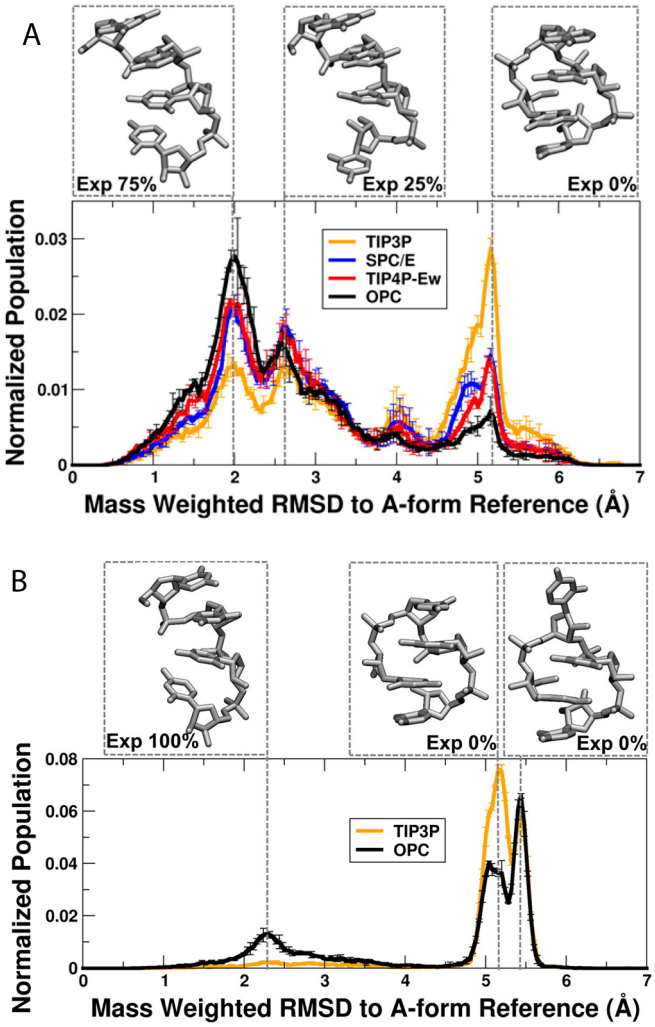


FIGURE 3.1: The figure shows the percentage distribution of RMSD values concerning the A-form structure for AMBER-ff12 + vdWbb (correction to vdW radii of nonbridging oxygen) for four water models. The histograms for the r(GACC) and r(CCCC) are taken from the lowest temperature replica (unbiased) and show in panel A and B respectively. The results on the bottom graph were obtained taking the two best water models from the top panel A. Error bars were calculated as the standard deviation from two independent runs. Figure adapted from Reference [114].

mimic RNA's physical properties, and are still under development. Despite all approximations, the performance of classical force fields has been readily tested and proved against experiments. Continuous FFs development and quality assessments are pushing the frontiers of MD simulations. Fundamental assumptions such as having fixed charges on atoms helps MD simulations become computationally viable. However, these same approximations might result in a lack of a fundamental physical basis to reproduce some molecular properties. Hence, fitting and tweaking parameters can only bring us so

far. Assuming that one is aware of the caveats and strengths of FF-based MD simulations of nucleic acids, there are many applications in which this technique can be extremely useful.

3.2.2 Mg^{2+} force fields

The inclusion of Mg^{2+} ions in RNA MD simulations is a delicate issue. Real charged molecules greatly affect the electronic environment around them. The high charge density of Mg^{2+} ions lead to strong polarization/charge-transfer effects. As an immediate consequence of this fact, Mg^{2+} ions alter the strength of the hydrogen bonds of the waters in its first coordination shell, and would also induce charge-transfer if bound to a RNA atom [123]. Figure 3.2 displays a schematic representation of charge-transfer effects upon Mg^{2+} binding to a non-bridging oxygen in the phosphate moiety of the RNA backbone. Pair-additive force fields are usually a combination of a 12-6 Lenard-Jones term and an electrostatic potential, which does not explicitly account for polarization and charge-transfer effects. To explicitly account for this effects, it would be necessary to use variable electrostatic terms and parameters that react to external electric fields [124].

There are many models employed to describe Mg^{2+} ions in MD simulations, the most common being the 12-6 LJ based ones. Despite their limitations, current nonbonded models are widely used in MD simulation. This could be due to their simple form, which can be adjusted to simulate desired experimental properties. It is believed that the most important properties of the Mg^{2+} ion aimed to be described are hydration free energies, distances between the ion and the water oxygen in the fist solvation shell, water exchange rate, and lastly the coordination number with water. It is obvious that these properties are correlated, and that are many combinations of parameters that might lead to the same values. The challenge is to find a set of parameters that offers a good compromise between all of them. Another point to be observed is that nonbonded models have limited transferability, meaning that different combination rules between FF and various water models result in different values of the desired properties.

Many efforts have been put into finding a suitable set of parameters for Mg^{2+} ions. In this work, we utilized the 12-6 LJ model for Mg^{2+} proposed by Allnèr, and Villa et al. [125] and Li and Merz et al. [126]. Li and Merz also introduced a 12-6-4 LJ model that, in principle, better describe ion polarization [127], which we briefly assessed (see Chapter 4). The first model (Alnnér-Villa) was designed so as to reproduce the water exchange rate of Mg^{2+} ions, in combination with the TIP3P [119] and SPC [118] water model. In Alnnér-Villa's work, the free-energy barrier of binding of Mg^{2+} ions to a dimethyl-phosphodiester, described by the CHARMM forcefield, was calculated using umbrella-sampling calculations, resulting in a good agreement

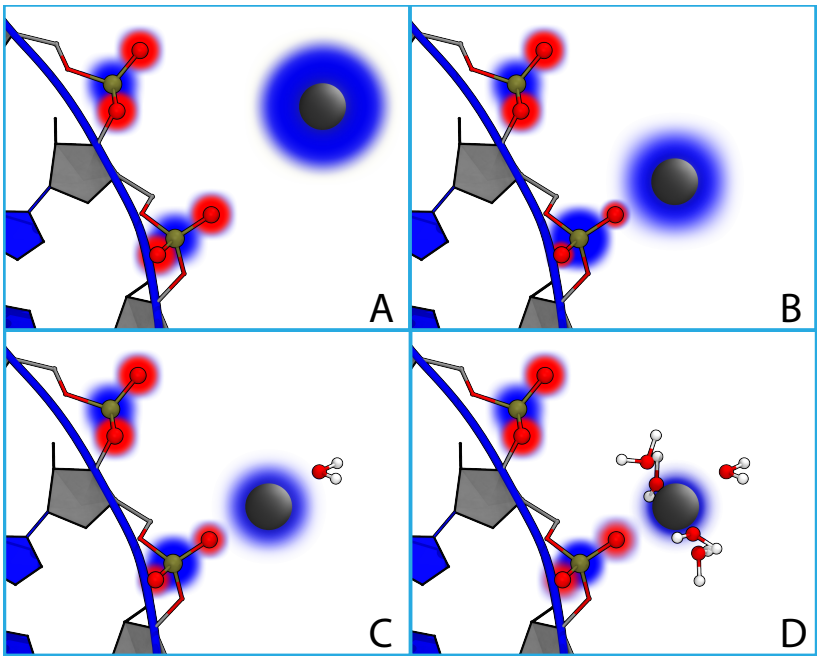


FIGURE 3.2: Schematic representation of charge-transfer effects upon Mg^{2+} binding to a non-bridging oxygen in the phosphate moiety of the RNA backbone. The different intensity of the shadings around the displayed atoms is a qualitative indication of how charges will be distributed in each scenario. Blue and red shadings depict the positive and negative charges, respectively. The red, white, ochre and black spheres are representations of the oxygen, hydrogen, phosphorous and magnesium atoms, respectively. It is worth noting that with the increase of the number of ligands in the inner coordination shell of Mg^{2+}, going from zero ligands in panel A, one in panel B, two in panel C and to six in panel D, the nominal charge of the cation is reduced. Also, the charge of the ligands is changed upon Mg^{2+} binding. The non-bridging phosphate oxygen donates charge to the Mg^{2+} ion while the phosphorous atom itself becomes slightly more positive. The changes in the charge on the non-bridging phosphate oxygens in panels B, C and D are related to the increase in the number of charge donors bound to the Mg^{2+} cation. Figure taken from [25].

with experiments (around 40 $\mathrm{kJ\,mol^{-1}}$). Li and Merz et.al work has a different approach, and calculated the best set of models to reproduce hydration free energies, distance between the ion and the water oxygen in the first solvation shell, and lastly the coordination number with water, for SPC, TIP3P, TIP4P [120] and TIP4P-ew [121] water models. Then, they proposed a compromise model for SPC, TIP3P, and TIP4P-ew, obtained by considering the best parameters for each property and aiming to minimize the error on the calculated properties. The parameters for the Li-Merz and Alnnér-Villa models are displayed on table 3.1

TABLE 3.1: Table containing σ and ϵ values for the 12-6 Lennard-Jones Mg^{2+} parameters for different water models

Mg^{2+} models	Water models					
	SPC		TIP3P		TIP4P-ew	
	σ (Å)	ϵ ($\mathrm{kJ\,mol^{-1}}$)	σ(Å)	ϵ($\mathrm{kJ\,mol^{-1}}$)	σ(Å)	ϵ($\mathrm{kJ\,mol^{-1}}$)
Allnèr-Villa [125]	2.77	0.0123428	2.77	0.0123428	–	–
Li-Merz [126]	1.36	0.0426867	1.36	0.0426867	1.353	0.0394048

3.3 Sampling Mg^{2+}-RNA interactions

In addition to the approximations discussed in previous Sections, one of the bottlenecks of MD simulations is the accessible timescale, which, in practical terms, is determined by the available computer resources. The standard MD recipe makes the molecular conformation of the system of interest evolve in real time by incremental steps in the order of fs. Timescales for Mg^{2+} binding on RNA could be from ns, for indirectly bound ions, up to ms, for directly bound ones. In a regular, single copy simulation of an RNA system in a solution containing Mg^{2+} ions, chelation events will rarely be observed, compromising the accuracy of thermodynamic and kinetic properties predicted from such calculation. Thus, to overcome sampling problems one might use the so-called rare event techniques, one of them being metadynamics.

Well-tempered metadynamics (WT-MetaD) is a methodology developed by Barducci, Bussi, and Parrinello [128] proposed in 2008. The algorithm is a variant of the original metadynamics approach [129] and, like its predecessor, can be coupled with MD-based calculations. WT-MetaD enhances the sampling of rare events that would otherwise not be observed enough times to generate reliable statistics. It can provide an estimation of a free energy $F(s)$ characterized by the existence of diverse metastable states and written with respect to a continuous function of the system coordinates x called a collective variable (CV) $s(x)$. The CVs are chosen based on *a priori* knowledge of the slow transition being investigated.

To estimate $F(s)$, WT-MetaD alters the dynamics of slow transitioning CVs by adding an adaptive bias potential ($V(s,t)$) built as a sum of Gaussians centered along the trajectory of the observable s:

$$V_G(s(x),t) = \int_0^t dt' \frac{w_G}{\tau_G} e^{(-\frac{V_G(s(x),t)}{k_B \Delta T})} e^{\frac{-\left[s(x)-s(x_{(t')})\right]}{2\sigma^2}}, \tag{3.3}$$

in which $w_G e^{(-\frac{V_G(s(x),t)}{k_B \Delta T})}$ is the height and σ the width of the Gaussians and τ_G the stride between the Gaussian deposition. In WT-MetaD, w_G decreases as $V_G(s(x))$ gradually increases, such that after some time the height of the Gaussians added is so small that the total bias potential will converge the potential will converge and the system will sample a quasi-static distribution. Such a potential will allow for a full exploration of the available space for the CV, it does so by reducing the probability of already visited states and scaling the barrier of the related transitions. Considering that the bias grows uniformly with time, it is possible to rigorously show [128, 130] that the bias

will tend to:

$$V(s, t \to \infty) = -\frac{\Delta T}{\Delta T + T} F(s) + C_{(t)}. \tag{3.4}$$

The conditional probability distribution of s will be:

$$P(s, t \to \infty) \propto \exp -\frac{F(s)}{k_B(\Delta T + T)}, \tag{3.5}$$

In wich ΔT sets the effective temperature for the CV.

The accuracy of the $F(s)$ estimate depends on the number of independent transitions observed, which is correlated with the simulation length, and with the choice of the parameters w_G, τ_G, and σ [131]. Additionally, the CV choice affects the convergence of the free-energy estimate independently of how correctly ω, τ_G, and σ are chosen. Thus, the selected CV has to be a good descriptor of the relevant metastable states and free-energy barriers of the process of interest [131].

It is possible to employ WT-MetaD simulations to describe direct-binding of Mg^{2+} ions to RNA. The interaction can be characterized using a combination of CVs, namely the distance between Mg^{2+} and the binding atomic position, such as an oxygen atom, and coordination number with water (CN_w). The distance well defines bound and unbound states to specific sites and at the same time enhances ion diffusion around RNA. Mg^{2+} diffusion is important since even when not directly bound to RNA, it might be trapped within its electrostatic potential minima and form indirect contacts that might be poorly sampled in short simulations. Biasing the Mg^{2+} coordination number with water enhances the first shell ligand exchanges which are necessary for the formation of Mg^{2+}-RNA complexes.

Another interesting point is that playing with different combinations of CVs or even changing the dimension of the bias potential applied could, in principle, allow a Mg^{2+} ion to bind to any electronegative atom in an RNA molecule. For example, one might try using multiple concurrent unidimensional WT-MetaD, as proposed for instance in reference [132], biasing separately the CN_w and all the distances to the possible binding sites, resulting in as many WT-MetaD biases as binding sites plus one. However, the binding mechanism is concerted, and changes in the distance and CN_w occur at the same time, having separated unidimensional WT-MetaD would significantly reduce the sampling efficiency. Moreover, distance alone is not a good CV to describe multiple equivalent metastable states, which is the case for Mg^{2+}-RNA binding. If multiple binding sites are simultaneously sampled, hysteresis could arise since same values of the distance between Mg^{2+} and

a target atom could correspond to different bound states, in various biding positions. A Mg^{2+} within 10 Å from a particular phosphate might be bound to a Guanine O6 or another neighboring phosphate moiety.

Besides the fact that obtaining relative free energies of different binding sites seems like a more interesting problem than having multiple Mg^{2+} competing to the same site but only being able to bind there, there might still be cases where the latter approach is interesting. In this way, assuming that only one Mg^{2+} ion binds at the time to the same position, a set of collective variables that describe well this interaction could be composed by the distance between the closest Mg^{2+} and the binding site and again the CN_w of each cation. In this case, hysteresis with respect to the distance might not be an issue. On the other hand, a practical difficulty could arise from having to calculate CN_w at every step, since its computational cost depends on the number of water molecules and scales linearly with the number of Mg^{2+} in solution. Usually, performing this calculation for a single Mg^{2+} ion slows down the simulation but not to the point that it becomes impractical, especially with the implementation of neighbor list searching [92]. Nonetheless, this approach would not be optimal, especially when other more practical solutions involving multiple replicas of the same system running in parallel are at our disposition.

WT-MetaD is suitable to describe the binding of a single Mg^{2+} ion with a single binding position, but complications arise when one aims to sample multiple ion-binding events to different binding sites concurrently. A possible way to tackle this issue is by using replica-exchange methods [133]. In this work, bias exchange metadynamics (BE-MetaD)[134] and Hamiltonian replica exchange (HREX) [135] were used. Both approaches derive from the same principle, where many simulations are performed at the same time using different control parameters and exchanging conformations at defined time interval with a certain probability calculated by the Metropolis-criterion [136].

BE-MetaD was proposed by Piana and Laio [134], and simulate each replica under a different metadynamics with its own set of CVs. After a defined time interval, exchange attempts between random replicas are performed, and their acceptance is calculated considering exclusively the replica-specific bias potential applied since each replica is described by the same potential function (FF) under the same temperature T. Considering n non-interacting replicas of a system, each of them under a time dependent bias $V^r(s(x),t) = V^r(s^r(x),t)$ applied in a WT-MetaD fashion to a replica-specific CV, $s^r(x)$,

$r = 1, \cdots, n$, the acceptance α of an exchange attempt is calculated as [134, 137]:

$$\alpha = min\left[1, \frac{e^{-\frac{1}{k_B T}\left(V^i(s(x_j),t)+V^j(s(x_i),t)\right)}}{e^{-\frac{1}{k_B T}(V^i(s(x_i),t)+S^j(s(x_j),t))}}\right], \tag{3.6}$$

This combined approach allows for the estimation of free energies when the dimensionality of the CV space is very large. In practice, the amount of replicas depends only on the number of CVs that one wants to investigate. The exchange between copies of the system introduces jumps in its configurations promoting further exploration of the CV space. That is, even if the WT-MetaD only addresses a few CVs per replica, by combining all the accessible individual replica CV spaces through exchanges, the system configurations will eventually explore the free-energy surface spanned by all CVs. Thus, it will significantly improve the free energy estimator, which is dependent on the correlation time of the dynamics in CV space [131].

Going back to the problem of sampling Mg^{2+}-RNA binding, BE-MetaD could, in principle, be straight-forwardly employed in the issue of sampling multiple RNA binding positions for a single Mg^{2+} ion. In principle, bi-dimensional WT-MetaD is suitable to describe the binding of a single Mg^{2+} ion with a single binding position, when applied using the distance between Mg^{2+} and the target binding position in combination with CN_w. Applying one independent WT-MetaD per binding site per replica would, in principle, solve the issue of sampling concurrently multiple binding sites. However, after a certain time the bias potential applied would allow for a hydrated Mg^{2+} to freely exchange water molecules while also accelerating its diffusion around RNA. Since the bias potential applied is similar to all the replicas and there is no barrier holding the ion from binding in other equivalent positions, nor force to promoting unbinding, Mg^{2+} could get stuck, by directly binding, to positions which are not the ones targeted to be sampled in that replica. Consequently, even exchanging conformations would only marginally improve sampling [26].

To tackle this matter, one might take advantage of the fact that any bias can be applied uniquely to each replica, and include a constraint to avoid Mg^{2+} binding on positions that are not being sampled on that specific replica. Then, when calculating the acceptance for a replica exchange, the extra penalty bias potential would also be taken into account. If Mg^{2+} binds to a prohibited position, the restraint potential will increase the difference in the energy between replicas leading to a high exchange acceptance such that these coordinates would migrate to a replica where this binding position is allowed.

The added biases, either from WT-MetaD or from the restrain Mg^{2+} binding, can be reweighted a posteriori, allowing to recover the equilibrium probability distribution [138, 139].

The next interesting point to assess is, how can one accelerate water exchange and Mg^{2+} diffusion and binding for multiple ions in many positions without making it impractically heavy in terms of computational cost. A possible answer is by enhancing water exchange tweaking the Mg^{2+} FF parameters with HREX while accelerating multiple Mg^{2+} binding events with BE-MetaD applied as described early in the Section. The difference in potential energy of replicas at the same temperature is the only parameter necessary to calculate the acceptance for swapping configuration, and could come from both re-scaling parts (or the whole) potential energy function of the system, or adding extra bias potentials using metadynamics.

HREX utilizes a similar approach to parallel tempering [133], where replicas are disposed in a sort of ladder and swaps are only between neighboring replicas. The lowest replica represents an ideal condition (experimental temperature), while the artificiality gradually progresses until reaching its maximum in the last replica (artificially high temperature), where the desired rare event has a higher probability of happening. In HREX, the first replica contains the standard FF parametrization the highest one a modified version of it, where transitions are more likely to be observed. The FF parameters in the highest replica might be also be standard MD FF parameters, but for a different particle or bond potential, for example Mg^{2+} in the lowest replica scaling to K^+ at the highest. Usually, the parameters of the FF are rescaled in a geometric progression in the replica ladder. In the Mg^{2+}-RNA case, the highest replica might be a Mg^{2+} ion with modified LJ radius to ensure water exchanges. In another words, if in the highest replica the LJ parameters of the Mg^{2+} allow for it to accommodate 7 ligands in its first coordination shell , it ensures that swap attempts down the replica ladder also mean ligand exchanges. The number of replicas would be chosen to find a compromise between a good exchange acceptance and computational cost and will depend on the number of modified particles.

Combining both approaches, BE-MetaD and HREX, one can sample Mg^{2+} water exchanges for multiple ions, while binding to multiple binding sites. Considering the most complex case in which multiple ions and various binding positions are concurrently sampled, adjustments in the definition of CVs would still be needed. Firstly, the number of replicas needed scales with the number of Mg^{2+} ions to be biased. All the Mg^{2+} ions are identical in all

the replicas, so applying WT-MetaD using the minimum distance between a binding site and a Mg^{2+} in solution as CV would promote binding, unbinding and ion diffusion. For multiple sites, one now can apply concurrently multiple one-dimensional WT-MetaD, which will be as many as the targeted binding sites. The bias introduced by the WT-MetaD can always be reweighted *a posteriori*.

In the next Chapters, I hope to present all the practical details needed to apply the aforementioned methodologies to sample Mg^{2+}-RNA binding.

3.4 Summary

In this Chapter, I discussed the basic of the methodologies used in this Thesis. A particular emphasis was given to the approximation and features of the approximations and models used in MD relevant to RNA simulations in water solutions containing Mg^{2+} ions. Mg^{2+} ligand exchanges happen in time scale not readly achievable for standard MD (Sections 3.2). Section 3.3 progressively explaining how enhanced sampling can be applied to describe Mg^{2+}-RNA binding. The Section starts describing how well-tempered metadynamics can be used to sample the simplest Mg^{2+}-RNA case (one ion directly binding to one site). Then, the discussion escalated to an innovative application of restraining potentials in bias exchange metadynamics which would then allow for a proper sampling of multiple binding positions concurrently. Lastly, it was presented a possible combination of bias exchange metadynamics with Hamiltonian replica exchange that ultimately would allow for many Mg^{2+} ions sampling to bind and unbind in multiple sites concurrently. The next Chapter will discuss results obtained by applying this methodologies.

Chapter 4

Force Field and Methodological Evaluation for describing Mg^{2+}-RNA binding

In this Chapter it will be discussed how the Mg^{2+}-RNA affinity is affected by the choice the Mg^{2+} parameters coupled with the water models and state-of-art RNA forcefields. In addition, it will be provided an assessment of convergence, strengths and caveats of different methodological approaches to sample RNA-Mg^{2+} binding through enhanced sampling. It was found that, the combination of Li-Merz Mg^{2+} parameters [126] with the Case et al. [117] corrections to the non-bridging phosphate oxygens of RNA is the best option for reproducing the experimental affinities. This result is consistent independently of use of TIP3P or OPC water models. Using replica exchange methods, such as HREX, in combination with multiple one-dimensional WT-MetaD, with the minimum distance between a Mg^{2+} and a binding site as collective variable, efficiently allow to capture Mg^{2+}-RNA interactions while being computationally accessible.

4.1 Introduction

Understanding the molecular behavior of ribonucleic acids in solution is an intrinsically difficult problem since even a well-defined sequence of nucleotides often can adopt multiple tridimensional structures within a tight range of folding energies of each other [140]. Different structures in an ensemble of RNA molecules can be interchangeable within thermal fluctuations. RNAs structure distributions are highly sensitive to interactions with proteins, co-factors, and ionic conditions [141]. For that matter, the effect of ions on RNA is particularly interesting since cations, such as Mg^{2+}, are essential for RNA

folding. Mg^{2+} ions are extensively present in molecular biology and have a unique capacity of shaping RNA structures [45, 142]. These cations compensate for the electrostatic repulsion of the RNA backbone and contribute to the overall decrease of entropy necessary for folding to occur [53, 143]. Most of the Mg^{2+} ions around RNA are loosely bound, present in the so-called ionic cloud that surrounds RNA molecules. This kind of ions indirectly interact with RNA atoms through water molecules forming complex hydrogen bond networks. Nevertheless, there is a small percentage of Mg^{2+} ions that tightly bind to RNA through direct coordination. In those cases, the directly bound Mg^{2+} ions are highly localized and have a slow ligand exchange rate, in the order of ms [14]. Due to the high residence times of directly bound cations, they are often correlated with the formation of well structured tertiary contacts [142]. Therefore, the proper accounting for ionic effects and positioning of magnesium ions in atomistic simulations can be determinant, if not essential, for a trustworthy description of nucleic acids in solution.

Biomolecular simulations face the fundamental issue that the more refined is a model, the more expensive a calculation become, thus the smaller will be the system scale and the reachable simulation time [144]. Even though the employment of massive resources in the development of dedicated hardware to perform MD lead to outstanding results, this technology is still not accessible to everyone [82, 83]. Thus, if not for enhanced sampling techniques, most of the modern molecular dynamics calculations would be limited to sampling conformations around a chosen RNA native structure [25]. Nonetheless, it is still possible to gather valuable insights of specific interactions through molecular dynamics. In the case of ions, a proper description of their effects on RNA molecules is hindered by two main factors, the precision of the models utilized, and the extent of sampling. With respect to the first problem, most of the current molecular dynamics simulations utilize ionic models consisting of an atom-centered point charged particle surrounded by pairwise Lennard-Jones based nonelectrostatic potentials. This kind of model is computationally efficient, however it can't directly include effects of polarization and charge transfer. In the case of Mg^{2+}, the ion's pairwise nonelectrostatic potential is fitted to reproduce meaningful properties such as water exchange rates, distance between the ion and the water oxygens as well as coordination number and hydration free-energies [126]. On top of that, Mg^{2+} ions parameters have to in constant improvement to couple well with the existing water and RNA force fields [89]. With respect to sampling issues, the indirectly bound ions are limited by their diffusion rates, which

are still accessible by standard molecular dynamics simulations [145]. Despite that, the exploration of all binding modes of Mg^{2+} is hindered by high free-energy barriers associated to water exchanges in the first coordination shell in combinations with ion competition and entropic effects [19, 26]. We believe that measuring binding affinities is a rigorous way to assess the synergy, or the lack of it, between different combinations of models for atomistic RNA molecular dynamics simulations.

4.2 Force field dependent binding affinities

To assess how RNA and Mg^{2+} force fields (FF) in combination with different water model influence the overall cation affinity to the RNA phosphate moiety, we performed a set of well-tempered metadynamics (WT-MetaD) [128] calculations for various FF combinations.

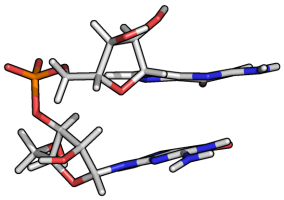


FIGURE 4.1: Structure of the guanine dinucleotide (GpG) taken from an ideal A-form helix.

The system composition was the same for all the simulations in the set, containing a guanine dinucleotide (GpG) without the 5′ phosphate. The initial structure for the GpG was generated by the *make-na* webserver [146] as if taken from an ideal A-form helix (shown in Figure 4.1). The dinucleotide was immersed in ≈ 2900 water molecules in a truncated dodecahedral also containing 0.1 $\mathrm{mol\,L^{-1}}$ of K^+ and Cl^- and one Mg^{2+} atom. Ions were initially included by substituting randomly selected water molecules. Water was modeled using the popular TIP3P [119] and the relatively recent OPC [122] force fields. TIP3P was chosen due to its popularity in the MD simulation community, while the OPC model has been shown to improve the agreement of RNA simulations with NMR data [114].

Molecular dynamics (MD) simulations were performed using GROMACS 4.6.7 [147]. Production runs were 1.0 µs long and performed in an NVT ensemble using a 2 fs time-step. The temperature was kept at 300 K by a stochastic velocity rescale thermostat [148]. Non-bonded and electrostatic interactions were calculated using the Verlet cutoff scheme and particle-mesh

Ewald respectively [149]. The FF used for K^+ was the one proposed by Cheatham et al. [150], and the GpG molecule and Cl^- were described using the AMBER-ff12 [104, 115, 116]. In some simulations, the parameter for the non-bridging phosphate oxygen from AMBER-ff12 was substituted by the one proposed by Case and collaborators in reference [117], and as mentioned in Section 3.2.1, when this modification is applied it will be noted by the "vdWbb" postfix. Mg^{2+} models were taken from the references Allnér et al. (Allnér-Villa) [125] and Li et al. (Li-Merz) [126], discussed previously on Section 3.2.2. It is important to note that Li and Merz et al. proposed different parameterizations to be used with different water models (SPC, TIP3P, TIP4P and TIP4P-ew), while Allnér and Villa et al. was designed based on SPC and TIP3P water models. The Li-Merz Mg^{2+} parameters for the TIP4P-ew were utilized in the simulations containing OPC water models since the models are very similar [122].

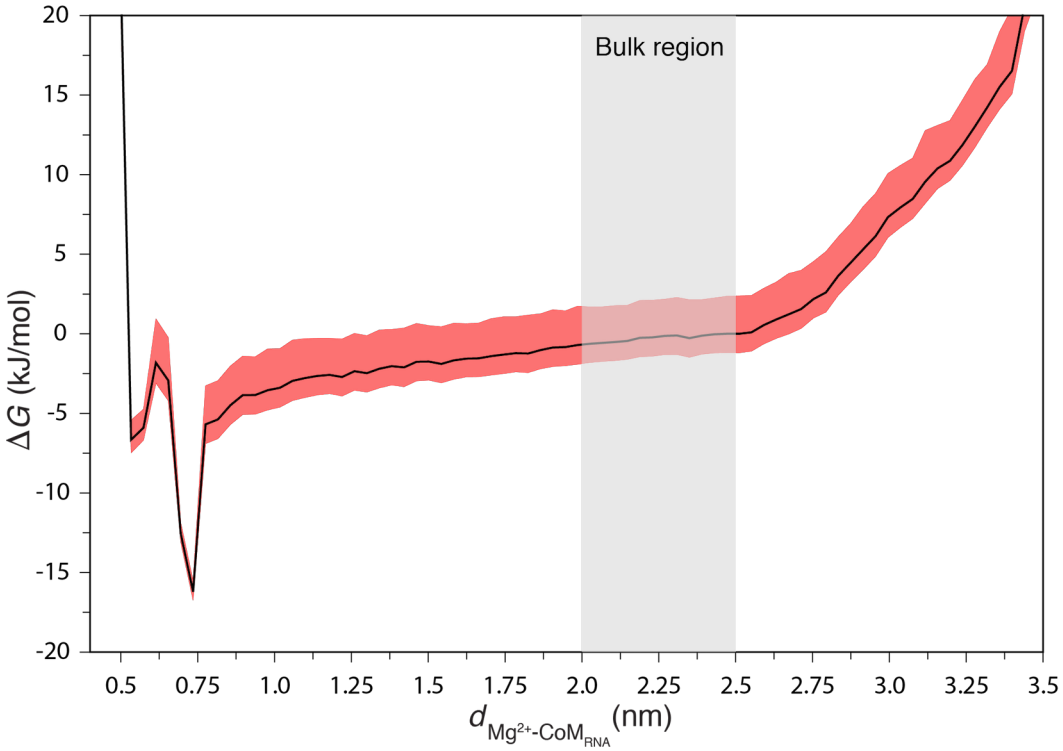


FIGURE 4.2: Free energy as a function of the distance between the biased Mg^{2+} and the center of mass of the GpG RNA fragment (CoM_{RNA}). ΔG is computed as $\Delta G_{(d)} = -k_B T \log P_{(d)} + k_B T \log d^2$, where $P_{(d)}$ is the probability to find Mg^{2+} at distance d from CoM_{RNA}. The error is shown in red shades and it was calculated using block analysis over 4 blocks. Since the Mg^{2+}-RNA distance is evaluated using the minimal image convention, it is not possible for this distance to be larger than half the box size. Thus, it can be seen that the free-energy profile consistently increases at that distance.

Well-tempered metadynamics (WT-MetaD) [128] was used to accelerate Mg^{2+}-$(PO_2)^-$ binding, applying bias on a two-dimensional collective variable (CV) composed of the distance between the Mg^{2+} ion and phosphate atom (d_{Mg-P}) in the GpG and the coordination number of Mg^{2+} with the water oxygen (CN_W). Coordination number is defined by the switching function

$$CN_W = \sum_i (1 + (d_{Mg-OW_i}/r_0)^6)^{-1}.$$

in which d_{Mg-OW_i} is the distance between Mg^{2+} and the i-th water oxygen (OW). r_0 is a cutoff that define if a water molecule is bound to Mg^{2+} and is set to 2.61 Å. The initial Gaussian height was set to 0.3 $\mathrm{kJ\,mol^{-1}}$, and the width 0.5 Å and 0.1 for d_{Mg-P} and CN_W respectively. Gaussians were deposited every 500 steps and the well-tempered bias factor (γ) was set to 15. PLUMED 2.2 was used to couple the WT-MetaD framework to the MD simulation and include any of the additional restraining biases. To optimize the sampling of Mg^{2+} binding to the phosphodiester moiety avoiding direct binding on other positions, repulsive harmonic potentials were activated every time the distance between the Mg^{2+} ion and the guanine O6 and N7 and the Cl^- ions was smaller than 3.0 Å. The force constant to the repulsive restraint was set to 50.0 $\mathrm{kJ\,mol^{-1}\,Å^{-1}}$. Also, the atomic positions for the RNA atoms were kept fixed, since a considerable extra amount of computational time would be needed to adequately sample all the GpG conformational space [151, 152]. The binding affinity (K_a) was calculated using the equilibrium distributions recovered from the WT-MetaD simulations. The K_a is computed as

$$K_a = \frac{w_b}{w_{shell}} V_{shell},$$

in which w_b is the accumulated weight of the bound states, defined by a cutoff of 2.6 Å on the distance between Mg^{2+} and a phosphodiester non-bridging oxygen (or equivalently 3.5 Å from the phosphate atom). The w_{shell} is the sum of the total weight in the bulk (shell) region, and V_{shell} is its volume. Weights are computed using the final WT-MetaD bias, as proposed in reference [153]. The bulk region is defined as the region in which the free energy with respect from the distance between Mg^{2+} and the GpG center of mass is flat (see Figure 4.2). The binding free energy at the standard 1M Mg^{2+} concentration was calculated accounting for the probability of the cation to be found in the bulk region [154] and is defined as $\Delta G = -k_B T \log K_a$. The statistical error was computed for all the reported quantities. For a free energies whose best estimate is X and the confidence interval is $[X - \Delta_1, X + \Delta_2]$, the

value is reported as $X^{+\Delta_2}_{-\Delta_1}$. Errors were computed by block averaging over 4 blocks without discarding any part of the simulation similarly to Reference [132]. Table 4.1 displays the binding free energies for different combinations of water, non-bridging phosphate oxygen, and Mg^{2+} FFs.

TABLE 4.1: Table containing Mg^{2+} direct-binding free-energies to the phosphodiester group in a guanine dinucleotide with respect to the different combinations of water, non-bridging phosphate oxygens and Mg^{2+} force fields. The ff12 and the ff12+vdWbb notations refer to the standard AMBER-ff12 FF and the Case et. al [117] versions to the phosphodiester non-bridging oxygen.

	TIP3P		OPC	
Mg^{2+} models	ff12+vdWbb	ff12	ff12+vdWbb	ff12
Li-Merz [126]	-12.5 $^{+1.1}_{-0.8}$	-19.3 $^{+1.3}_{-0.8}$	-15.7 $^{+3.3}_{-1.4}$	-20.0 $^{+3.3}_{-1.5}$
Allnér-Villa [125]	-14.6 $^{+1.0}_{-0.7}$	-30.2 $^{+0.7}_{-0.6}$	-28.1 $^{+1.5}_{-0.9}$	-29.7 $^{+1.6}_{-1.0}$

The experimental value for the standard binding free energy of Mg^{2+} cations to a phosphodiester moiety, estimated by Sigel et al. [155] using potentiometric titrations, is -6.0 $\pm$ 1.7 $\mathrm{kJ\,mol^{-1}}$. Although none of the available force fields can quantitatively reproduce the experiment, the AMBER-ff12+vdWbb FF combination lead to a better agreement with the experimental estimation of the free energy, for all the choices of water and Mg^{2+} parameters, with the exception of the Allnér-Villa parameters combined with the OPC water model. The Allnér-Villa parameters were designed to reproduce water exchange rates and the free-energy barrier of binding with an RNA phosphodiester, in combination with three-point charge water models (TIP3P and SPC), which might explain why it performs poorly when combined with OPC, but not in the TIP3P/AMBER-ff12 combination. Interestingly, even though the Li-Merz parameters were devised to reproduce a compromise between the average Mg^{2+}-water oxygen distances and hydration free-energies, disregarding the interaction with a phosphodiester moiety, it still consistently perform better than Allnér-Villa.

In another work, Li and Merz Jr. [127] suggested that a different formulation for the LJ function, including a term proportional to r^{-4}, might help to mitigate the (counter)effects of not explicitly considering polarization leading to a more accurate description of the interaction with water and highly charged species. We also calculated the affinity for the 12-6-4 model in combination with the TIP3P water and AMBER-ff12, which result in an affinity

of -14.0 $^{+0.9}_{-0.7}$ $\mathrm{kJ\,mol^{-1}}$. The 12-6-4 model performs slightly better than its 12-6 counterpart. However, this calculation was performed in a smaller simulation box ($\approx$ 1200 waters) and thus might not be directly equivalent to the other ones. The 12-6-4 functional form is not supported by GROMACS, and the extra term was added on the fly using the PLUMED pluging. The extra term was calculated for all the interactions between Mg^{2+} and the water oxygens, G-O6, G-N7 and the phosphodiester oxygens with the appropriate $C4$ value for each interaction pair. The extra potential term is a function of distance, and has to be calculated at every step, which substantially slows down the simulation. This result indicates that the standard 12-6 LJ formulation, while simpler and less computationally demanding, might still be a viable one in comparison to more refined functional forms.

It is noticeable that, considering the Li-Merz set, the difference in the affinity coming from the vdWbb modification to the AMBER-ff12 is within statistical error, but this does not hold for the Allnér-Villa parameter. The reason might reside on a balance between the strength of the Mg^{2+}-water / $(PO_2)^-$-water interactions, which leads to nontrivial effects. These results reinforce that 12-6 LJ parameters are not very portable across water models, and even RNA FF, and should be used with attention.

4.3 Methods to calculate converged Mg^{2+} $(PO_2)^-$ affinities

In the previous Section, the binding free energies of a single Mg^{2+} ion with the phosphodiester moiety of a guanine dinucleotide was presented as obtained through long WT-MetaD simulations. Equilibrium probability distributions obtained from these simulations helped assessing FF accuracy. Additionally, this approach could directly assist experimental interpretations. As an example, the relative affinity between directly and indirectly bound Mg^{2+} ions to a site could be used to further clarify NMR data, which often rely on the use of cations such as $Co(NH_3)_6^{3+}$ and Mn^{+2} to indirectly identify Mg^{2+} binding positions [71, 79]. In principle, it would be straight-forward to apply the methodology previously presented to sample a Mg^{2+} binding to a different, and more interesting, RNA molecule. Nonetheless, there are many applications for which sampling thoroughly a single binding position is not enough to properly describe ion binding. Biologically relevant RNA molecules have a number of binding positions that are often related. Sampling exclusively one binding position at a time would nullify possible correlation effects between binding sites in the estimated free energies. For example, the binding of the 5'-exon-intron recognition duplex of the *Sc*.ai5γ self-splicing ribozyme contains three metal ion binding positions that dictates the kinetics of the exon-intron binding according to their relative occupancy [156].

Concurrently sampling multiple binding positions with many Mg^{2+} ions using MD simulations would allow tackling more relevant and complex biological problems. Interestingly, it also steeply escalates the complexity of the sampling approach needed (see Section 3.3). Here, it will be discussed a set of simulations performed with elaborate methodological combinations to deal with multi-site and muti-Mg^{2+} concurrently sampling.

The set of simulations performed here follow the same MD protocol and is composed of the same system described in the previous Section. The AMBER-ff12 + vdWbb [104, 115–117] FF version was used in combination with TIP3P [119] water and Allnér-Villa [125] parameters for the Mg^{2+} cation. The Allnér-Villa parameters were chosen based on our previous work [26], that in this Thesis, is presented later in Chapter 5.

Table 4.2 summarizes the different methodological approaches for each simulation. The bias-exchange metadynamics (BE-MetaD) simulation (scheme A) was performed accordingly with our work (Chapter 5) [26]. This approach

TABLE 4.2: Table summarizing the methodologies utilized to sample Mg^{2+} binding to a GpG dinucleotide. In which, d_{Mg-X} is the minimum distance between a Mg^{2+} ion and an GpG atom X, which could be a P, 1G-O6, 2G-O6, 1G-N7, or 2G-N7 atom, and r is the replica index. The bias penalty potentials (restraints) do not allow direct Mg^{2+} binding to the Y GpG atom.

Label	Methods	CVs	Num. of Replicas	Restraints	Sim. time (ns)	Num. of biased Mg^{2+}
A	BE-MetaD	d_{Mg-X_r}, CN_W	5	d_{Mg-Y_r}, $\forall Y_r \neq X_r$	1125	1
B	HREX + WT-MetaD	d_{Mg-P}	6	–	2614	1
C	HREX + multi WT-MetaD	d_{Mg-X}	6	–	2614	1
D	HREX + WT-MetaD	d_{Mg-P}	9	–	3510	2

allows for concurrently sampling of multiple binding positions by a single Mg^{2+} ion. Each replica runs a two-dimensional WT-MetaD on the distance between Mg^{2+} and a target binding site together with its coordination number with water (CN_W). The number of replicas only depends on the number of putative binding sites to be sampled. In this approach 5 replicas were used to sample directly binding on phosphodiester moiety, the guanine O6 and N7 atoms. A group of replica-specific restraining bias potentials keeps Mg^{2+} from binding directly on the binding sites not targeted by that replica. The WT-MetaD bias, after a certain time, allows for the ion to freely diffuse around the RNA molecule while also easily exchanging its first coordination shell waters. The added restrained potentials avoid Mg^{2+} being stuck in positions that are not directly affected by the metadynamics bias. Exchange attempts between random replicas are accepted based on the difference on the total sum of the biases applied to each replica. However, this approach is designed for a single Mg^{2+} ion binding to a single site per replica, which does not directly consider possible correlations between binding sites.

Methodological schemes B, C and D utilize the same principle. Hamiltonian replica exchange (HREX), scaling up the σ parameter of the LJ function describing Mg^{2+} ions, enhances ligand exchanges in the first coordination shell of Mg^{2+}, while multiple concurrent one-dimensional WT-MetaD promotes ion diffusion (see Section 3.3). This approach does not require a two-dimensional CV to describe Mg^{2+} binding, since first shell ligand exchanges are now sampled independently from the distance. Schemes B, and C only differ in the number of concurrent WT-MetaD. Schemes B and D both only bias the minimum distance between Mg^{2+} and the phosphate atom, however scheme D has an extra Mg2+ ion.

The maximum value of σ was chosen to ensure that first-shell ligand exchanges are possible within thermal fluctuations in the highest replica, consequently improving Ergodicity in the lowest replica through conformational swaps. Mg^{2+} FFs are designed to reproduce the experimental Mg^{2+} CN_W of 6. Thus, ensuring that the highest replica at some point has 7 ligands significantly increases the probability that conformational swaps down the replica ladder also correspond to ligand exchanges. Based on this, we choose the minimum σ value necessary to have equilibrium between hexahydrate and heptahydrate forms with close to equal populations of the artificial ionic species. The number of replicas was decided in terms of a compromise between a good acceptance (larger than 15%) and computational cost. σ was scaled from 2.77 Å (original FF) to 3.30 Å in a geometric progression across the 6-replica ladder. Similarly to conventional parallel tempering, the number of replicas required to span a given range of sigma is proportional to the square root of the number of modified particles, which here are the Mg^{2+} ions. We thus increase the number of replicas proportionally to $\sqrt{n_{Mg}}$. For all the HREX simulations (Scheme B, C and D) exchanges were attempted every 400 steps (0.8 ps).

TABLE 4.3: Table containing Mg^{2+} direct-binding free energies to a guanine dinucleotide (GpG). The different methodological schemes (A to D) are summarized in Table 4.2. The binding free energies of the G-O6 and G-N7 are calculated from the average affinity over the two guanines nucleosides, since they are equivalent. It was not observed Mg^{2+} directly binding to the G-N3 atom in any of the simulations.

Binding position	Methodological schemes				
	A	B	C	D	
				Mg-1	Mg-2
G-N7	$35.7\ ^{+0.1}_{-0.1}$	$86.7\ ^{\inf}_{-1.7}$	$34.3\ ^{+0.5}_{-0.4}$	–	–
G-O6	$12.2\ ^{+0.3}_{-0.3}$	$8.1\ ^{+3.2}_{-0.8}$	$13.0^{+0.3}_{-0.2}$	$36.5\ ^{\inf}_{-1.7}$	$11.9\ ^{+7.1}_{-1.7}$
G-O1P	$-16.4\ ^{+1.0}_{-0.7}$	$-17.7\ ^{+0.6}_{-0.5}$	$-19.7\ ^{+0.2}_{-0.2}$	$-15.6\ ^{+0.8}_{-0.6}$	$-16.9\ ^{+0.8}_{-0.6}$
G-O2P	$-5.4\ ^{+2.4}_{-1.2}$	$-13.6\ ^{+0.7}_{-0.5}$	$-15.5\ ^{+0.1}_{-0.1}$	$-13.8\ ^{+0.4}_{-0.4}$	$-15.1\ ^{+0.7}_{-0.5}$

Data from different replicas were incorporated using the weighted histogram analysis method (WHAM) [138] combining the sum of the last bias from all the WT-MetaD [153] and the difference in the total potential energy, arising from the scaled LJ interactions, concerning the unmodified replica.

This procedure is closely related to the standard way used to analyze BE-MetaD simulations [139] but was here performed in a binless fashion that allows a large number of bias potentials to be simultaneously reweighted.

The binding free energies obtained from each method are presented in Table 4.3. The direct binding affinity of Mg^{2+} to the O1P atom dominates over the O2P, as G-O6 is also preferred over G-N7. The binding affinities of the nucleobase sites G-O6 and G-N7 are very small compared to the phosphodiester moiety. Zheng et al. performed an analysis of the binding frequency of Mg^{2+} ions in RNA structures deposited in the PDB that also show the same trends. [70]. The initial GpG atomic positions are taken from an ideal A-form duplex and are kept frozen throughout the simulation. The A-duplex is the most recurrent RNA motif, this could explain why experimental trends are recovered.

The methodological schemes A and C have dedicated bias potentials to promote sampling in positions G-O6 and G-N7. In contrast B and D, do not promote binding to the nucleobase. The clear difference between the estimated binding free energy in the A-C and B-D methodological pairs indicates that those sites can be very poorly sampled, even on a μs simulation which Mg^{2+} can, in principle, freely exchange first-shell ligands. Despite of that, the schemes A and C show a relatively good agreement reinforcing that at least the affinity ranks are consistent. Also, the difference in the binding affinity of the G-O6 and G-N7 is an indication that errors are underestimated.

The possible underestimation of the errors calculated through block analysis can be further assessed by analyzing the methodological scheme D. The difference in the free energies of binding between Mg-1 and Mg-2 could only come from statistical errors since the two cations are exactly equivalent. Despite that, the fact that free-energy values $\pm$ the error do not overlap is an indication of how much the errors are being underestimated. Taking the average difference between the affinities and comparing with the average errors values tells us that the underestimation of the error is about 0.5 $\mathrm{kJ\,mol^{-1}}$.

Interestingly, the scheme B, which utilizes multiple concurrent one-dimensional WT-MetaD has, in average, smaller errors than all the other approaches. A fair comparison can be made against the approach B, since the only difference is the extra one-dimensional WT-MetaD applied on the distances between Mg^{2+} and 1G-O6, 2G-O6, 1G-N7 and 2G-N7 atoms. The errors in the O1P and O2P affinities are higher that in the ones obtained using approach B, suggesting that, for this situation, having multiple concurrent metadynamics

help enhancing the sampling even on seemingly uncorrelated binding positions.

In summary, it is both practically and methodologically possible to simulate multiple RNA binding positions concurrently together with many Mg^{2+} binding. Applying multiple one-dimensional WT-MetaD together with HREX is a suitable combination for tackling Mg^{2+}-RNA binding issues. Apart from the normal computational cost of a MD simulation, the impact on the computational cost of adding extra WT-MetaD using the distance between Mg^{2+} and a putative binding position is irrelevant. Thus, the expensiveness of these approaches (B, C or D) increases with the number of replicas needed, which consequently scales with the square root of the number of Mg^{2+} ions to be addressed. The minimum number of replicas for addressing a single Mg^{2+} ion depends on its LJ parameters and requires previous short MD simulations to determine it. We believe that the methods and results presented here could be applied to more complex systems and biologically relevant problems.

4.4 Summary

In this Chapter I discussed the two main issues with simulating Mg^{2+}-RNA binding, namely the quality of the force fields and proper sampling. Section 4.1 is a short introduction with a few directions on why and how one could tackle Mg^{2+}-RNA binding problems. The next two Sections are dedicated to the methodology, results and discussion of the obtained binding free energies against various methodological approaches. In summary, the common 12-6 LJ parameters does not quantitatively agree with experimental affinities and also are not very portable across water models, and even RNA FF, and should be used with attention. With respect to sampling, rescaling the α LJ parameter from the ion non-bonded function using Hamiltonian replica exchanges allow for overcoming the very high dehydration free-energy barrier. This methodological approach, in combination with multiple concurrent well-tempered metadynamics, using the minimum distance between a Mg^{2+} ion and the binding sites to be addressed results in a computationally efficient sampling of Mg^{2+}-rna binding.

Chapter 5

Dissecting Mg^{2+}-RNA interactions

Interaction with divalent cations is of paramount importance for RNA structural stability and function. In this chapter it will be reported a detailed molecular dynamics study of all the possible binding sites for Mg^{2+} on a RNA duplex, including both direct (inner sphere) and indirect (outer sphere) binding. In order to tackle sampling issues, we develop a modified version of bias-exchange metadynamics which allows us to simultaneously compute affinities with previously unreported statistical accuracy. Results correctly reproduce trends observed in crystallographic databases. Based on this, we simulate a carefully chosen set of models that allows us to quantify the effects of competition with monovalent cations, RNA flexibility, and RNA hybridization. Our simulations reproduce the decrease and increase of Mg^{2+} affinity due to ion competition and hybridization respectively, and predict that RNA flexibility has a site dependent effect. This suggests a non trivial interplay between RNA conformational entropy and divalent cation binding. The results and methods presented in this chapter are based on the reference [26].

5.1 Introduction

The relevance of RNA in molecular biology has systematically grown since the discovery that it can not only transfer genetic information, but also catalyze chemical reactions [157] and RNA is now considered a key player in many of the regulatory networks of the cell [158]. Functions such as catalysis and regulation of gene expression rely on the peculiar structure and dynamics of RNA molecules. The folding of RNA three-dimensional structure stands in a delicate balance between canonical interactions and strong electrostatics mediated by the presence of cations. Cations, together with water, are indeed crucial to compensate for the large repulsion between the charged

phosphate groups present in the RNA backbone. They allow for the formation of tertiary contacts [13, 14], and can also provide entropic stabilization to RNA motifs [53]. Among cations, Mg^{2+} is particularly relevant because of its double charge and small radius and abundance in the cell [159]. Mg^{2+} can be both directly and indirectly bound to RNA, that is RNA atoms can be part of Mg^{2+} inner coordination sphere or interact through hydrogen bonds with its hydration sphere [160]. The inner sphere ions mainly contribute to the formation of specific structural motifs [52]. The outer sphere ones might also bind to specific motifs and additionally take part to the ion atmosphere and stabilize RNA structures by screening electrostatic repulsion [41]. Several experimental works have provided valuable insights on the thermodynamics of RNA-Mg^{2+} interactions in solution [51, 73, 161, 162]. Titration experiments have been used to characterize the overall affinity of RNA for Mg^{2+} [155]. Affinities for individual Mg^{2+} binding sites on RNA nucleosides and small RNA motifs have also been reported [19]. The role of functional metal ions on small RNAs have also been investigated through nuclear magnetic resonance spectroscopy [163–165]. However, the precise characterization of typical Mg^{2+} binding sites in large RNA molecules has largely been obtained by analyzing crystal structures [76, 166]. A recent database survey allowed for a classification of all the binding modes observed in crystallographic structures [70]. Molecular modeling could in principle provide a powerful tool to bridge the gap between detailed crystallographic structures and solution experiments [167]. In this respect, several works at different level of resolution have been reported, ranging from quantum-chemistry calculations [52, 88, 168–170], to explicit solvent molecular dynamics (MD) [97, 125, 171], implicit solvent methods [172, 173], and coarse-grained models [174]. Among the computational methods, MD presents an excellent compromise between accuracy and computational cost, though the development of an appropriate parametrization for Mg^{2+}-RNA interactions is still a debated topic [89, 145]. Moreover, due to the high energetic barriers involved in Mg^{2+}-RNA and Mg^{2+}-water interaction [175], which brings the lifetime of these complexes to the ms timescale, coupling of MD with enhanced sampling methods is required.

In this Chapter, it will be shown a unique combination of enhanced sampling techniques together with a recently published parametrization [125] for Mg^{2+} to compute its affinity on a flexible RNA duplex. RNA duplexes are the most recurring motifs observed in ribosomal RNA [30]. The computed

affinities for all the relevant binding sites are compared with previously reported thermodynamic data and with an statistical analysis of the protein databank (PDB). Furthermore, by performing simulations on an appropriately chosen set of model systems it was possible to investigate the interplay between Mg^{2+}-RNA binding affinity and competition with monovalent ions, RNA flexibility, and RNA hybridization.

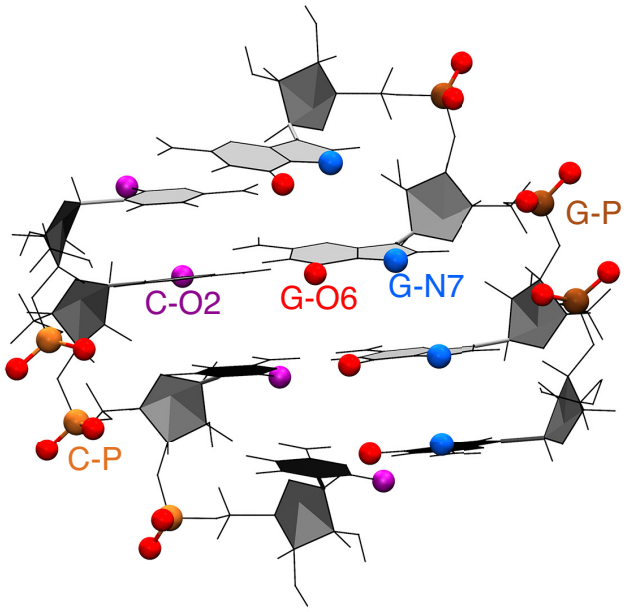


FIGURE 5.1: A-form RNA duplex with sequence $\frac{GGGG}{CCCC}$. Target Mg^{2+} binding sites are highlighted.

5.2 Methods

The initial RNA structure was generated using the *make-na* webserver [146] as an ideal A-form helix with sequence $\frac{GGGG}{CCCC}$ (see Figure 5.1). Molecular dynamics (MD) simulations of the duplex were performed using GROMACS 4.6.7 [147]. RNA was described using the AMBER-ff12 force field with parmbsc0 and χOL corrections [104, 115, 116]. The choice of the FF is based on the discussion presented on the 3.2.1. These parameters are available at http://github.com/srnas/ff. The modeling of the monovalent ions (K^+ and Cl^-), was done using the parameters proposed by Cheatham and cooworkers [150]. For Mg^{2+} we used a parametrization developed in ref. [125], which was also discussed further previously on chapters 3 and 4. The duplex was solvated in a truncated dodecahedral box filled with explicit TIP3P water molecules [176]. Ions were

added by substituting randomly selected water molecules. Bonds were constrained using the LINCS algorithm, and the integration of the equations of motion were performed with a 2 fs timestep. The temperature was set to 300K, and it was kept constant by a stochastic velocity rescale thermostat [148]. Non-bonded interactions were calculated using the Verlet cutoff scheme, and electrostatics using particle-mesh Ewald [149]. The cutoff was initially set to 1 nm and is adjusted adaptively so as to balance the load of real-space and reciprocal-space contributions. Pressure was kept constant at 1bar using the Berendsen barostat for the equilibration phase [177] and the Parrinello-Rahman scheme during the production runs [178]. Simulations with a rigid duplex had the RNA atomic positions kept fixed, and were performed at constant volume.

Enhanced sampling simulations were then performed combining GROMACS with PLUMED 2.2 [179]. A modified version of bias-exchange metadynamics (BE-MetaD)[134] was used to sample all the possible sites for inner sphere Mg^{2+} binding. In each replica we applied a bias potential according to MetaD in its well-tempered formulation,[128, 129] acting simultaneously on two collective variables, namely coordination number with water (CN_W) and distance d_i between Mg^{2+} and i-th target binding sites, resulting in as many replicas as potential binding sites. A bidimensional free-energy surface profile corresponding to binding on one of the duplex phosphate moieties is shown on Figure 5.2.

In this thesis we identify the possible binding sites for both inner and outer sphere binding using the name of the corresponding ligand (as highlighted in Figure 5.1). To this aim, we only considered binding sites with an expected affinity large enough to require enhanced sampling (phosphates, O6 and N7 in guanine, O2 in cytosine, see Figure 5.1). We chose these sites based on preliminary simulations that we performed for all 4 RNA nucleosides and a guanine dinucleoside monophosphate and also on available experimental information, such as the frequency that Mg^{2+} appear close to the a RNA atom in the PDB and potentiometric titrations. [70, 155] This procedure resulted in 18 replicas and a total simulation time of 9 μs (18 replicas $\times$ 0.5μs). With the exception of a control simulation all the MD runs contained only one Mg^{2+} ion, which is the one being biased by the MetaD. In the control simulation including multiple Mg^{2+} ions only one of them was biased. The coordination number of the remaining Mg^{2+} was restrained so as to avoid them to directly bind RNA. In simulations performed with a larger box to study RNA hybridization the total time was 3.6 μs (18 replicas $\times$ 0.2μs). On the table 5.1

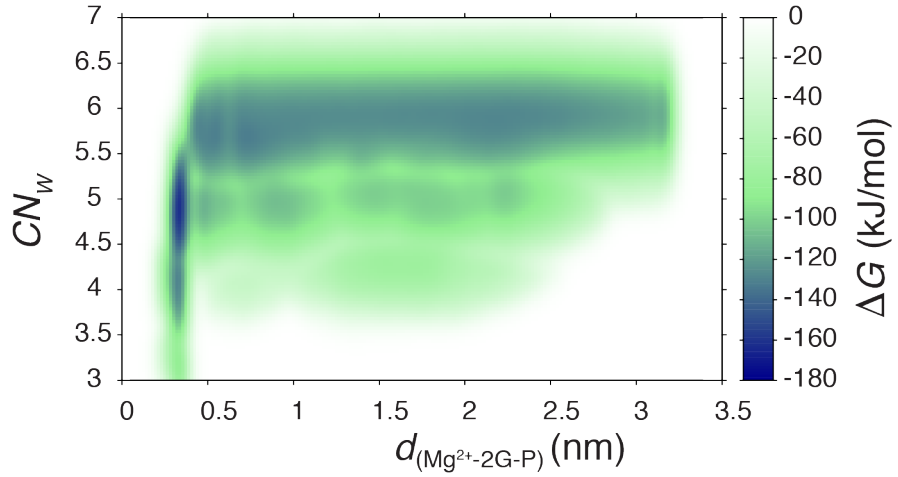


FIGURE 5.2: Sample free-energy profile as a function of the distance between Mg^{2+} and 2G-P (phosphate of second guanine) $d_{(Mg^{2+}\text{-}2G\text{-}P)}$ and of the coordination number of Mg^{2+} with water (CN_W), as obtained from the simulation with flexible RNA and explicit K^+ ions. The profile results from the bias potential accumulated on a single replica, and is thus subject to the penalizing restraints applied on that replica. An unbiased profile can be obtained using the reweighting procedure discussed in the methods, on the main text. In this figure it can be seen that Mg^{2+} can bind on other positions which are not the ones being biased by metadynamics on that replica ($CN_W < 5.5$ and $d > 0.5$). However, theres is a large penalty bias on those, which avoids the cation being trapped in those positions but still allows for sampling.

it is summarized the number of atoms, number of water molecules, number of ions and length for each simulation. Additionally, penalty bias potentials were added to avoid the competition of different binding sites in the same replica. This is not usually done in BE-MetaD but was required here to avoid Mg^{2+} to be trapped in unbiased binding sites. We also tried using the conventional BE-MetaD approach, but the sampling was undermined due to the Mg^{2+} being stuck in other positions. To ensure that the RNA helical structure was maintained through the enhanced sampling simulations, restraints were added to the distance and angles of all the hydrogen bonds corresponding to the four Watson-Crick base pairs. All the replicas were run simultaneously, and the acceptance rate was calculated taking into account the bias potentials introduced by the MetaD and penalty potentials on the unbiased binding sites. All the replicas have the same temperature, and the difference between the ensembles comes from the bias introduced by the MetaD and from the penalty potentials, which are unique for each replica. The acceptance for an exchange between replica i and replica j is thus evaluated as

$$\alpha = \min\left(1, e^{-\frac{V_i(q_j)+V_j(q_i)-V_i(q_i)-V_j(q_j)}{k_B T}}\right)$$

TABLE 5.1: Table containing the list of the studied systems, its components and the total simulation time. The simulations vary in the following parameters: small box (*sb*) vs large box (*lb*), rigid (fixed RNA atomic positions) vs flexible (restraints only on the hydrogen bonds of the Watson-Crick base pairs, for more details on the restraints used refer to the methods) and K^+ or KCl vs Uniform positive background (UB+). On the GpG the RNA atoms were not frozen, instead the RMSD with respect to an equivalent RNA duplex fragment was restrained to 0 nm.

System Description		Num. of atoms	Num. of ions			Num. of waters	V_{box}(nm^3)	Sim. time (µs)
			Mg^{2+}	K^+	Cl^-			
Nucleosides	Adenosine	1545	1	0	0	504	15.2	0.5
	Cytidine	2623	1	0	0	864	26.6	0.5
	Guanosine	1480	1	0	0	482	14.8	1.0
	Uridine	1698	1	0	0	556	17.02	1.0
	Rigid GpG	1814	1	0	0	582	18.3	1.0
GGGG CCCC	Rigid *sb* UB+	6577	1	0	0	2106	66.6	9.0
	Rigid *sb* K^+	6569	1	4	0	2102	66.6	9.0
	Rigid *lb* K^+	73298	1	4	0	24345	735.1	3.0
	SS Rigid *lb* K^+	73256	1	4	0	24339	735.1	3.0
	Flexible *sb* UB+	6580	1	0	0	2107	65.9	9.0
	Flexible *sb*	6572	1	4	0	2103	65.8	9.0
	Flexible *sb* KCl	6556	1	8	4	2095	65.8	9.0
	Flexible *lb* KCl	33333	1	24	20	11010	337.4	9.0
	Flexible *lb* KCl + MgCl	33323	4	22	24	11005	336.1	9.0
GGCC CCGG	GGCC Flexible *sb* K^+	6563	1	4	0	2100	65.9	9.0

Here V_i is the bias potential acting on replica i, including both the MetaD potential and the penalty bias potential, and q_i are the coordinates for the replica i. Ergodicity was thus ensured by accelerating the binding and unbinding events on all the possible binding sites with significant free-energy barriers. A sample input file for one of the model systems is provided in Figure A.1.

The K_a related to individual binding sites were calculated using the equilibrium distributions recovered from the BE-MetaD simulations. The simulations were reweighted using the umbrella sampling relationship [180], combining the last bias from MetaD [153] and the penalty bias potentials which were unique to each replica. Data from different replicas were combined with the weighted histogram analysis method (WHAM) [138]. This procedure is closely related to the standard way used to analyse BE-MetaD simulations [139] but was here performed in a binless fashion that allows a large number of restraints to be simultaneously reweighted. The binding free energy at the standard 1M Mg^{2+} concentration was calculated accounting for the

probability of the cation to be found in the bulk region [154]. The bulk region was defined as a spherical shell around the center of mass of the RNA duplex (CoM_{RNA}) in which the free-energy profile as a function of the Mg^{2+}-CoM_{RNA} distance was flat as shown in Figure 5.3.

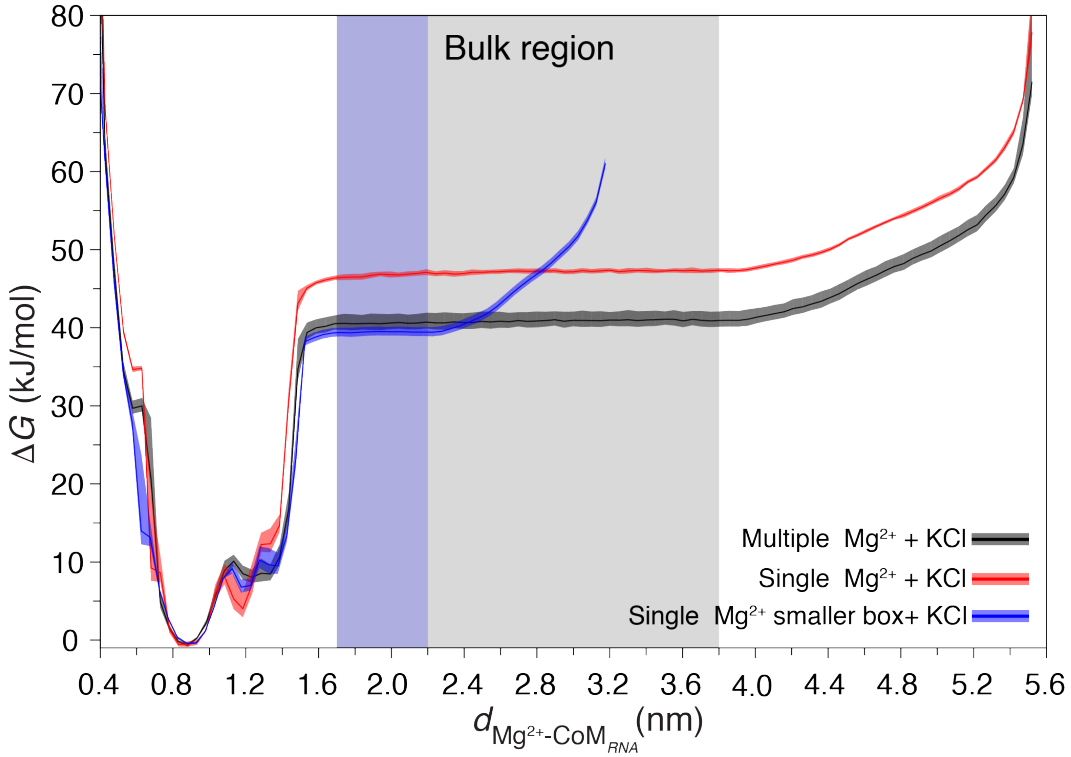


FIGURE 5.3: Free energy as a function of the distance between the biased Mg^{2+} and the center of mass of RNA (CoM_{RNA}). The curves shown on the figure were obtained from the control simulations performed on a flexible $\frac{GGGG}{CCCC}$ duplex contained in a large simulation box (≈11000 explicit water molecules) with a buffer of KCl at 0.1 M concentration (red curve), then the same box and a buffer with KCl and $MgCl_2$ at 0.1 M and 0.05 M concentration respectively (black curve) and lastly on a smaller box (≈2100 explicit water molecules) and a buffer of KCl at 0.1 M concentration (blue curve). The free energy (ΔG) is computed as $\Delta G_{(d)} = -k_B T \log P_{(d)} + k_B T \log d^2$, where $P_{(d)}$ is the probability to find Mg^{2+} at distance d from CoM_{RNA}. The error is shown by the red, gray and blue shades of the respective curves and it was calculated using block analysis over 4 blocks. The free-energy curves are aligned on the minimum and the bulk region is defined by the flat free energy part (shown by the light-gray and light-blue boxes) corresponding to the Mg^{2+} unbound states. Since the Mg^{2+}-RNA distance is evaluated using the minimal image convention, it is not possible for this distance to be larger than half the box size. Thus, it can be seen that the free-energy profile consistently increases at that distance.

The total weight w_i corresponding to the i-th binding site was obtained by accumulating the corresponding WHAM weights, and the affinities were

computed as

$$K_a^i = \frac{w_i}{w_{shell}} V_{shell}$$

Here w_{shell} is the total weight accumulated in the bulk (shell) region and V_{shell} is its volume. Mg^{2+} binding free energies were then defined as $\Delta G_i = -k_B T \log K_a^i$. Since K_a is expressed in molar units, a positive ΔG indicates that at a nominal Mg^{2+} concentration of 1 M one would expect the probability of finding a Mg^{2+} bound to be smaller than the probability of finding no Mg^{2+} bound.

The following set of simulations were designed and performed in order to evaluate the effect of the box size and of the ionic composition of the buffer: (I) a flexible duplex in a large box (≈11000 explicit water molecules) and a buffer of KCl at 0.1 M concentration; (II) a flexible duplex with the same box and a buffer with KCl and $MgCl_2$ at 0.1 M and 0.02 M concentration respectively; (III) a flexible duplex with a smaller box (≈2100 explicit water molecules) and a buffer of KCl at 0.1 M concentration. In this way, by comparing the affinities and the free energy of the biased Mg^{2+} against the center of mass of the RNA duplex (see Figure 5.3) we could single out the effect of the having extra Mg^{2+} in the bulk ((I) vs (II)) and the effects of the box size in the Mg^{2+} affinity ((I) vs (III)).

To dissect the contributions to Mg^{2+}-RNA binding we performed calculations on the following systems using a box with ≈2100 explicit water molecules: (a) a flexible duplex with and without explicit K^+; (b) a rigid duplex with and without explicit K^+; (c) a rigid duplex and two rigid separated strands with sequences GGGG and CCCC in a larger simulation box (≈24000 explicit water molecules). In this latter case the K^+ concentration was ≈0.01 M. When not using explicit monovalent ions, K^+ was replaced by a uniform positive background (UB+). This combination of setups allowed for the following factors to be considered: ion competition, RNA flexibility, and RNA hybridization. All the simulations followed the same protocols described above.

It is important to consider that a proper description of the kinetic and thermodynamic behavior of Mg^{2+} cations is very difficult to achieve without explicitly taking polarization and charge transfer effects into account [52, 124]. Several models have been introduced to effectively include polarization either using standard force field terms [125, 126], by means of *ad hoc* modified Lennard-Jones potentials [181], or within a Drude model [182]. We here decided to opt for the parameters developed in reference [125] which were optimized to improve Mg^{2+} kinetic behavior in water and interaction with

phosphate. We already used these parameters in previous applications to model ATP-bound Mg^{2+} and to describe the effect of Mg^{2+} on tertiary contacts in a riboswitch [183, 184]. We notice that a proper balance in Mg^{2+}-RNA interaction is not granted by available force fields [89]. For this reason, we checked the robustness of the reported results by using a reweighting procedure. We applied a pragmatic correction, adding *a posteriori* a contribution to the interaction between Mg^{2+} and individual binding sites on RNA proportional to a switching function $V_{correction} = \sum_i \lambda_i (1 + (d_i/R_0)^6)^{-1}$. Here d_i is the distance between Mg^{2+} and the i-th target binding site. R_0 is a cutoff radius that defines the directly bound state and is chosen to correspond to the barrier separating inner and outer sphere binding. λ_i are Lagrangian multipliers found with an iterative procedure so as to enforce the experimental value of the affinity on individual binding sites. Affinities calculated on nucleosides as well as λ_i and R_0 values are reported in Table 5.2. The weight used to compute the affinities are then corrected by a factor $e^{-\frac{V_{correction}}{k_B T}}$. This procedure follows the MaxEnt prescription [185] stating that the minimal correction to a force field so as to enforce the average value of an observable should be proportional to the same observable. This procedure is expected to provide results comparable to those reported in ref. [181]. The difference between the results with or without these corrections is discussed when appropriate.

5.3 Mg^{2+} binding on a flexible duplex

The main output of our simulations is the binding affinity of Mg^{2+} on all the possible binding sites in a RNA duplex. In this section it will be reported a detailed analysis of the obtained affinities and the correspondence with frequencies from the PDB. A comparison with the thermodynamic data available for nucleosides and a dinucleoside monophosphate is reported in Table 5.2.

In the next Sections will be shown a set of simulations performed in different conditions to dissect the important contributions to Mg^{2+}-RNA binding. The reported affinities were calculated by averaging over atoms of the same type in the central bases of the duplex, so as to mitigate terminal effects. For all the reported quantities we also computed statistical errors. For a quantity whose best estimate is X and the confidence interval is $[X - \Delta_1, X + \Delta_2]$, the value is reported as $X^{+\Delta_2}_{-\Delta_1}$. Errors are computed by block averaging over 4 blocks without discarding any part of the simulation. When relevant we also

TABLE 5.2: Affinities (ΔG_N) calculated for target binding sites on nucleosides (G, C, A and U) and a dinucleoside monophosphate (GpG), which is used here to probe the phosphate moiety. The experimental values reported here (ΔG_{exp}) are taken from Sigel and Sigel, Acc. Chem. Res. 2010, 43, 974. Lagrangian multipliers λ_i and cutoff values (R_0) are also reported. λ_i values were found with an iterative procedure and enforce the calculated affinity to be identical to the experimental one. R_0 is the cutoff distance which define the Mg^{2+} direct bound state.

Binding Sites		ΔG_N (kJ/mol)	ΔG_{exp} (kJ/mol)	λ_i (kJ/mol)	R_0 (nm)
G	O6	$11.7\ ^{0.4}_{-0.4}$	-2.6	-13.72	0.267
	N7	$37.2\ ^{0.7}_{-0.5}$	-2.6	-13.72	0.280
C	O2	$18.2\ ^{inf}_{-4.4}$	-2.9	-5.85	0.245
Phosp.	O1P	$-30.3\ ^{0.5}_{-0.4}$	-4.3	25.27	0.261
	O2P	$-28.6\ ^{2.6}_{-1.3}$	-4.3	25.27	0.261
A	N1	$58.7\ ^{0.4}_{-0.4}$	—	—	0.270
	N7	$44.5\ ^{0.5}_{-0.4}$	-2.0	-55.74	0.270
	N3	$55.0\ ^{5.2}_{-1.6}$	—	—	0.270
U	O2	$23.4\ ^{0.2}_{-0.2}$	-2.0	-17.60	0.263
	O4	$11.3\ ^{0.01}_{-0.01}$	-2.0	-17.60	0.263

discuss the results obtained by applying a correction that enforces the experimental affinity on all the binding sites of a nucleoside (MaxEnt correction, see 5.2).

Table 5.3 reports the binding affinity for Mg^{2+} of a flexible RNA duplex in presence of explicit K^+ and Cl^- ions. Reported results are obtained with the large simulation box (≈ 11000 water molecules). With our approach one can obtain affinities for both inner and outer sphere binding on all the possible binding sites. For the sake of clarity we repeat here that outer sphere binding is defined by any state in which at least one explicit water molecule is between Mg^{2+} and a electronegative donor.

Affinity for inner sphere binding is clearly dominated by the phosphates, with a preference for the strand composed of guanines. We observed a significant preference for direct binding on G-O2P with respect to the G-O1P. This is not seen for cytosines, where the difference between C-O1P and C-O2P is within statistical error Nitrogens that are involved in base pairing never formed direct contacts with Mg^{2+}. Affinity of C-O2 was extremely low, being

surpassed by the O2′ in the sugar moiety. The only atoms in the nucleobase displaying significant affinities were G-O6 and G-N7. All these observations are in striking agreement with interaction frequencies observed in the PDB taken from ref. [70] that are also reported in Table 5.3. The only exception is the inversion in the binding free energies of C-O2 and O2′. Our underestimation of the affinity of C-O2 might be biased by our choice to simulate a RNA duplex. Indeed, affinity of C-O2 is expected to be increased when cytosine is not involved in a canonical base pair. Interestingly, in a simulation performed on a isolated nucleoside (see Table 5.2) the affinity of C-O2 was significantly larger. Still, it is possible that neighboring Cytosines hinder the affinity of inner-sphere binding Magnesium(II) to the O2 atom, since we observed that in the rigid simulations containing two separated single strands (SS Rigid lb , described in Table 5.1), we did not observe binding in the central C-O2 atoms but only in the terminal ones. It might be that this is an artifact since the RNA atoms were frozen, or the duplex was too short.

When analyzing the outer sphere binding, both nucleobase and phosphate backbone contribute to the overall affinity (see Table 5.3). Also in this case, there is a preference for the strand composed of guanines, and binding on G-O2P is more favorable with respect to binding on G-O1P. The affinity of G-O6 and G-N7 is comparable to the affinity of phosphates. Also the sugar oxygens have a relatively large affinity. These observations agree with the interaction frequencies observed in the PDB. The only two exceptions are related to O5′ and O3′, for which the reported frequency is low and to the G-O6 and G-N7 affinities which are in the inverse order but within error of each other. The former discrepancy could be related to the fact that, at variance with our approach, in the reported experimental frequency the outer sphere binding with phosphates were excluded from the count on the O5′ and O3′ interaction frequencies. The latter discrepancy could be related to the sequence we choose to sample. It must be also noticed that binding of Mg^{2+} on G-N7 can happen in a large variety of equivalent and consequently non-comparable structural contexts. On another note, it there are structures deposited on the PDB wrongly assigning magnesium(II) cations bound to G-N7. [78] Moreover, the difference in the reported experimental frequencies is very small.

One might be tempted to convert the observed frequencies into binding free energies that can be quantitatively compared with our results. Even though the correlation is good ($R^2 = 0.61$ for inner sphere and $R^2 = 0.67$ for outer sphere binding), indicating that the ranking is consistent, the slope

of the fitting line is very far from unity, as shown in Figure 5.4. This might be due to imbalance in the force field [89] as well as to the fact that PDB distributions are not necessarily representative of the canonical or any thermodynamic ensemble. Additionally, it is not clear how much these frequencies could be used to anticipate location of Mg^{2+} ions in solution.

Remarkably, our calculation can recapitulate the most important trends observed in experimental frequencies, namely: preference for G with respect to C; preference for major-groove with respect to minor groove; relative preference between all the relevant binding sites.

TABLE 5.3: Calculated Mg^{2+} affinities on a duplex and PDB frequencies from ref. [70]. Frequencies for sugar and phosphate moieties were reported independently of the base identity.

Binding sites			Inner sphere		Outer sphere	
			ΔG^{inner} (kJ/mol)	F^{inner}_{PDB}	ΔG^{outer} (kJ/mol)	F^{outer}_{PDB}
Bases	G	N1	—	—	7.6	0.22
		N2	—	—	4.0	0.11
		N3	—	0.002	5.6	0.12
		N9	—	—	0.6	0.01
		N7	25.2	1.35	-10.9	3.62
		O6	22.4	1.45	-10.3	3.84
	C	N1	—	—	15.2	0.008
		N3	—	0.01	21.8	0.33
		O2	59.9	0.14	4.6	0.36
Sugar	G	O2′	33.4	0.07	0.4	0.54
		O3′	—	0.04	-5.3	0.55
		O4′	—	0.004	5.7	0.07
		O5′	—	0.04	-8.6	0.61
	C	O2′	34.4	0.07	1.0	0.54
		O3′	—	0.04	-3.9	0.55
		O4′	—	0.004	4.7	0.07
		O5′	—	0.04	-4.3	0.61
Phosp.	G	O1P	-43.5	4.19	-8.4	1.91
		O2P	-48.1	4.99	-10.5	2.78
	C	O1P	-22.8	4.19	-7.0	1.91
		O2P	-28.2	4.99	-6.6	2.78

5.4 Effects on the Mg^{2+}-RNA binding affinity

We then repeated the calculations in several different conditions with the aim of dissecting important contributions to RNA-Mg^{2+} binding affinity. Simulations with explicit K^+ ions were compared with equivalent ones using a uniform positive background (UB+) in order to account for ion competition effects. In the same spirit, simulations with flexible RNA were compared

against equivalent ones with rigid RNA, in order to account for flexibility and conformational entropy effects. These simulations are performed with a box containing approximately 2100 water molecules, which is large enough to observe a clearly flat free-energy profile as a function of the distance of the Mg^{2+} from the CoM_{RNA}. Profiles obtained with control simulations are shown in Figure 5.3. The hybridization effects on Mg^{2+} affinity can be clarified by comparing affinities obtained by simulations done on single stranded (ssRNA) and double stranded RNA (dsRNA). The conditions for each set of simulation were previously discussed in detail, and can be seen in Section 5.2.

5.4.1 Ion competition

We first use our simulations to quantify how much the competition with K^+ influences the RNA affinity for Mg^{2+}. To this aim we compared the affinities on individual binding sites using either an uniform positive background (UB+) or explicit K^+ ions. To avoid sampling complications related to the interaction of counterions and coions, we only added 4 monovalent cations so as to neutralize the system. Inner sphere binding free-energies reported in Figure 5.6 A shows the effect of explicit K^+ ions on Mg^{2+}-RNA affinity. Here it is possible to appreciate that competition of K^+ ions decreases the overall Mg^{2+} affinity, both when RNA is kept rigid and when it is modeled as flexible. The change in the total binding free energy is quantified as $10.5\ ^{+1.0}_{-0.7}$ kJ/mol for flexible RNA and as $9.5\ ^{+0.7}_{-0.6}$ kJ/mol for rigid RNA. The effect of the presence of K^+ ions on the affinity can be also rationalized by measuring the conditional Mg^{2+} affinity on individual sites when there is a K^+ ion in proximity of the same site. Results are reported in Figure 5.5 and are consistent with the fact that the decreased affinity is an effect of the competition between the two species for the same binding site. We notice that the effect of competition is local and propagates in a few case to the nearest neighbor binding sites, suggesting that also a short model duplex can be used to quantify ion competition.

The binding free energies for the indirectly bound ions are also reported (Figure 5.6C) and follow a similar trend being reduced by $4.9\ ^{+0.03}_{-0.03}$ kJ/mol for flexible RNA and $9.3\ ^{+0.06}_{-0.06}$ kJ/mol for rigid RNA. Errors are much smaller here since the number of binding and unbinding events is significantly larger in the case of outer-sphere binding.

Equivalent data obtained including MaxEnt corrections are shown in Appendix A.2. The changes in binding free energies due to flexibility are are not affected by the corrections on the Mg^{2+} parametrization, indicating that these results are robust with respect to the imbalance between binding on nucleobases and phosphates observed in the original force field.

5.4.2 RNA flexibility

We then compared the results obtained with a rigid RNA molecule with those obtained with a flexible one. The flexible RNA molecule had minor restraints so as to conserve its secondary structure, but still could undergo significant local deformations. Flexibility effects on the affinity of the inner sphere bound ions are reported in Figure 5.6B. The effect of flexibility is not trivial. In the system where cations are explicitly included, the affinity of Mg^{2+} on flexible RNA is decreased by 3.9 $^{+0.9}_{-0.7}$ kJ/mol with respect to rigid RNA. In the system where cations are replaced with a UB+ the affinity of Mg^{2+} on flexible RNA with respect to rigid is decreased by 2.9 $^{+0.8}_{-0.6}$ kJ/mol.

The values of affinity for the indirectly bound ions (Figure 5.6D) follow the same direction, decreasing by 4.3 $^{+0.05}_{-0.05}$ kJ/mol in the simulation with explicit K^+ ions and 8.7 $^{+0.04}_{-0.04}$ kJ/mol to the one with a UB+.

Also these values are barely affected by the corrections on the Mg^{2+} parametrization, indicating that these results are robust with respect to the imbalance between the binding affinity of nucleobases and phosphates observed in the original force field. Equivalent data obtained including MaxEnt corrections are shown in Table A.2.

By dissecting the contribution of the individual binding sites to the overall affinity, it can be seen that the central phosphate of the guanine (G-P), which contributes most to the overall affinity, has a greater affinity for Mg^{2+} when RNA is kept frozen. This is true for all the three G-Ps.

Interestingly, the affinity on the nucleobase binding sites located in major groove (G-O6 and G-N7) is affected by flexibility with an opposite trend. The lower affinity in the ideal rigid structure indicates that the duplex should undergo slight rearrangements so as to bind Mg^{2+} on the major groove. Figures 5.7 and 5.8 show the conditional probability distributions of all RNA backbone dihedrals consequent to Mg^{2+} direct binding for all phosphates. No significant rearrangement can be appreciated on the backbone dihedrals. We notice however that a very small repositioning of the phosphates could lead to a significant change in the electrostatic interaction with Mg^{2+} that would

explain the observed differences. Therefore, the structural integrity of the duplex was maintained even when Mg^{2+} was directly bound to RNA. It is also relevant to say that for ≈1% of the simulation time flexible RNA underwent a reversible transitions to ladder-like structures [186]. This is consistent with what has been observed in recent simulations of restrained RNA duplexes [187]. Reversibility was checked by monitoring the continuous trajectories so as to avoid false transitions to be observed just due to replica exchanges. To assess the impact of these structures on Mg^{2+} binding, we recomputed all the affinities by excluding all the snapshots where at least one of the glycosidic torsions was in the range $(-90°, 0°)$, which corresponds to the *high anti* conformation observed in ladder-like structures. Only the affinities for O5′ at the 5′ termini were affected. All the other affinities where within the statistical error from the calculation including all the data, indicating that transitions to ladder-like structures is not correlated with Mg^{2+} binding. Since only the affinities for non-terminal nucleotides are discussed in this work, the presence of ladder-like structure does not affect the reported results.

5.4.3 RNA hybridization

Finally, we compare the affinity of Mg^{2+} with a double stranded RNA (dsRNA) against the one with a pair of single stranded RNAs (ssRNA) with the same sequence. The calculations were performed for both systems in a box that was large enough to contain the two separated strands and in identical ionic conditions. Also these simulations were performed using neutralizing cations only, so that the affinities reported for the duplex presented in this Section corresponded to a lower ionic strength in comparison with those presented above. In all these simulations, RNA was kept rigid. Indeed, sampling all the conformations available for a ssRNA is a formidable task [132, 188] and would have made virtually impossible to obtain converged values for the binding affinities. Moreover, the capability of current force fields to correctly reproduce the conformational ensembles of ssRNA has been questioned [101, 152, 187, 189]. To allow for the affinities to be comparable, it was necessary to treat also the dsRNA as rigid. Affinities for inner sphere binding are reported in Figure 5.9A. Overall the affinity in the dsRNA was larger indicating that hybridization and Mg^{2+} binding act cooperatively. In other words, when a Mg^{2+} ion is interacting with RNA, the hybridization free energy is expected to be decreased by $9^{+1.4}_{-0.9}$ kJ/mol, further stabilizing the duplex. We also notice that the overall affinity on the dsRNA is dominated by direct interactions

with the phosphate. However, it is interesting to see this effect on individual binding sites. In Figure 5.9 it can be seen that the affinity with the G-O6 is affected by hybridization in the opposite manner, so that affinity in the ssRNA is enhanced. This is consistent with the fact that electronegative atoms in the base are more accessible to divalent ions. However, since the contribution of bases to the overall affinity on the dsRNA is negligible with respect the contribution of phosphates, this effect is not visible in the overall affinity.

It is also possible to compare the affinity of Mg^{2+} ions which are directly bound with that of ions that are indirectly bound. As it can be seen in Figure 5.9B, hybridization increases the stability of indirect binding sites as well, by $9.7\ ^{+0.1}_{-0.1}$ kJ/mol.

5.5 Conclusion

In this chapter it was present extensive MD simulations investigating the binding affinity of Mg^{2+} on all the possible binding sites in a short RNA duplex. Calculations are performed using state-of-the-art force fields and a modified version of bias-exchange metadynamics.

The enhanced sampling method we used is an improved version of the bias-exchange metadynamics (BE-MetaD) procedure designed for this application. At variance with the original approach, we here added different penalizing potentials in each replica to avoid binding events that would trap Mg^{2+} in undesired positions. This procedure was necessary here to achieve converged affinities. The idea of forbidding different events in different replicas can be straightforwardly generalized to cases where one wants to study competing rare events under the same conditions, and could provide a significant improvement in the applicability of BE-MetaD to the description of complex free-energy landscapes. To help reproducibility of our results and application of the procedure to different systems, we included a sample input file in Figure A.1.

Overall, this procedure provides statistically converged affinities for the modeled systems. To further check convergence, we applied the same protocol on a symmetric duplex. The calculated affinities are shown in Table 5.4). This check would show if the reported statistical error were underestimated, since a difference in the affinity between the sides of a perfectly symmetric duplex could only arise from statistical uncertainty.

Our results rely on a few methodological choices and approximations that we here discuss in detail. It is known that the current RNA force fields may

not properly describe unstructured single-stranded oligonucleotides, as discussed previously on Chapter 3.2.1. [101, 187, 189] In this work we focused in the investigation on dsRNAs and on rigid ssRNA. With respect to the Mg^{2+} ion itself we assessed our chosen force field by testing the effect of an *a posteriori* adjustment of its interactions with RNA so as to enforce binding affinities to be in agreement with potentiometric titration experiments on nucleosides. Similarly to a recently published parametrization, proposed by [181] this procedure did not affect the ion-water and ion-ion interactions. However, the qualitative consistency between the results with and without the corrections indicates robustness with respect to Mg^{2+} force field choice. Another point to consider is that in our simulations we assumed a single Mg^{2+} ion binding to RNA at any time, implying an infinite Mg^{2+} dilution. To verify the effect of the neglecting extra Mg^{2+} in the buffer we performed an extra control simulation including an appropriate $MgCl_2$ buffer. Our results on the competition between K^+ and Mg^{2+} suggest that it would be very difficult for multiple Mg^{2+} ions to bind on the same site. However, the double charge of Mg^{2+} could allow for longer range interactions, affecting the affinity of the neighboring binding sites. One in principle should thus simulate a replica corresponding to each pair of potentially cooperative binding sites. Ideally, this could be done after an initial screening where the most important binding sites have been identified. Additionally, since some simulation boxes only had enough ions to counterbalance the negative charge of RNA backbone, some of our results did not consider the effect of anions. To verify the effect of the neglecting Cl^- in the buffer we performed an extra control simulation including an appropriate KCl buffer. We notice that in all these simulations the number of ions rather than their chemical potential is kept constant. This limitation could be overcome using a very large simulation box and an extra potential to control ionic strength in the spirit of ref. [190]. Additionally, the reported control simulations performed with a large box allow the possible artifacts related to box size to be assessed. Here the Mg^{2+} binding affinity might be affected by a larger effective cation concentration in the vicinity of the RNA (see Figure 5.3). However, our results show that relative binding affinities are virtually independent on this effect. Finally the reported results were obtained using a single RNA sequence in a A-form helix. Sequence and structure dependent effects will be the subject of a further investigation. This will be further discussed in later Chapter, in which a similar methodology used here, is applied to more complex system than a duplex.

Our results show that the overall affinity of the inner sphere (direct) bound

Mg^{2+} cations on a RNA duplex is largely dominated by the interaction with phosphates. On the contrary, outer sphere (indirect) bound Mg^{2+} cations interacts more strongly with the nucleobases. We observe that there is a significant preference for inner Mg^{2+} binding on the guanines with respect to cytosines. Interestingly, this is consistent in all our simulations including the ones performed with two separated single strands in the A-helix. This suggests that indirect contacts with guanine N7 might provide extra stabilization. Additionally, we see an overall preference between the three moieties of a nucleotide in the following order: phosphate > bases > sugar. The affinities on specific binding atoms follows the trend O2P > O1P > G-O6 > G-N7 > sugar hydroxyls > C-O2. This trend is not the same when considering the outer sphere contacts, where it is changed to G-N7 ≈ G-O6 ≈ O2P > O1P > sugar hydroxyls. Our procedure captures the experimental trends observed in the PDB binding frequencies both for inner and outer sphere binding. It must be noticed that the PDB survey reported in ref. [70] considers binding with a variety of RNA motifs. However, the most common RNA motif present in the PDB is the A-helix, which is the same motif addressed by our study. Although Mg^{2+} is expected to mostly bind on specific structures and to stabilize tertiary contacts, the comparison of our study with the discussed PDB analysis indirectly confirms that the patterns of the electrostatic field in the neighborhood of a helix are representative for the general trend observed in structured RNAs. It must be also observed that the interpretation of primary X-ray data is not trivial and the assignment of many of the reported density peaks to Mg^{2+} ions have been recently challenged. [24] However, whereas these errors could affect the interpretation of specific important structures, we expect the overall statistics to be reliable. Moreover, we notice that, although the ranking are correctly reproduced by our calculations, the reported frequencies are not proportional to $e^{-\frac{\Delta G^i_{Mg^{2+}}}{k_B T}}$. It is not clear whether the frequencies from the PDB can be assumed as representative of a Boltzmann ensemble. This discrepancy could also be related to an imbalance inherent to the force field in the description of interactions of Mg^{2+} with phosphates and bases which has also been reported in refs. [89, 181]. Ions in MD simulations are usually described by charged Van der Waals spheres. Although this model has proven to be very useful, its accuracy is still debated. The main source of doubt comes from the fact that usual MD does not explicitly includes polarizations effects.

Our procedure also allowed for the dissection of the effect of ion competition, RNA flexibility and RNA hybridization to Mg^{2+} affinity. We found that,

for both inner and outer sphere binding, ion competition and RNA flexibility reduce Mg^{2+} overall binding affinity while hybridization increases it.

The effect of ion competition on the inner sphere binding was independent of solute flexibility and amounted to ≈ 10 kJ/mol. The same was true for outer sphere binding on a rigid RNA. However, the effect for outer sphere binding on flexible RNA was significantly smaller (≈ 4 kJ/mol). This indicates that local rearrangements that are possible in the flexible RNA could compensate for the repulsion between the cations.

Interestingly, RNA flexibility decreased its affinity for Mg^{2+}. We recall that the total affinity is dominated by the phosphates. Affinity on the nucleobases was on the other hand increased by flexibility. We argue that the flexibility of RNA might affect binding affinity in two opposite ways. First, the enthalpic contribution to the affinity could be increased by RNA flexibility when local rearrangements lead to more favorable RNA-Mg^{2+} interactions. On the other hand, binding of RNA with Mg^{2+} constrains RNA leading to a loss in its conformational entropy. [145] Interestingly, it has been recently suggested that multivalent cations make RNA helices more rigid. [191] The simulation of rigid RNA allowed us to explicitly ignore changes in the RNA conformational entropy. We observe that nucleobases are significantly constrained by Watson-Crick pairing and require a local rearrangement so as to bind Mg^{2+}. Conversely, phosphates are accessible for Mg^{2+} binding even in a rigid RNA model. We hypothesize that in the case of Mg^{2+} binding on phosphates the second effect dominates over the first effect leading to a decreased affinity in the flexible model.

During the revision process for publication of this work (see ref [26]), an alternative approach was proposed to find Mg^{2+} binding sites with large affinity using a grand-canonical Monte Carlo scheme. [192] Since the approach presented here addresses the same problem in an orthogonal direction, the two schemes might be combined so as to allow for an even more efficient simulation protocol.

In this Chapter it was presented a computational approach to the detailed and atomistic characterization of Mg^{2+}-RNA binding. To this aim, we introduced a modified version of bias-exchange metadynamics. The results reproduce statistics observed in structural databases and allow for a dissection of the most important contribution to Mg^{2+}-RNA interactions, shading a new light on the interplay between RNA flexibility and binding with divalent cations. We foresee the application of our computational approach to the characterization of Mg^{2+} binding sites on a repertoire of RNA motifs and

sequences. Although there is still controversy regarding Mg^{2+} ion parameters and the accuracy of the simple Lennard-Jones based models employed, the trends in the binding affinities would likely be consistent. Additionally, the introduced procedure could be used as a benchmark to compare several models. More generally, our procedure could be used to trustfully quantify affinities of ions or small ligands when multiple competing binding sites have to be simultaneously assessed.

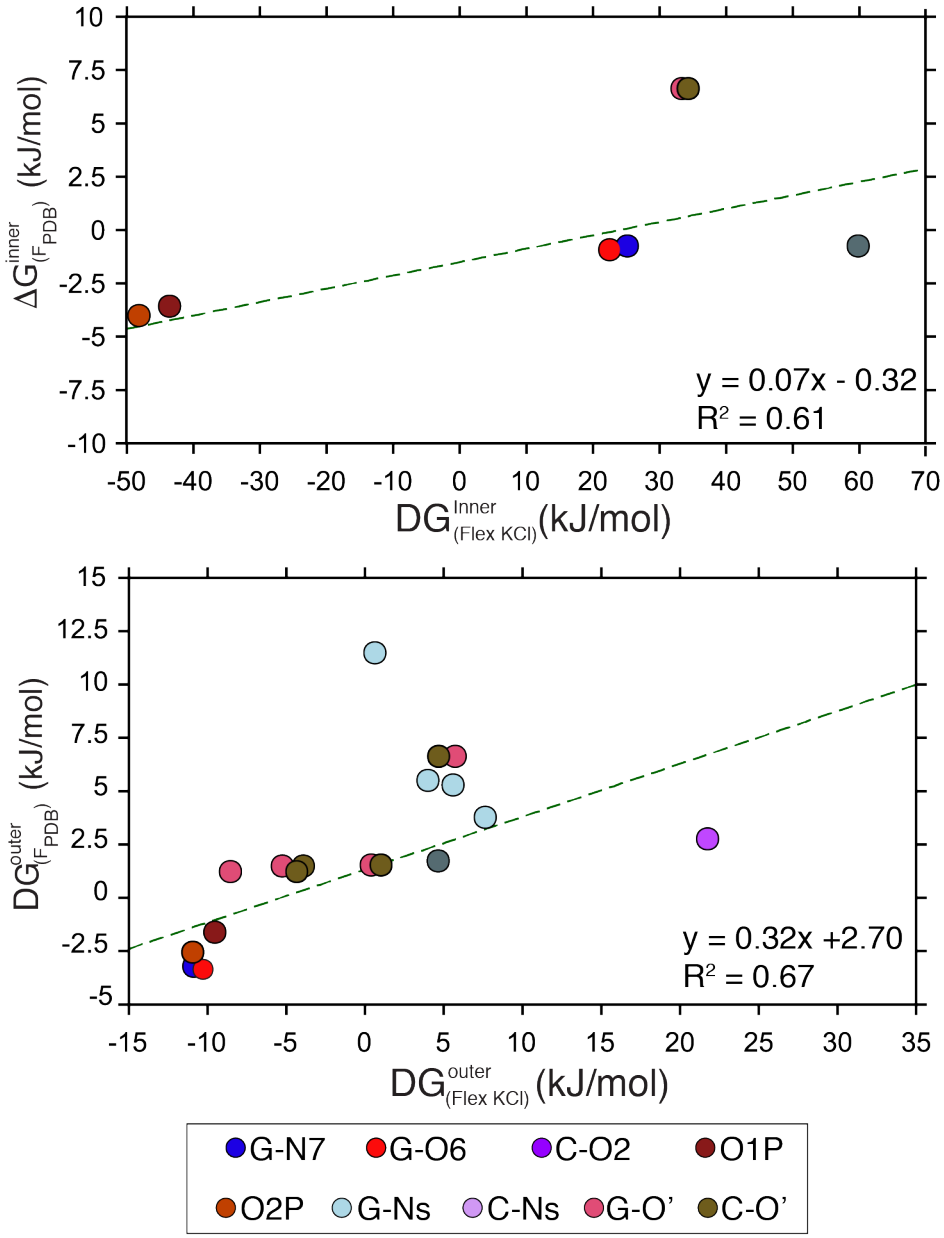


FIGURE 5.4: Mg^{2+} affinities against the estimated free energies from the PDB interaction frequencies. $\Delta G^i_{F_{PDB}} = -k_B T \log F^i_{PDB}$, being F^i_{PDB} the frequency of Mg^{2+} interaction measured on the PDB, with i indicating innner or outer shell binding. The dotted line is obtained through linear regression of the data. We notice that the reported $\Delta G^i_{F_{PDB}}$ are subject to an arbitrary shift due to the normalization of the corresponding frequencies, so that only the slope of this line can be interpreted. The guanine nitrogens (N1, N2, N3 and N9) are labeled as G-Ns and the cytosine oxygens as C-O2. The sugar oxygens (O2′, O3′, 04′ and O5′) are labeled as G-O′ for guanine and C-O′ for cytosine. Error-bars are not shown, but the error estimation for each of the affinities plotted here can be found in the table A.1.

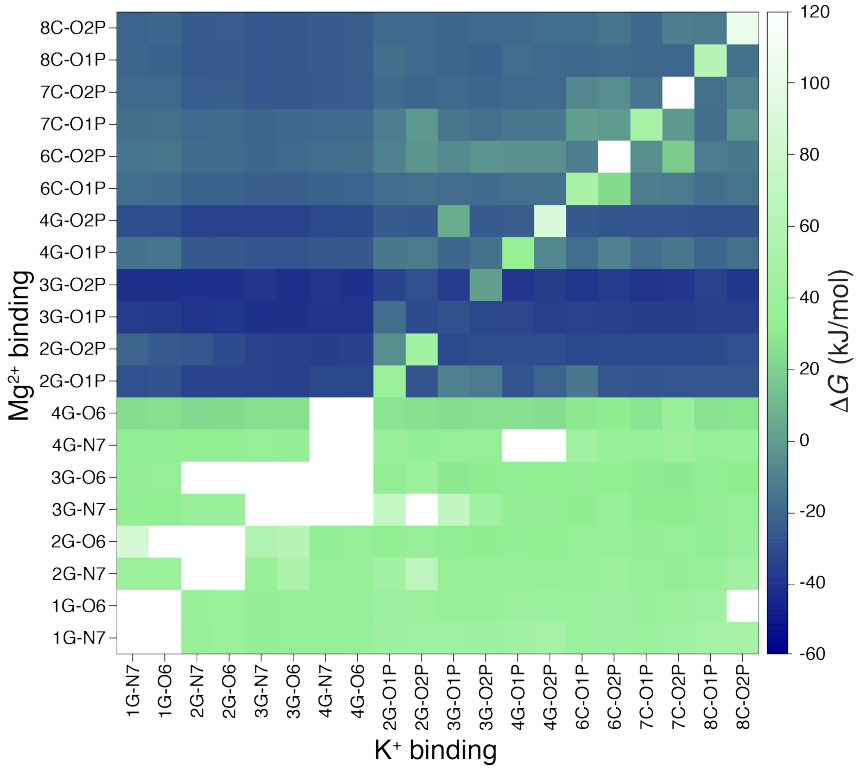


FIGURE 5.5: Conditional Mg^{2+} affinities upon K^+ binding. Each square represents in a color scale the affinity of Mg^{2+} on a specific binding site (vertical axis) when a ịs close to another binding site (horizontal axis). We notice that binding sites that are spatially close (e.g. sites in the same nucleobase, G-O6 and G-N7 from consecutive nucleotides, and some of the neighboring phosphates) are affected by K^+ binding. We also notice that Mg^{2+} affinity is accumulated considering only simulation snapshots where K^+ was bound on a specific site. Since some of the sites are rarely occupied by K^+, statistical errors for those sites are larger with respect to the reported ones. The affinities shown in this figure were calculated from the simulation on a large box with a buffer of KCl at 0.1 M concentration. The same graph was obtained for the simulations used to analyze the effects of ion competition on the affinity. The pattern of the conditional Mg^{2+} affinities upon K^+ binding remained qualitatively identical to the one shown here.

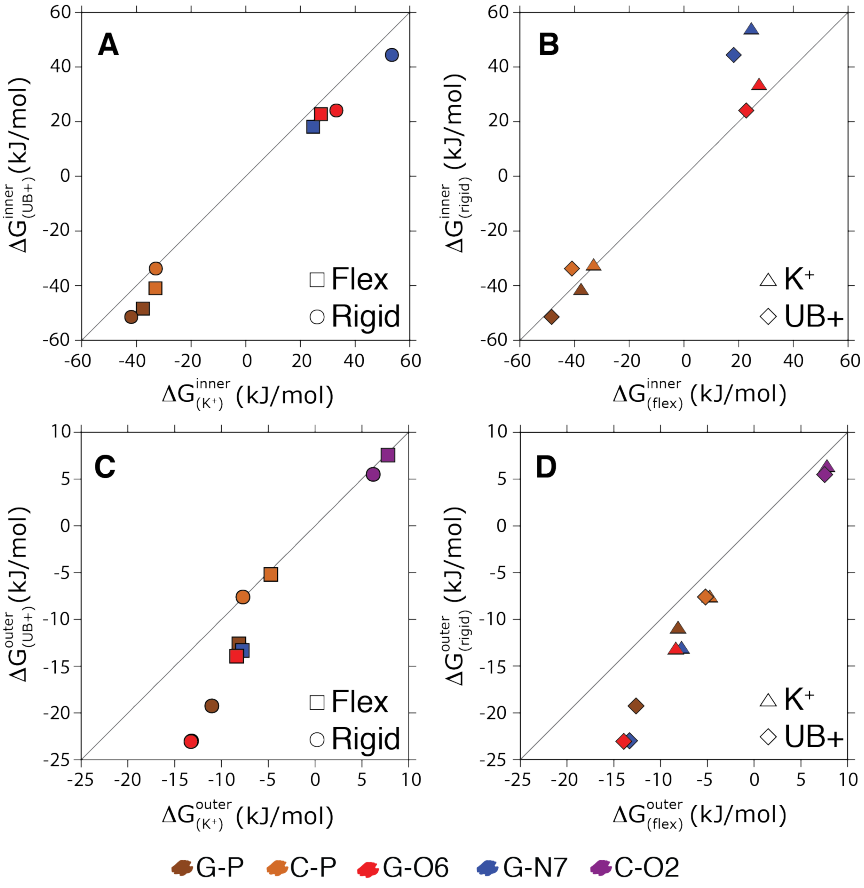


FIGURE 5.6: Specific Mg^{2+} binding affinities on a $\frac{GGGG}{CCCC}$ duplex under different simulation conditions. The affinities were obtained in a flexible and rigid duplex both with explicit K^+ ions and without, thus with a uniform positive background (UB+). Plots A and C show the effect of ion competition (K^+ *vs* UB+) for inner and outer sphere Mg^{2+} binding respectively. Plots B and D show the effect of flexibility (flex *vs* rigid) for inner and outer sphere Mg^{2+} binding respectively.

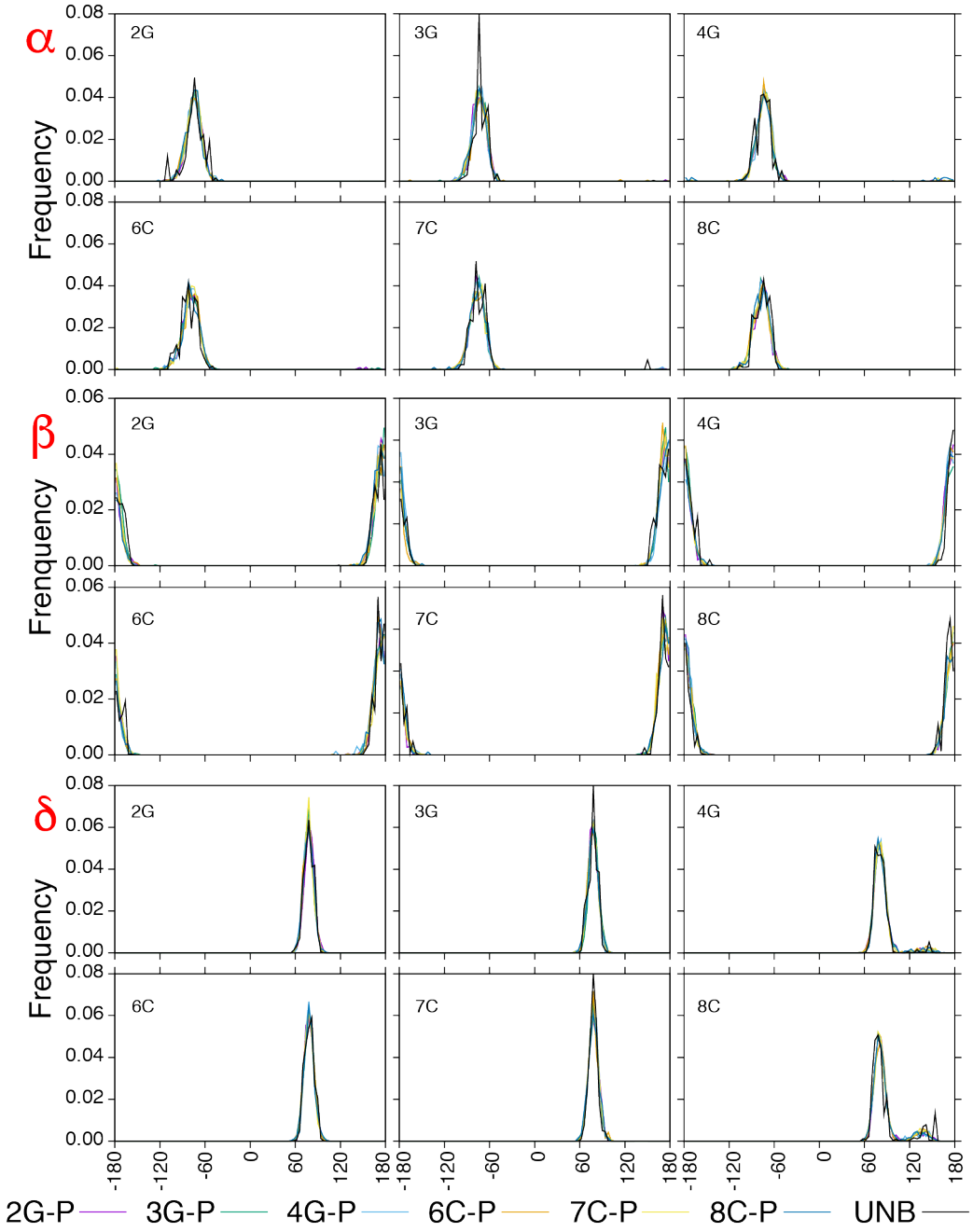


FIGURE 5.7: Conditional distributions for α, β and δ backbone dihedral angles upon Mg^{2+} binding on the phosphates. Each panel shows the distribution of the dihedral from the indicated residue conditioned to Mg^{2+} being bound on a different phosphate group (different colors). The black line indicates the distribution when Mg^{2+} is not bound.

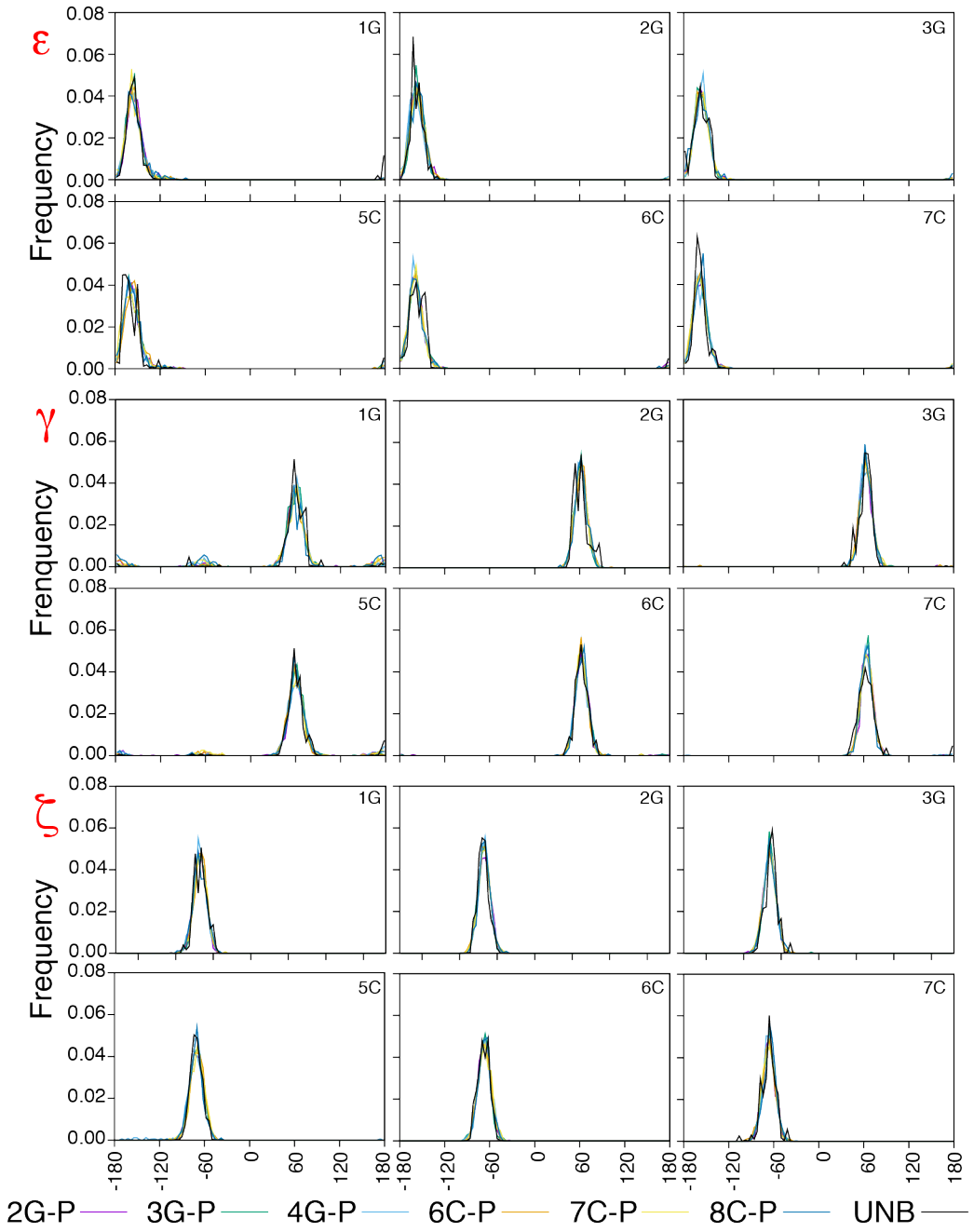
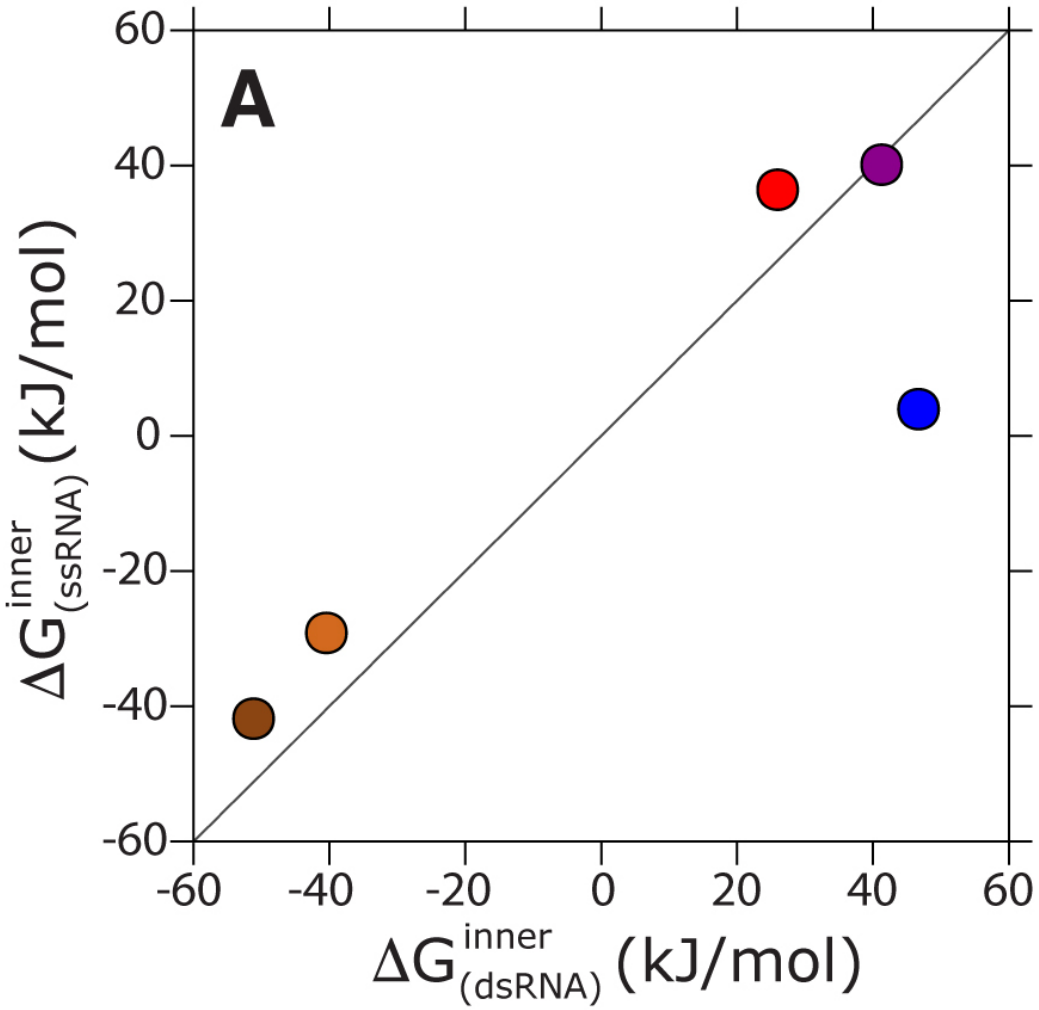


FIGURE 5.8: Conditional distributions for ε, γ and ζ backbone dihedral angles upon Mg^{2+} binding on the phosphates. Each panel shows the distribution of the dihedral from the indicated residue conditioned to Mg^{2+} being bound on a different phosphate group (different colors). The black line indicates the distribution when Mg^{2+} is not bound.


A
60
40
20
0
-20
-40
-60
-60
-40
-20
0
20
40
60
$\Delta G^{inner}_{(ssRNA)}$ (kJ/mol)
$\Delta G^{inner}_{(dsRNA)}$ (kJ/mol)

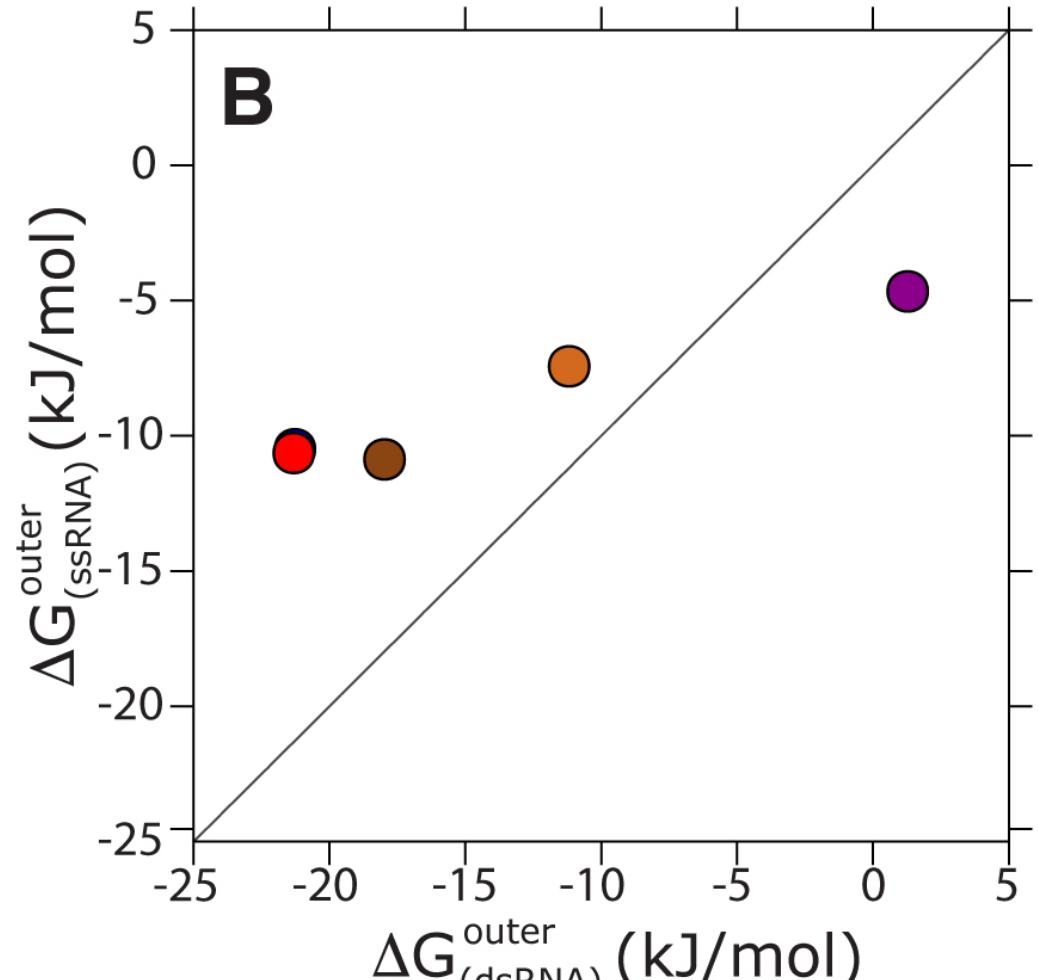

B
5
0
-5
-10
-15
-20
-25
-25
-20
-15
-10
-5
0
5
$\Delta G^{outer}_{(ssRNA)}$ (kJ/mol)
$\Delta G^{outer}_{(dsRNA)}$ (kJ/mol)

TABLE 5.4: Affinities on a $\frac{GGCC}{CCGG}$ duplex. The difference on the affinities of the same binding site on the 3′-5′ side with respect to the 5′-3′ is only due to statistical errors, since the duplex is completely symmetrical. Here we report errors and affinities in kJ/mol and only for binding sites for the central GC pairs. Notice that the affinities are within error in the large majority of cases, confirming that the calculation provides good statistical convergence. The errors on the K_a on the $i-th$ binding site (ϵ_i) were calculated over 4 blocks being $\epsilon_i = \frac{\sigma_i}{\sqrt{n-1}}$. σ_i is the standard deviation of the K_a calculated for each block n. The affinities here are show as $\Delta G_i \, {}^{+\Delta_2}_{-\Delta_1}$, in which $\Delta_1 = -k_B T \log\left(K_a^i - \epsilon_i\right) - \Delta G_i$ and $\Delta_2 = -k_B T \log\left(K_a^i + \epsilon_i\right) - \Delta G_i$.

.

Binding Sites			$\Delta G^{inner}_{nocorr}$ (kJ/mol)		ΔG^{inner}_{corr} (kJ/mol)		ΔG^{outer} (kJ/mol)	
			3′ - 5′	5′ - 3′	3′ - 5′	5′ - 3′	3′ - 5′	5′ - 3′
Bases	G	N1	—	—	—	—	$22.9\,{}^{2.9}_{-1.3}$	$32.9\,{}^{7.7}_{-1.7}$
		N2	—	—	—	—	$1.1\,{}^{0.3}_{-0.2}$	$1.2\,{}^{0.3}_{-0.3}$
		N3	—	—	—	—	$6.7\,{}^{0.1}_{-0.1}$	$6.7\,{}^{0.2}_{-0.2}$
		N7	$37.2\,{}^{0.3}_{-0.3}$	$41.3\,{}^{0.5}_{-0.4}$	$25.0\,{}^{0.4}_{-0.3}$	$28.9\,{}^{0.5}_{-0.4}$	$-6.6\,{}^{0.2}_{-0.1}$	$-6.7\,{}^{0.1}_{-0.1}$
		N9	—	—	—	—	$14.8\,{}^{0.7}_{-0.6}$	$15.9\,{}^{1.6}_{-1.0}$
		O6	$22.5\,{}^{0.4}_{-0.3}$	$21.1\,{}^{0.6}_{-0.5}$	$8.5\,{}^{0.3}_{-0.3}$	$6.9\,{}^{0.6}_{-0.5}$	$-7.9\,{}^{0.1}_{-0.1}$	$-8.0\,{}^{0.1}_{-0.1}$
	C	N1	—	—	—	—	$15.8\,{}^{0.9}_{-0.7}$	$20.4\,{}^{3.3}_{-1.4}$
		N3	—	—	—	—	$12.3\,{}^{1.5}_{-0.9}$	$12.1\,{}^{0.3}_{-0.3}$
		N4	—	—	—	—	$-7.4\,{}^{0.1}_{-0.1}$	$-7.4\,{}^{0.1}_{-0.1}$
		O2	$59.3\,{}^{0.5}_{-0.4}$	$58.7\,{}^{0.3}_{-0.3}$	$54.7\,{}^{0.5}_{-0.4}$	$54.1\,{}^{0.4}_{-0.3}$	$0.7\,{}^{0.3}_{-0.3}$	$0.7\,{}^{0.3}_{-0.3}$
Sugar	G	O2′	—	$33.5\,{}^{5.0}_{-1.6}$	—	$33.5\,{}^{5.0}_{-1.6}$	$1.4\,{}^{0.3}_{-0.3}$	$1.5\,{}^{0.2}_{-0.2}$
		O3′	—	—	—	—	$-4.0\,{}^{0.1}_{-0.1}$	$-4.0\,{}^{0.1}_{-0.1}$
		O4′	—	—	—	—	$6.1\,{}^{0.1}_{-0.1}$	$6.1\,{}^{0.3}_{-0.3}$
		O5′	—	—	—	—	$-4.4\,{}^{0.1}_{-0.1}$	$-4.4\,{}^{0.1}_{-0.1}$
	C	O2′	$32.1\,{}^{1.9}_{-1.1}$	$34.6\,{}^{2.9}_{-1.3}$	$32.1\,{}^{1.9}_{-1.1}$	$34.5\,{}^{2.9}_{-1.3}$	$-0.5\,{}^{0.2}_{-0.1}$	$-0.6\,{}^{0.2}_{-0.2}$
		O3′	—	—	—	—	$-4.1\,{}^{0.1}_{-0.1}$	$-3.8\,{}^{0.1}_{-0.1}$
		O4′	—	—	—	—	$4.3\,{}^{0.2}_{-0.2}$	$4.3\,{}^{0.2}_{-0.2}$
		O5′	—	—	—	—	$-4.7\,{}^{0.1}_{-0.1}$	$-4.3\,{}^{0.1}_{-0.1}$
Phosp.	C	O1P	$-31.8\,{}^{1.3}_{-0.9}$	$-32.9\,{}^{1.3}_{-0.9}$	$-8.1\,{}^{1.3}_{-0.8}$	$-9.0\,{}^{1.3}_{-0.8}$	$-6.5\,{}^{0.1}_{-0.1}$	$-6.3\,{}^{0.1}_{-0.1}$
		O2P	$-34.8\,{}^{1.5}_{-0.9}$	$-32.9\,{}^{1.8}_{-1.0}$	$-10.4\,{}^{1.5}_{-0.9}$	$-8.4\,{}^{1.7}_{-1.0}$	$-6.0\,{}^{0.1}_{-0.1}$	$-5.9\,{}^{0.1}_{-0.1}$

Chapter 6

Conclusions and Perspectives

In this Thesis, state-of-the-art simulations techniques for rare events were applied to sampling Mg^{2+} binding to RNA with molecular dynamics and assessing current force field accuracy. Results indicate that current Mg^{2+} force fields do not reproduce well experimental values of the Mg^{2+}-phosphodiester binding free energy and the measured affinities are sensitive to water models utilized and different versions of the RNA force field. Furthermore, we proposed an innovative combination of enhanced sampling techniques to sample many Mg^{2+} ions potentially binding to multiple RNA is an expensive and complex computational task. Simulations performed to investigate Mg^{2+} binding to a four-base-pairs RNA duplex thoroughly, indicate that the relative affinity between the phosphodiester moieties and the RNA bases are not affected by the choice of Mg^{2+} force field. Also, the obtained affinities qualitatively agree with experimental distributions of Mg^{2+} ions bound to RNA obtained by an analysis of structures deposited on the PDB. This indicates that, although not able to give thermodynamic stabilities in quantitative agreement with experiments, these simulations can be used to reliably predict Mg^{2+} binding sites. Lastly, effects of solute flexibility, ion competition and RNA hybridization on the Mg^{2+}-RNA affinity were also quantified, indicating that ion competition and RNA hybridization decreases and increases the RNA overall affinity, while flexibility affects affinity in a non-trivial way due to entropic effects.

The methodologies and results proposed here could be applied to describing Mg^{2+} affinities to different RNA structures or Mg^{2+} dependent biological problems. Relative affinities for all the possible Mg^{2+} binding positions of known metal-binding structures such as the GNRA tetraloops and S-turns could be directly compared with available NMR data [164, 193]. In addition, one might apply the approaches proposed in this thesis to understand better why Mg^{2+} ions interact differently with RNA and DNA [42, 171], or

even to possibly describing the Mg^{2+} dependency of some biological processes, such as gene regulation performed by riboswitches or the intro-exon recognition site binding in ribozymes [156, 194, 195]. Some work in those directions are already in progress, although it not presented here. Furthermore, the enhanced sampling approaches proposed in this work could be applied to a different ligand-binding pair where multiple concurrent and relatively equivalent binding sites need to be explored. The next step could be towards understanding how Mg^{2+} ions affect the free-energy landscape of RNA folding for small molecules, for which it might be usefulG to integrate state-of-the-art enhanced sampling techniques used to assess RNA folding landscape [196] and the methods proposed here.

Appendix A

Appendix 1

```
# Start of cvs.dat file

# insert inside {} the correspondent index number of the described atoms.
# @{atom type}-{#residue}
MOLINFO STRUCTURE=my-initial-configuration.pdb MOLTYPE=rna

# Essential when calculating CoM
WHOLEMOLECULES STRIDE=10 ENTITY0={first RNA atom}-{last RNA atom}

## Assign atoms to group names.

# Chooses only the the water oxygens for coordination number with water (CN_W)
OW:      GROUP ATOMS={first water oxygen}-{last water oxygen}:3

Mg:      GROUP ATOMS={Mg}
rna:     GROUP ATOMS={first RNA atom}-{last RNA atom}
P:       GROUP ATOMS=@P-2
OP:      GROUP ATOMS=@O1P-2,O2P-2
N31:     GROUP ATOMS=@N3-1
N32:     GROUP ATOMS=@N3-2
N71:     GROUP ATOMS=@N7-1
N72:     GROUP ATOMS=@N7-2
O2s1:    GROUP ATOMS=@O2'-1
O2s2:    GROUP ATOMS=@O2'-2
O3s2:    GROUP ATOMS=@O3'-2
O4s1:    GROUP ATOMS=@O4'-1
O4s2:    GROUP ATOMS=@O4'-2
O5s1:    GROUP ATOMS=@O5'-1
O61:     GROUP ATOMS=@O6-1
O62:     GROUP ATOMS=@O6-2

# Setting collective variables (Distance and coordination number with water oxygens)
# Here one could add the distance between the divalent cation and a putative binding
# position in which enhancing sampling is needed.

dP:     DISTANCE ATOMS=Mg,P
dN71:   DISTANCE ATOMS=Mg,N71
dN72:   DISTANCE ATOMS=Mg,N72
dO61:   DISTANCE ATOMS=Mg,O61
dO62:   DISTANCE ATOMS=Mg,O62

cn:     COORDINATION GROUPA=Mg GROUPB=OW R_0=0.261 NLIST NL_CUTOFF=0.6 NL_STRIDE=100

rnacom: COM ATOMS=rna

#End of cvs.dat file

### ========================================================================== ###
# One will need as many plumed.dat files as replicas to simulate
# For each plumed file there will be one MetaD acting on the distance between the
# Mg2+ and the target binding site and the coordination number of Mg2+ with water.
# Notice how the lines are commented, the penalty potentials act on all the binding
# positions but the one enhanced on this replica.
# The same pattern will be repeated for all the other replicas.
### ========================================================================== ###

RESTART                  # Necessary when one is performing the simulations by parts.
RANDOM_EXCHANGES         # Helps achieve better ergodicity.
INCLUDE FILE=cvs.dat     # Feeds the file with the declared CV to all the replicas.


# Well tempered metadynamics

metadP:    METAD ARG=dP,cn   SIGMA=0.05,0.1 HEIGHT=0.3 PACE=500 TEMP=300 BIASFACTOR=15 FILE=HILLS_dP
#metadN71: METAD ARG=dN71,cn SIGMA=0.05,0.1 HEIGHT=0.3 PACE=500 TEMP=300 BIASFACTOR=15 FILE=HILLS_dN71
#metadN72: METAD ARG=dN72,cn SIGMA=0.05,0.1 HEIGHT=0.3 PACE=500 TEMP=300 BIASFACTOR=15 FILE=HILLS_dN72
#metadO61: METAD ARG=dO61,cn SIGMA=0.05,0.1 HEIGHT=0.3 PACE=500 TEMP=300 BIASFACTOR=15 FILE=HILLS_dO61
#metadO62: METAD ARG=dO62,cn SIGMA=0.05,0.1 HEIGHT=0.3 PACE=500 TEMP=300 BIASFACTOR=15 FILE=HILLS_dO62

# One can improve the computational performance of Plumed by using grids when calculating and storing
# the MetaD potentials. For more information refer to Plumed manual.

# Penalty potentials on the binding positions which are not biased on this specific replica.

#lwalldP:    LOWER_WALLS ARG=dP     AT=0.380   EPS=0.1  EXP=1  KAPPA=-500
lwalldN71:  LOWER_WALLS ARG=dN71   AT=0.300   EPS=0.1  EXP=1  KAPPA=-500
lwalldN72:  LOWER_WALLS ARG=dN72   AT=0.300   EPS=0.1  EXP=1  KAPPA=-500
lwalldO61:  LOWER_WALLS ARG=dO61   AT=0.300   EPS=0.1  EXP=1  KAPPA=-500
lwalldO62:  LOWER_WALLS ARG=dO62   AT=0.300   EPS=0.1  EXP=1  KAPPA=-500
```

FIGURE A.1: Sample PLUMED input file for a GpG dinucleoside monophosphate.

TABLE A.1: Reported inner sphere Mg^{2+} affinities calculated for the possible binding sites on a $\frac{GGGG}{CCCC}$ duplex. Columns are displaying affinities and errors in kJ/mol for the following simulations: flexible RNA with explicit K^+ (Flex K^+), flexible RNA with uniform positive background (Flex UB+), rigid RNA with explicit K^+ (rigid K^+), rigid RNA with uniform positive background (rigid UB+), rigid RNA on a large simulation box with explicit K^+ (dsRNA) and two separated rigid single strands (GGGG and CCCC) on the same conditions as the latter (ssRNA).

Binding sites			Flex K^+	Flex UB+	Rigid K^+	Rigid UB+	dsRNA	ssRNA
Bases	G	N1	—	—	—	—	—	—
		N2	—	—	—	—	—	—
		N3	—	—	—	—	—	—
		N7	$27.5\ ^{0.5}_{-0.4}$	$22.8\ ^{0.4}_{-0.4}$	$33.2\ ^{0.7}_{-0.5}$	$24.1\ ^{1.0}_{-0.7}$	$26.0\ ^{0.8}_{-0.6}$	$36.4\ ^{1.1}_{-0.8}$
		N9	—	—	—	—	—	—
		O6	$24.6\ ^{0.4}_{-0.3}$	$18.2\ ^{0.2}_{-0.2}$	$53.5\ ^{0.8}_{-0.6}$	$44.4\ ^{0.7}_{-0.5}$	$46.7\ ^{1.7}_{-1.0}$	$3.9\ ^{0.3}_{-0.2}$
	C	N1	—	—	—	—	—	—
		N3	—	—	—	—	—	$50.3\ ^{2.6}_{-1.2}$
		O2	$61.4\ ^{0.5}_{-0.4}$	$59.5\ ^{0.9}_{-0.6}$	—	—	—	$10.4\ ^{0.5}_{-0.4}$
Sugar	G	O2′	—	$36.0\ ^{1.6}_{-1.0}$	—	$78.3\ ^{inf}_{-1.8}$	—	—
		O3′	—	—	—	—	—	—
		O4′	—	—	—	—	—	—
		O5′	—	—	—	—	—	—
	C	O2′	$37.5\ ^{2.2}_{-1.1}$	$40.4\ ^{8.1}_{-1.7}$	$49.8\ ^{0.5}_{-0.5}$	$47.8\ ^{0.7}_{-0.6}$	$41.3\ ^{4.9}_{-1.5}$	$40.1\ ^{3.5}_{-1.4}$
		O3′	—	—	—	—	—	—
		O4′	—	—	—	—	—	—
		O5′	—	—	—	—	—	—
Phosp.	G	O1P	$-30.7\ ^{inf}_{-1.7}$	$-39.5\ ^{2.8}_{-1.3}$	$-31.1\ ^{0.9}_{-0.7}$	$-28.4\ ^{5.8}_{-1.6}$	$8.6\ ^{inf}_{-1.9}$	$-38.1\ ^{2.7}_{-1.3}$
		O2P	$-37.5\ ^{0.9}_{-0.7}$	$-48.3\ ^{0.8}_{-0.6}$	$-41.8\ ^{0.6}_{-0.5}$	$-51.4\ ^{0.4}_{-0.3}$	$-51.2\ ^{1.1}_{-0.8}$	$-41.2\ ^{1.0}_{-0.7}$
	C	O1P	$-31.6\ ^{0.5}_{-0.4}$	$-38.5\ ^{2.8}_{-1.3}$	$-29.2\ ^{0.9}_{-0.7}$	$-30.2\ ^{2.7}_{-1.3}$	$-36.8\ ^{3.1}_{-1.3}$	$-29.1\ ^{0.3}_{-0.3}$
		O2P	$-31.0\ ^{2.6}_{-1.2}$	$-39.8\ ^{1.7}_{-1.0}$	$-32.3\ ^{1.3}_{-0.9}$	$-33.0\ ^{1.4}_{-0.9}$	$-39.8\ ^{4.5}_{-1.5}$	$-16.1\ ^{inf}_{-1.9}$

TABLE A.2: The reported Mg^{2+} affinities were calculated using a correction to the the force field described in the methods. Columns are displaying affinities and and errors in kJ/mol for the following simulations: flexible RNA with explicit K^+ (Flex K^+), flexible RNA with uniform positive background (Flex UB+), rigid RNA with explicit K^+ (rigid K^+), rigid RNA with uniform positive background (rigid-UB+), rigid RNA on a large simulation box with explicit K^+ (dsRNA) and two separated rigid single strands (GGGG and CCCC) on the same conditions as the latter (ssRNA).

Binding sites			Flex K^+	Flex UB+	Rigid K^+	Rigid UB+	dsRNA	ssRNA
Bases	G	N1	—	—	—	—	—	—
		N2	—	—	—	—	—	—
		N3	—	—	—	—	—	—
		N7	14.8 $^{0.5}_{-0.4}$	10.3 $^{0.3}_{-0.3}$	21.2 $^{0.6}_{-0.5}$	12.4 $^{0.8}_{-0.6}$	13.8 $^{0.6}_{-0.5}$	24.0 $^{1.1}_{-0.8}$
		N9	—	—	—	—	—	—
		O6	10.4 $^{0.3}_{-0.3}$	3.9 $^{0.2}_{-0.2}$	40.9 $^{0.7}_{-0.5}$	31.9 $^{0.5}_{-0.4}$	34.2 $^{1.6}_{-1.0}$	-10.3 $^{0.2}_{-0.2}$
	C	N1	—	—	—	—	—	—
		N3	—	—	—	—	—	49.6 $^{2.4}_{-1.2}$
		O2	56.9 $^{0.5}_{-0.4}$	54.9 $^{0.9}_{-0.6}$	—	—	—	5.6 $^{0.5}_{-0.4}$
Sugar	G	O2′	—	36.1 $^{1.6}_{-1.0}$	—	78.3 $^{inf}_{-1.8}$	—	—
		O3′	—	—	—	—	—	—
		O4′	—	—	—	—	—	—
		O5′	—	—	—	—	—	—
	C	O2′	37.4 $^{2.2}_{-1.1}$	40.4 $^{8.4}_{-1.7}$	49.8 $^{0.6}_{-0.5}$	47.8 $^{0.7}_{-0.6}$	41.3 $^{4.9}_{-1.5}$	40.1 $^{3.5}_{-1.4}$
		O3′	—	—	—	—	—	—
		O4′	—	—	—	—	—	—
		O5′	—	—	—	—	—	—
Phosp.	G	O1P	-7.0 $^{inf}_{-1.7}$	-15.7 $^{2.8}_{-1.3}$	-7.4 $^{0.9}_{-0.7}$	-3.7 $^{5.7}_{-1.6}$	28.1 $^{inf}_{-1.9}$	-14.4 $^{2.5}_{-1.2}$
		O2P	-13.3 $^{0.9}_{-0.6}$	-24.1 $^{0.7}_{-0.6}$	-17.2 $^{0.6}_{-0.5}$	-26.8 $^{0.4}_{-0.3}$	-26.6 $^{1.1}_{-0.8}$	-17.4 $^{1.0}_{-0.7}$
	C	O1P	-7.9 $^{0.5}_{-0.4}$	-14.8 $^{2.7}_{-1.3}$	-5.3 $^{0.9}_{-0.7}$	-6.3 $^{2.7}_{-1.3}$	-12.9 $^{3.0}_{-1.3}$	-5.6 $^{0.2}_{-0.2}$
		O2P	-6.6 $^{2.6}_{-1.2}$	-15.5 $^{1.7}_{-1.0}$	-7.5 $^{1.3}_{-0.8}$	-8.3 $^{1.4}_{-0.9}$	-15.1 $^{4.4}_{-1.5}$	8.0 $^{inf}_{-1.9}$

Bibliography

[1] Francis Crick. "Central dogma of molecular biology". *Nature* 227.5258 (1970), pp. 561–563.

[2] Walter Gilbert. "Origin of life: The RNA world". *nature* 319.6055 (1986).

[3] Raymond Gesteland and J Atkins. "ThE RNA World" (1993).

[4] William G Scott, Abraham Szöke, Josh Blaustein, Sara M O'Rourke, and Michael P Robertson. "RNA catalysis, thermodynamics and the origin of life". *Life* 4.2 (2014), pp. 131–141.

[5] José Gallego and Gabriele Varani. "Targeting RNA with small-molecule drugs: therapeutic promise and chemical challenges". *Accounts of chemical research* 34.10 (2001), pp. 836–843.

[6] Han Ying, Tal Z Zaks, Rong-Fu Wang, Kari R Irvine, Udai S Kammula, Francesco M Marincola, Wolfgang W Leitner, and Nicholas P Restifo. "Cancer therapy using a self-replicating RNA vaccine". *Nature medicine* 5.7 (1999), p. 823.

[7] Qingfei Jiang, Leslie A Crews, Frida Holm, and Catriona HM Jamieson. "RNA editing-dependent epitranscriptome diversity in cancer stem cells". *Nature Reviews Cancer* 17.6 (2017), pp. 381–392.

[8] Kelly Kruger, Paula J. Grabowski, Arthur J. Zaug, Julie Sands, Daniel E. Gottschling, and Thomas R. Cech. "Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena". *Cell* 31.1 (1982), pp. 147–157.

[9] VA Erdmann, MZ Barciszewska, A Hochberg, Nathan De Groot, and J Barciszewski. "Regulatory RNAs". *Cellular and Molecular Life Sciences* 58.7 (2001), pp. 960–977.

[10] Maumita Mandal and Ronald R Breaker. "Gene regulation by riboswitches". *Nature reviews. Molecular cell biology* 5.6 (2004), p. 451.

[11] EJJH Domingo and JJ Holland. "RNA virus mutations and fitness for survival". *Annual Reviews in Microbiology* 51.1 (1997), pp. 151–178.

[12] Neocles B Leontis and Eric Westhof. "Analysis of RNA motifs". *Current opinion in structural biology* 13.3 (2003), pp. 300–308.

[13] Ignacio Tinoco and Carlos Bustamante. "How RNA folds". *Journal of molecular biology* 293.2 (1999), pp. 271–281.

[14] Jan Lipfert, Sebastian Doniach, Rhiju Das, and Daniel Herschlag. "Understanding nucleic acid–ion interactions". *Annual review of biochemistry* 83 (2014), p. 813.

[15] Jamie H Cate, Raven L Hanna, and Jennifer A Doudna. "A magnesium ion core at the heart of a ribozyme domain". *Nature structural biology* 4.7 (1997), pp. 553–558.

[16] Ross Shiman and David E Draper. "Stabilization of RNA tertiary structure by monovalent cations". *Journal of molecular biology* 302.1 (2000), pp. 79–91.

[17] A Stein and DM Crothers. "Conformational changes of transfer RNA. The role of magnesium (II)". *Biochemistry* 15.1 (1976), pp. 160–168.

[18] Anna Marie Pyle. "Ribozymes: a distinct class of metalloenzymes". *Science* 261.5122 (1993), pp. 709–714.

[19] Eva Freisinger and Roland K.O. Sigel. "From nucleotides to ribozymes-A comparison of their metal ion binding properties". *Coordination Chemistry Reviews* 251.13-14 (2007), pp. 1834 –1851.

[20] Roland K. O. Sigel. "Group II intron ribozymes and metal ions–a delicate relationship". *European journal of inorganic chemistry* 2005.12 (2005), pp. 2281–2292.

[21] Ruben L Gonzalez and Ignacio Tinoco. "Solution structure and thermodynamics of a divalent metal ion binding site in an RNA pseudoknot". *Journal of molecular biology* 289.5 (1999), pp. 1267–1282.

[22] A Jack, JE Landner, D Rhodes, RS Brown, and A Klug. "A crystallographic study of metal-binding to yeast phenylalanine transfer RNA". *Journal of molecular biology* 111.3 (1977), pp. 315–328.

[23] Loren Dean Williams. "Between Objectivity and Whim: Nucleic Acid Structural Biology". *DNA Binders and Related Subjects*. Ed. by Michael J. Waring and Jonathan B. Chaires. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 77–88.

[24] Filip Leonarski, Luigi D'Ascenzo, and Pascal Auffinger. "Binding of metals to purine N7 nitrogen atoms and implications for nucleic acids: A CSD survey". *Inorganica Chimica Acta* 452 (2016), pp. 82–89.

[25] Jiří Šponer, Giovanni Bussi, Miroslav Krepl, Pavel Banáš, Sandro Bottaro, Richard A. Cunha, Alejandro Gil-Ley, Giovanni Pinamonti, Simón Poblete, Petr Jurečka, Nils G. Walter, and Michal Otyepka. "RNA Structural Dynamics as Captured by Molecular Simulations: A Comprehensive Overview". *Chemical Reviews* (under revision).

[26] Richard A. Cunha and Giovanni Bussi. "Unraveling Mg^{2+}-RNA binding with atomistic molecular dynamics". *RNA* 23.5 (2017), pp. 628–638.

[27] David H Mathews, Jeffrey Sabina, Michael Zuker, and Douglas H Turner. "Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure". *Journal of molecular biology* 288.5 (1999), pp. 911–940.

[28] Alexander P Gultyaev, René CL Olsthoorn, Cornelis WA Pleij, and Eric Westhof. "RNA structure: pseudoknots". *eLS* (2012).

[29] CAG Haasnoot, HP Westerink, GA Van der Marel, and JH Van Boom. "Discrimination between A-type and B-type conformations of double helical nucleic acid fragments in solution by means of two-dimensional nuclear Overhauser experiments". *Journal of Biomolecular Structure and Dynamics* 2.2 (1984), pp. 345–360.

[30] R R Gutell, N Larsen, and C R Woese. "Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective." *Microbiological Reviews* 58.1 (1994), pp. 10–26.

[31] Fatih Ozsolak and Patrice M Milos. "RNA sequencing: advances, challenges and opportunities". *Nature reviews. Genetics* 12.2 (2011), p. 87.

[32] Erika Check Hayden. "Genome sequencing: the third generation". *Nature* 457.7231 (2009), pp. 768–9.

[33] Jane S Richardson, Bohdan Schneider, Laura W Murray, Gary J Kapral, Robert M Immormino, Jeffrey J Headd, David C Richardson, Daniela Ham, Eli Hershkovits, Loren Dean Williams, et al. "RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution)". *Rna* 14.3 (2008), pp. 465–481.

[34] Jesse Stombaugh, Craig L Zirbel, Eric Westhof, and Neocles B Leontis. "Frequency and isostericity of RNA base pairs". *Nucleic acids research* 37.7 (2009), pp. 2294–2312.

[35] Stephen Jefferson Sharp, Jerone Schaack, Lyan Cooley, Debroh Johnson Burke, and Dieter Soil. "Structure and transcription of eukaryotic tRNA gene". *Critical Reviews In Biochemistry* 19.2 (1985), pp. 107–144.

[36] Andrei Korostelev, Dmitri N Ermolenko, and Harry F Noller. "Structural dynamics of the ribosome". *Current opinion in chemical biology* 12.6 (2008), pp. 674–683.

[37] Yingfu Li and Ronald R Breaker. "Deoxyribozymes: new players in the ancient game of biocatalysis". *Current opinion in structural biology* 9.3 (1999), pp. 315–323.

[38] Alexander F Palazzo and Eliza S Lee. "Non-coding RNA: what is functional and what is junk?" *Frontiers in genetics* 6 (2015).

[39] John S. Mattick and Igor V. Makunin. "Small regulatory RNAs in mammals". *Human Molecular Genetics* 14 (2005), R121–R132.

[40] Alexander Serganov and Evgeny Nudler. "A decade of riboswitches". *Cell* 152.1 (2013), pp. 17–24.

[41] David E. Draper. "A guide to ions and RNA structure". *RNA* 10.3 (2004), pp. 335–343.

[42] Suzette A Pabit, Xiangyun Qiu, Jessica S Lamb, Li Li, Steve P Meisburger, and Lois Pollack. "Both helix topology and counterion distribution contribute to the more effective charge screening in dsRNA compared with dsDNA". *Nucleic acids research* 37.12 (2009), pp. 3887–3896.

[43] Kathleen Hall, Phillip Cruz, Ignacio Tinoco, Thomas M Jovin, and Johan H van de Sande. "'Z-RNA'- a left-handed RNA double helix". *Nature* 311.5986 (1984), pp. 584–586.

[44] Mariusz Popenda, Jan Milecki, and Ryszard W Adamiak. "High salt solution structure of a left-handed RNA double helix". *Nucleic acids research* 32.13 (2004), pp. 4044–4054.

[45] Desirae Leipply and David E. Draper. "Dependence of RNA Tertiary Structural Stability on Mg^{2+} Concentration: Interpretation of the Hill Equation and Coefficient". *Biochemistry* 49.9 (2010), pp. 1843–1853.

[46] Peter L Privalov, Anatoly I Dragan, and Colyn Crane-Robinson. "Interpreting protein/DNA interactions: distinguishing specific from non-specific and electrostatic from non-electrostatic components". *Nucleic acids research* 39.7 (2010), pp. 2483–2491.

[47] PE Cole, SK Yang, and DM Crothers. "Conformational changes of transfer ribonucleic acid. Equilibrium phase diagrams". *Biochemistry* 11.23 (1972), pp. 4358–4368.

[48] Melissa Maderia, Laura M Hunsicker, and Victoria J DeRose. "Metal-Phosphate Interactions in the Hammerhead Ribozyme Observed by 31P NMR and Phosphorothioate Substitutions". *Biochemistry* 39.40 (2000), pp. 12113–12120.

[49] Vincent B Chu, Yu Bai, Jan Lipfert, Daniel Herschlag, and Sebastian Doniach. "A repulsive field: advances in the electrostatics of the ion atmosphere". *Current opinion in chemical biology* 12.6 (2008), pp. 619–625.

[50] Anton S Petrov, Gene Lamm, and George R Pack. "Calculation of the binding free energy for magnesium–RNA interactions". *Biopolymers* 77.3 (2005), pp. 137–154.

[51] David E. Draper. "RNA Folding: Thermodynamic and Molecular Descriptions of the Roles of Ions". *Biophysical Journal* 95.12 (2008), pp. 5489 –5495.

[52] Anton S. Petrov, Jessica C. Bowman, Stephen C. Harvey, and Loren Dean Williams. "Bidentate RNA-magnesium clamps: On the origin of the special role of magnesium in RNA folding". *RNA* 17.2 (2011), pp. 291–297.

[53] Julie L. Fiore, Erik D. Holmstrom, and David J. Nesbitt. "Entropic origin of Mg^{2+}+-facilitated RNA folding". *Proceedings of the National Academy of Sciences of U.S.A.* 109.8 (2012), pp. 2902–2907.

[54] David E Draper, Dan Grilley, and Ana Maria Soto. "Ions and RNA folding". *Annual Review of Biophysics and Biomolecular Structure* 34 (2005), pp. 221–243.

[55] M Thomas Record Jr, Wentao Zhang, and Charles F Anderson. "Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: a practical guide to recognizing and interpreting polyelectrolyte effects, Hofmeister effects, and osmotic effects of salts." *Advances in protein chemistry* 51 (1998), p. 281.

[56] Vinod K Misra and David E Draper. "On the role of magnesium ions in RNA stability". *Biopolymers* 48.2-3 (1998), pp. 113–135.

[57] Anton S Petrov, George R Pack, and Gene Lamm. "Calculations of Magnesium- Nucleic Acid Site Binding in Solution". *The Journal of Physical Chemistry B* 108.19 (2004), pp. 6072–6081.

[58] Cuauhtémoc Garcıa-Garcıa and David E Draper. "Electrostatic interactions in a peptide–RNA complex". *Journal of molecular biology* 331.1 (2003), pp. 75–88.

[59] Magdalena Gebala, Steve Bonilla, Namita Bisaria, and Daniel Herschlag. "Does cation size affect occupancy and electrostatic screening of the nucleic acid ion atmosphere?" *Journal of the American Chemical Society* 138.34 (2016), pp. 10925–10934.

[60] George D Markham, Jenny P Glusker, and Charles W Bock. "The arrangement of first-and second-sphere water molecules in divalent magnesium complexes: Results from molecular orbital and density functional theory and from structural crystallography". *The Journal of Physical Chemistry B* 106.19 (2002), pp. 5118–5134.

[61] GW Neilson and N Skipper. "K^+ coordination in aqueous solution". *Chemical physics letters* 114.1 (1985), pp. 35–38.

[62] Lavanya M Ramaniah, Marco Bernasconi, and Michele Parrinello. "Ab initio molecular-dynamics simulation of K^+ solvation in water". *The Journal of chemical physics* 111.4 (1999), pp. 1587–1591.

[63] Philippe Brion and Eric Westhof. "Hierarchy and dynamics of RNA folding". *Annual review of biophysics and biomolecular structure* 26.1 (1997), pp. 113–137.

[64] Xingwang Fang, Tao Pan, and Tobin R Sosnick. "A thermodynamic framework and cooperativity in the tertiary folding of a Mg^{2+}-dependent ribozyme". *Biochemistry* 38.51 (1999), pp. 16840–16846.

[65] Gregory Bokinsky, David Rueda, Vinod K Misra, Maria M Rhodes, Andrew Gordus, Hazen P Babcock, Nils G Walter, and Xiaowei Zhuang. "Single-molecule transition-state analysis of RNA folding". *Proceedings of the National Academy of Sciences* 100.16 (2003), pp. 9302–9307.

[66] Rhiju Das, Kevin J Travers, Yu Bai, and Daniel Herschlag. "Determining the Mg^2+ stoichiometry for folding an RNA metal ion core". *Journal of the American Chemical Society* 127.23 (2005), pp. 8272–8273.

[67] Larisa E Kapinos, Bert P Operschall, Erik Larsen, and Helmut Sigel. "Understanding the acid–base properties of adenosine: the intrinsic basicities of N1, N3 and N7". *Chemistry-a European Journal* 17.29 (2011), pp. 8156–8164.

[68] Fawzia M Al-Sogair, Bert P Operschall, Astrid Sigel, Helmut Sigel, Joachim Schnabl, and Roland KO Sigel. "Probing the metal-ion-binding strength of the hydroxyl group". *Chemical reviews* 111.8 (2011), pp. 4964–5003.

[69] Michèle C Erat, Jonathan Coles, Cinzia Finazzo, Bernd Knobloch, and Roland KO Sigel. "Accurate analysis of Mg^{2+} binding to RNA: From classical methods to a novel iterative calculation procedure". *Coordination Chemistry Reviews* 256.1 (2012), pp. 279–288.

[70] Heping Zheng, Ivan G. Shabalin, Katarzyna B. Handing, Janusz M. Bujnicki, and Wladek Minor. "Magnesium-binding architectures in RNA crystal structures: validation, binding preferences, classification and motif detection". *Nucleic Acids Research* 43.7 (2015), pp. 3789–3801.

[71] Victoria J DeRose. "Characterization of nucleic acid metal ion binding by spectroscopic techniques". *Nucleic Acid-Metal Ion Interactions*. Royal Society of Chemistry Cambridge, UK, 2008, pp. 154–179.

[72] Magdalena Gebala, George M Giambaşu, Jan Lipfert, Namita Bisaria, Steve Bonilla, Guangchao Li, Darrin M York, and Daniel Herschlag. "Cation–anion interactions within the nucleic acid ion atmosphere revealed by ion counting". *Journal of the American Chemical Society* 137.46 (2015), pp. 14705–14715.

[73] Serdal Kirmizialtin, Suzette A. Pabit, Steve P. Meisburger, Lois Pollack, and Ron Elber. "RNA and Its Ionic Cloud: Solution Scattering Experiments and Atomically Detailed Simulations". *Biophysical Journal* 102.4 (2012), pp. 819 –828.

[74] Max Greenfeld and Daniel Herschlag. "Probing nucleic acid–ion interactions with buffer exchange-atomic emission spectroscopy". *Methods in enzymology* 469 (2009), pp. 375–389.

[75] Suzette A Pabit, Steve P Meisburger, Li Li, Joshua M Blose, Christopher D Jones, and Lois Pollack. "Counting ions around DNA with anomalous small-angle X-ray scattering". *Journal of the American Chemical Society* 132.46 (2010), pp. 16334–16336.

[76] Pascal Auffinger, Neena Grover, and Eric Westhof. "Metal ion binding to RNA". *Metal Ions in Life Sciences* 9 (2011), pp. 1–35.

[77] Heping Zheng, Mahendra D Chordia, David R Cooper, Maksymilian Chruszcz, Peter Müller, George M Sheldrick, and Wladek Minor. "Validating metal binding sites in macromolecule structures using the CheckMyMetal web server". *Nature protocols* 9.1 (2014), p. 156.

[78] Filip Leonarski, Luigi D'Ascenzo, and Pascal Auffinger. "Mg^{2+} ions: do they bind to nucleobase nitrogens?" *Nucleic Acids Research* 45.2 (2017), pp. 987–1004.

[79] JA Cowan. "Metallobiochemistry of RNA. $Co(NH_3)_6^{3+}$ as a probe for $Mg_{(aq)}^{2+}$ binding sites". *Journal of inorganic biochemistry* 49.3 (1993), pp. 171–175.

[80] Gonzalo Colmenarejo and Ignacio Tinoco. "Structure and thermodynamics of metal binding in the P5 helix of a group I intron ribozyme". *Journal of molecular biology* 290.1 (1999), pp. 119–135.

[81] Michèle C Erat and Roland KO Sigel. "Determination of the intrinsic affinities of multiple site-specific Mg2+ ions coordinated to domain 6 of a group II intron ribozyme". *Inorganic chemistry* 46.26 (2007), pp. 11224–11234.

[82] David E Shaw, Martin M Deneroff, Ron O Dror, Jeffrey S Kuskin, Richard H Larson, John K Salmon, Cliff Young, Brannon Batson, Kevin J Bowers, Jack C Chao, et al. "Anton, a special-purpose machine for molecular dynamics simulation". *Communications of the ACM* 51.7 (2008), pp. 91–97.

[83] David E Shaw, JP Grossman, Joseph A Bank, Brannon Batson, J Adam Butts, Jack C Chao, Martin M Deneroff, Ron O Dror, Amos Even, Christopher H Fenton, et al. "Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer". *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. IEEE Press. 2014, pp. 41–53.

[84] David E Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, Michael P Eastwood, Joseph A Bank, John M Jumper, John K Salmon, Yibing Shan, et al. "Atomic-level characterization of the structural dynamics of proteins". *Science* 330.6002 (2010), pp. 341–346.

[85] Wilfred F van Gunsteren, Dirk Bakowies, Riccardo Baron, Indira Chandrasekhar, Markus Christen, Xavier Daura, Peter Gee, Daan P Geerke, Alice Glättli, Philippe H Hünenberger, et al. "Biomolecular modeling: goals, problems, perspectives". *Angewandte Chemie International Edition* 45.25 (2006), pp. 4064–4092.

[86] A Rahman. "Correlations in the motion of atoms in liquid argon". *Physical Review* 136.2A (1964), A405.

[87] J Andrew McCammon, Bruce R Gelin, and Martin Karplus. "Dynamics of folded proteins". *Nature* 267.5612 (1977), pp. 585–590.

[88] Lorenzo Casalino and Alessandra Magistrato. "Structural, dynamical and catalytic interplay between Mg^{2+} ions and RNA. Vices and virtues of atomistic simulations". *Inorganica Chimica Acta* (2016). in press.

[89] Maria T. Panteva, George M. Giambaşu, and Darrin M. York. "Comparison of structural, thermodynamic, kinetic and mass transport properties of Mg^{2+} ion models commonly used in biomolecular simulations". *Journal of Computational Chemistry* 36.13 (2015), pp. 970–982.

[90] HJC Berendsen and WF Van Gunsteren. "Practical algorithms for dynamic simulations". *Molecular-dynamics simulation of statistical-mechanical systems* (1986), pp. 43–65.

[91] David J Adams, Eveline M Adams, and Graham J Hills. "The computer simulation of polar liquids". *Molecular Physics* 38.2 (1979), pp. 387–400.

[92] Loup Verlet. "Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules". *Physical review* 159.1 (1967), p. 98.

[93] William C Swope, Hans C Andersen, Peter H Berens, and Kent R Wilson. "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters". *The Journal of Chemical Physics* 76.1 (1982), pp. 637–649.

[94] Michael P Allen and Dominic J Tildesley. *Computer simulation of liquids*. Oxford university press, 2017.

[95] In-Chul Yeh and Gerhard Hummer. "System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions". *The Journal of Physical Chemistry B* 108.40 (2004), pp. 15873–15879.

[96] In-Chul Yeh and Gerhard Hummer. "Diffusion and electrophoretic mobility of single-stranded RNA from molecular dynamics simulations". *Biophysical journal* 86.2 (2004), pp. 681–689.

[97] Serdal Kirmizialtin, Alexander R.J. Silalhi, Ron Elber, and Marcia O. Fenley. "The Ionic Atmosphere around A-RNA: Poisson-Boltzmann and Molecular Dynamics Simulations". *Biophysical Journal* 102.4 (2012), pp. 829 –838.

[98] Jiří Šponer, Judit E Šponer, Arnošt Mládek, Pavel Banáš, Petr Jurečka, and Michal Otyepka. "How to understand quantum chemical computations on DNA and RNA systems? A practical guide for non-specialists". *Methods* 64.1 (2013), pp. 3–11.

[99] Ivan Ivani, Pablo D Dans, Agnes Noy, Alberto Pérez, Ignacio Faustino, Adam Hospital, Jürgen Walther, Pau Andrio, Ramon Goñi, Alexandra Balaceanu, et al. "Parmbsc1: a refined force-field for DNA simulations". *Nature methods* 13.1 (2016), p. 55.

[100] Harry F Noller. "RNA structure: reading the ribosome". *Science* 309.5740 (2005), pp. 1508–1514.

[101] David E. Condon, Scott D. Kennedy, Brendan C. Mort, Ryszard Kierzek, Ilyas Yildirim, and Douglas H. Turner. "Stacking in RNA: NMR of Four Tetramers Benchmark Molecular Dynamics". *Journal of Chemical Theory and Computation* 11.6 (2015), pp. 2729–2742.

[102] Bo Zhao and Qi Zhang. "Characterizing excited conformational states of RNA by NMR spectroscopy". *Current Opinion in Structural Biology* 30 (2015). Folding and binding/Nucleic acids and their protein complexes, pp. 134 –146.

[103] Alexander D MacKerell, Bernard Brooks, Charles L Brooks, Lennart Nilsson, Benoit Roux, Youngdo Won, and Martin Karplus. "CHARMM: the energy function and its parameterization". *Encyclopedia of computational chemistry* (1998).

[104] Wendy D Cornell, Piotr Cieplak, Christopher I Bayly, Ian R Gould, Kenneth M Merz, David M Ferguson, David C Spellmeyer, Thomas Fox, James W Caldwell, and Peter A Kollman. "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules". *Journal of the American Chemical Society* 117.19 (1995), pp. 5179–5197.

[105] Junmei Wang, Piotr Cieplak, and Peter A Kollman. "How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?" *Journal of computational chemistry* 21.12 (2000), pp. 1049–1074.

[106] Piotr Cieplak, Wendy D Cornell, Christopher Bayly, and Peter A Kollman. "Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins". *Journal of Computational Chemistry* 16.11 (1995), pp. 1357–1377.

[107] Jiří Šponer, Jerzy Leszczyński, and Pavel Hobza. "Nature of nucleic acid-base stacking: nonempirical ab initio and empirical potential characterization of 10 stacked base dimers. Comparison of stacked and H-bonded base pairs". *The Journal of Physical Chemistry* 100.13 (1996), pp. 5590–5596.

[108] Judit E Šponer, Nad'a Špačková, Petr Kulhánek, Jerzy Leszczynski, and Jiří Šponer. "Non-Watson-Crick base pairing in RNA. quantum chemical analysis of the cis Watson-Crick/sugar edge base pair family". *The Journal of Physical Chemistry A* 109.10 (2005), pp. 2292–2301.

[109] Thomas E Cheatham III, Piotr Cieplak, and Peter A Kollman. "A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat". *Journal of Biomolecular Structure and Dynamics* 16.4 (1999), pp. 845–862.

[110] Jiří Šponer, Arnošt Mládek, Judit E Šponer, Daniel Svozil, Marie Zgarbová, Pavel Banáš, Petr Jurečka, and Michal Otyepka. "The DNA and RNA sugar–phosphate backbone emerges as the key player. An overview of quantum-chemical, structural biology and simulation studies". *Physical Chemistry Chemical Physics* 14.44 (2012), pp. 15257–15277.

[111] Alexey Savelyev and Alexander D MacKerell. "All-atom polarizable force field for DNA based on the classical drude oscillator model". *Journal of computational chemistry* 35.16 (2014), pp. 1219–1239.

[112] Justin A Lemkul and Alexander D MacKerell Jr. "Polarizable Force Field for DNA Based on the Classical Drude Oscillator: I. Refinement using Quantum Mechanical Base Stacking and Conformational Energetics". *Journal of Chemical Theory and Computation* 13.5 (2017), pp. 2053–2071.

[113] Justin A Lemkul and Alexander D MacKerell Jr. "Polarizable Force Field for DNA Based on the Classical Drude Oscillator: II. Microsecond Molecular Dynamics Simulations of Duplex DNA". *Journal of Chemical Theory and Computation* 13.5 (2017), pp. 2072–2085.

[114] Christina Bergonzo and Thomas E Cheatham III. "Improved force field parameters lead to a better description of RNA structure". *Journal of chemical theory and computation* 11.9 (2015), pp. 3969–3972.

[115] Alberto Pérez, Iván Marchán, Daniel Svozil, Jiří Šponer, Thomas E Cheatham, Charles A Laughton, and Modesto Orozco. "Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers". *Biophysical Journal* 92.11 (2007), pp. 3817–3829.

[116] Marie Zgarbová, Michal Otyepka, Jiří Šponer, Arnošt Mládek, Pavel Banáš, Thomas E Cheatham III, and Petr Jurečka. "Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles". *Journal of Chemical Theory and Computation* 7.9 (2011), pp. 2886–2902.

[117] T Steinbrecher, J Latzer, and DA Case. "Revised AMBER parameters for bioorganic phosphates". *Journal of chemical theory and computation* 8.11 (2012), pp. 4405–4412.

[118] HJC Berendsen, JR Grigera, and TP Straatsma. "The missing term in effective pair potentials". *Journal of Physical Chemistry* 91.24 (1987), pp. 6269–6271.

[119] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. "Comparison of simple potential functions for simulating liquid water". *The Journal of chemical physics* 79.2 (1983), pp. 926–935.

[120] Jose LF Abascal and Carlos Vega. "A general purpose model for the condensed phases of water: TIP4P/2005". *The Journal of chemical physics* 123.23 (2005), p. 234505.

[121] Hans W Horn, William C Swope, Jed W Pitera, Jeffry D Madura, Thomas J Dick, Greg L Hura, and Teresa Head-Gordon. "Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew". *The Journal of chemical physics* 120.20 (2004), pp. 9665–9678.

[122] Saeed Izadi, Ramu Anandakrishnan, and Alexey V Onufriev. "Building water models: a different approach". *The journal of physical chemistry letters* 5.21 (2014), pp. 3863–3871.

[123] Jordi Muñoz, Šponer, Jiří, Pavel Hobza, Modesto Orozco, and F Javier Luque. "Interactions of hydrated Mg^{2+} cation with bases, base pairs, and nucleotides. Electron topology, natural bond orbital, electrostatic, and vibrational study". *The Journal of Physical Chemistry B* 105.25 (2001), pp. 6051–6060.

[124] Daniel Spångberg and Kersti Hermansson. "Many-body potentials for aqueous Li^+, Na^+, Mg^{2+}, and Al^{3+}: Comparison of effective three-body potentials and polarizable models". *The Journal of Chemical Physics* 120.10 (2004), pp. 4829–4843.

[125] Olof Allnér, Lennart Nilsson, and Alessandra Villa. "Magnesium ion–water coordination and exchange in biomolecular simulations". *Journal of Chemical Theory and Computation* 8.4 (2012), pp. 1493–1502.

[126] Pengfei Li, Benjamin P Roberts, Dhruva K Chakravorty, and Kenneth M Merz Jr. "Rational design of particle mesh Ewald compatible Lennard-Jones parameters for+ 2 metal cations in explicit solvent". *Journal of Chemical Theory and Computation* 9.6 (2013), pp. 2733–2748.

[127] Pengfei Li and Kenneth M Merz Jr. "Taking into account the ion-induced dipole interaction in the nonbonded model of ions". *Journal of chemical theory and computation* 10.1 (2013), pp. 289–297.

[128] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. "Well-tempered metadynamics: A smoothly converging and tunable free-energy method". *Physical Review Letters* 100.2 (2008), p. 020603.

[129] Alessandro Laio and Michele Parrinello. "Escaping free-energy minima". *Proceedings of the National Academy of Sciences of U.S.A.* 99.20 (2002), pp. 12562–12566.

[130] James F Dama, Michele Parrinello, and Gregory A Voth. "Well-tempered metadynamics converges asymptotically". *Physical review letters* 112.24 (2014), p. 240602.

[131] Alessandro Laio, Antonio Rodriguez-Fortea, Francesco Luigi Gervasio, Matteo Ceccarelli, and Michele Parrinello. "Assessing the accuracy of metadynamics". *The journal of physical chemistry B* 109.14 (2005), pp. 6714–6721.

[132] Alejandro Gil-Ley and Giovanni Bussi. "Enhanced Conformational Sampling Using Replica Exchange with Collective-Variable Tempering". *Journal of Chemical Theory and Computation* 11.3 (2015), pp. 1077–1085.

[133] Yuji Sugita, Akio Kitao, and Yuko Okamoto. "Multidimensional replica-exchange method for free-energy calculations". *The Journal of Chemical Physics* 113.15 (2000), pp. 6042–6051.

[134] Stefano Piana and Alessandro Laio. "A bias-exchange approach to protein folding". *The Journal of Physical Chemistry B* 111.17 (2007), pp. 4553–4559.

[135] Hiroaki Fukunishi, Osamu Watanabe, and Shoji Takada. "On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction". *The Journal of chemical physics* 116.20 (2002), pp. 9058–9067.

[136] Cameron Abrams and Giovanni Bussi. "Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration". *Entropy* 16.1 (2013), pp. 163–199.

[137] Giovanni Bussi, Francesco Luigi Gervasio, Alessandro Laio, and Michele Parrinello. "Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics". *Journal of the American Chemical Society* 128.41 (2006), pp. 13435–13441.

[138] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method". *Journal of Computational Chemistry* 13.8 (1992), pp. 1011–1021.

[139] Fabrizio Marinelli, Fabio Pietrucci, Alessandro Laio, and Stefano Piana. "A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations". *PLOS Computational Biology* 5.8 (Aug. 2009), pp. 1–18.

[140] Jörg Rinnenthal, Janina Buck, Jan Ferner, Anna Wacker, Boris Fürtig, and Harald Schwalbe. "Mapping the Landscape of RNA Dynamics with NMR Spectroscopy". *Accounts of Chemical Research* 44.12 (2011), pp. 1292–1301.

[141] Hashim M Al-Hashimi and Nils G Walter. "RNA dynamics: it is about time". *Current Opinion in Structural Biology* 18.3 (2008), pp. 321 –329.

[142] Anna Pyle. "Metal ions in the structure and function of RNA". *JBIC Journal of Biological Inorganic Chemistry* 7.7 (2002), pp. 679–690.

[143] Vinod K. Misra, Ross Shiman, and David E. Draper. "A thermodynamic framework for the magnesium-dependent folding of RNA". *Biopolymers* 69.1 (2003), pp. 118–136.

[144] Steven O. Nielsen, Rosa E. Bulo, Preston B. Moore, and Bernd Ensing. "Recent progress in adaptive multiscale molecular dynamics simulations of soft matter". *Phys. Chem. Chem. Phys.* 12 (39 2010), pp. 12401–12414.

[145] Christina Bergonzo, Kathleen B Hall, and Thomas E Cheatham. "Divalent Ion Dependent Conformational Changes in an RNA Stem-Loop Observed by Molecular Dynamics". *Journal of Chemical Theory and Computation* (2016).

[146] Thomas J. Macke and David A. Case. "Modeling Unusual Nucleic Acid Structures". *Molecular Modeling of Nucleic Acids* (1998), pp. 379–393.

[147] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R. Shirts, Jeremy C. Smith, Peter M. Kasson, David van der Spoel, Berk Hess, and Erik Lindahl. "GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit". *Bioinformatics* 29.7 (2013), pp. 845–854.

[148] Giovanni Bussi, Davide Donadio, and Michele Parrinello. "Canonical sampling through velocity rescaling". *The Journal of Chemical Physics* 126.1 (2007), p. 014101.

[149] Tom Darden, Darrin York, and Lee Pedersen. "Particle mesh Ewald: An N log (N) method for Ewald sums in large systems". *The Journal of Chemical Physics* 98.12 (1993), pp. 10089–10092.

[150] In Suk Joung and Thomas E Cheatham III. "Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations". *The journal of Physical Chemistry B* 112.30 (2008), pp. 9020–9041.

[151] Andrea Cesari, Alejandro Gil-Ley, and Giovanni Bussi. "Combining simulations and solution experiments as a paradigm for RNA force field refinement". *Journal of chemical theory and computation* 12.12 (2016), pp. 6192–6200.

[152] Alejandro Gil-Ley, Sandro Bottaro, and Giovanni Bussi. "Empirical Corrections to the Amber RNA Force Field with Target Metadynamics". *Journal of Chemical Theory and Computation* 12.6 (2016), pp. 2790–2798.

[153] Davide Branduardi, Giovanni Bussi, and Michele Parrinello. "Metadynamics with adaptive Gaussians". *Journal of Chemical Theory and Computation* 8.7 (2012), pp. 2247–2254.

[154] Michael K Gilson, James A Given, Bruce L Bush, and J Andrew McCammon. "The statistical-thermodynamic basis for computation of binding affinities: a critical review." *Biophysical Journal* 72.3 (1997), p. 1047.

[155] Roland K. O. Sigel and Helmut Sigel. "A Stability Concept for Metal Ion Coordination to Single-Stranded Nucleic Acids and Affinities of Individual Sites". *Accounts of Chemical Research* 43.7 (2010), pp. 974–984.

[156] Danny Kowerko, Sebastian LB König, Miriam Skilandat, Daniela Kruschel, Mélodie CAS Hadzic, Lucia Cardo, and Roland KO Sigel. "Cation-induced kinetic heterogeneity of the intron–exon recognition in single group II introns". *Proceedings of the National Academy of Sciences* 112.11 (2015), pp. 3403–3408.

[157] Jennifer A Doudna and Thomas R Cech. "The chemical repertoire of natural ribozymes". *Nature* 418.6894 (2002), pp. 222–228.

[158] Kevin V Morris and John S Mattick. "The rise of regulatory RNA". *Nature reviews. Genetics* 15.6 (2014), p. 423.

[159] Romina Oliva and Luigi Cavallo. "Frequency and Effect of the Binding of Mg^{2+}, Mn^{2+}, and Co^{2+} Ions on the Guanine Base in Watson-Crick and Reverse Watson-Crick Base Pairs". *The Journal of Physical Chemistry B* 113.47 (2009), pp. 15670–15678.

[160] Jessica C Bowman, Timothy K Lenz, Nicholas V Hud, and Loren Dean Williams. "Cations in charge: magnesium ions in RNA folding and catalysis". *Current opinion in structural biology* 22.3 (2012), pp. 262–272.

[161] C. V. Bizarro, A. Alemany, and F. Ritort. "Non-specific binding of Na^+ and Mg^{2+} to RNA determined by force spectroscopy methods". *Nucleic Acids Research* 40.14 (2012), pp. 6922–6935.

[162] Michele C. Erat and Roland K. O. Sigel. "2 Methods to Detect and Characterize Metal Ion Binding Sites in RNA". *Structural and Catalytic Roles of Metal Ions in RNA*. Vol. 9. The Royal Society of Chemistry, 2011, pp. 37–100.

[163] Bernd Hoffmann, G. Thomas Mitchell, Patrick Gendron, Franćois Major, Angela A. Andersen, Richard A. Collins, and Pascale Legault. "NMR structure of the active conformation of the Varkud satellite ribozyme cleavage site". *Proceedings of the National Academy of Sciences of U.S.A.* 100.12 (2003), pp. 7003–7008.

[164] Dean O. Campbell, Patricia Bouchard, Geneviéve Desjardins, and Pascale Legault. "NMR Structure of Varkud Satellite Ribozyme Stem-Loop V in the Presence of Magnesium Ions and Localization of Metal-Binding Sites," *Biochemistry* 45.35 (2006), pp. 10591–10605.

[165] Simona Bartova, Maria Pechlaner, Daniela Donghi, and Roland K. O. Sigel. "Studying metal ion binding properties of a three-way junction RNA by heteronuclear NMR". *JBIC Journal of Biological Inorganic Chemistry* 21.3 (2016), pp. 319–328.

[166] D Rey Banatao, Russ B Altman, and Teri E Klein. "Microenvironment analysis and identification of magnesium binding sites in RNA". *Nucleic Acids Research* 31.15 (2003), pp. 4450–4460.

[167] Tamar Schlick. *Molecular modeling and simulation: an interdisciplinary guide: an interdisciplinary guide*. Vol. 21. Springer Science & Business Media, 2010.

[168] Stefan Kolev, Petko St. Petkov, Miroslav Rangelov, and Georgi N. Vayssilov. "Ab Initio Molecular Dynamics of Na+ and Mg^{2+} Countercations at the Backbone of RNA in Water Solution". *ACS Chemical Biology* 8.7 (2013), pp. 1576–1589.

[169] Anan Tongraar and Bernd Michael Rode. "The role of non-additive contributions on the hydration shell structure of Mg^{2+} studied by Born-Oppenheimer ab initio quantum mechanical/molecular mechanical molecular dynamics simulation". *Chemical Physics Letters* 346.5-6 (2001), pp. 485 –491.

[170] Vojtěch Mlỳnskỳ, Nils G Walter, Jiří Šponer, Michal Otyepka, and Pavel Banáš. "The role of an active site Mg^{2+} in HDV ribozyme self-cleavage: insights from QM/MM calculations". *Physical Chemistry Chemical Physics* 17.1 (2015), pp. 670–679.

[171] Feng Pan, Christopher Roland, and Celeste Sagui. "Ion distributions around left- and right-handed DNA and RNA duplexes: a comparative study". *Nucleic Acids Research* 43 (2014), pp. 7260–7269.

[172] Ryan L Hayes, Jeffrey K Noel, Ana Mandic, Paul C Whitford, Karissa Y Sanbonmatsu, Udayan Mohanty, and José N Onuchic. "Magnesium Dependence of the RNA Free Energy Landscape". *Biophysical Journal* 108.2 (2015), 235a.

[173] Paul S. Henke and Chi H. Mak. "An implicit divalent counterion force field for RNA molecular dynamics". *The Journal of Chemical Physics* 144.10 (2016).

[174] Changbong Hyeon and D. Thirumalai. "Capturing the essence of folding and functions of biomolecules using coarse-grained models". *Nature Communications* 2 (Sept. 2011), p. 487.

[175] Anne Bleuzen, Lothar Pittet Pierre-Andréand Helm, and André E. Merbach. "Water exchange on magnesium(II) in aqueous solution: a variable temperature and pressure 17O NMR study". *Magnetic Resonance in Chemistry* 35.11 (1997), pp. 765–773.

[176] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. "Comparison of simple potential functions for simulating liquid water". *The Journal of Chemical Physics* 79.2 (1983), pp. 926–935.

[177] Herman JC Berendsen, J Pl M Postma, Wilfred F van Gunsteren, ARHJ DiNola, and JR Haak. "Molecular dynamics with coupling to an external bath". *The Journal of Chemical Physics* 81.8 (1984), pp. 3684–3690.

[178] Michele Parrinello and Aneesur Rahman. "Polymorphic transitions in single crystals: A new molecular dynamics method". *Journal of Applied Physics* 52.12 (1981), pp. 7182–7190.

[179] Gareth A Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. "PLUMED 2: New feathers for an old bird". *Computer Physics Communications* 185.2 (2014), pp. 604–613.

[180] Glenn M Torrie and John P Valleau. "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling". *Journal of Computational Physics* 23.2 (1977), pp. 187–199.

[181] Maria T. Panteva, George M. Giambaşu, and Darrin M. York. "Force Field for Mg^{2+}, Mn^{2+}, Zn^{2+}, and Cd^{2+} Ions That Have Balanced Interactions with Nucleic Acids". *The Journal of Physical Chemistry B* 119.50 (2015), pp. 15460–15470.

[182] Haibo Yu, Troy W. Whitfield, Edward Harder, Guillaume Lamoureux, Igor Vorobyov, Victor M. Anisimov, Jr. Alexander D. MacKerell, and Benoît Roux. "Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field". *Journal of Chemical Theory and Computation* 6.3 (2010), pp. 774–786.

[183] Andrea Pérez-Villa, Maria Darvas, and Giovanni Bussi. "ATP dependent NS3 helicase interaction with RNA: insights from molecular simulations". *Nucleic Acids Research* 45 (2015), pp. 5883–5891.

[184] Francesco Di Palma, Sandro Bottaro, and Giovanni Bussi. "Kissing loop interaction in adenine riboswitch: insights from umbrella sampling simulations". *BMC bioinformatics* 16.9 (2015), S6.

[185] Jed W Pitera and John D Chodera. "On the use of experimental observations to bias simulated ensembles". *Journal of Chemical Theory and Computation* 8.10 (2012), pp. 3445–3451.

[186] Pavel Banáš, Daniel Hollas, Marie Zgarbová, Petr Jurečka, Modesto Orozco, III Thomas E. Cheatham, Jiří Šponer, and Michal Otyepka. "Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins". *Journal of Chemical Theory and Computation* 6.12 (2010), pp. 3836–3849.

[187] Christina Bergonzo, Niel M. Henriksen, Daniel R. Roe, and Thomas E. Cheatham. "Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields". *RNA* 21.9 (2015), pp. 1578–1590.

[188] Christina Bergonzo, Niel M. Henriksen, Daniel R. Roe, Jason M. Swails, Adrian E. Roitberg, and III Thomas E. Cheatham. "Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide". *Journal of Chemical Theory and Computation* 10.1 (2014), pp. 492–499.

[189] Sandro Bottaro, Alejandro Gil-Ley, and Giovanni Bussi. "RNA folding pathways in stop motion". *Nucleic Acids Research* 44.12 (2016), p. 5883.

[190] C. Perego, M. Salvalaglio, and M. Parrinello. "Molecular dynamics simulations of solutions at constant chemical potential". *The Journal of Chemical Physics* 142.14, 144113 (2015).

[191] Aleksander V. Drozdetski, Igor S. Tolokh, Lois Pollack, Nathan Baker, and Alexey V. Onufriev. "Opposing Effects of Multivalent Ions on the Flexibility of DNA and RNA". *Phys. Rev. Lett.* 117 (2 2016), p. 028101.

[192] Justin A. Lemkul, Sirish Kaushik Lakkaraju, and Alexander D. MacKerell. "Characterization of Mg^{2+} Distributions around RNA in Solution". *ACS Omega* 1.4 (2016), pp. 680–688.

[193] Jamie H Cate and Jennifer A Doudna. "Metal-binding sites in the major groove of a large ribozyme domain". *Structure* 4.10 (1996), pp. 1221–1229.

[194] Takahiro Yamauchi, Daisuke Miyoshi, Takafumi Kubodera, Akira Nishimura, Susumu Nakai, and Naoki Sugimoto. "Roles of Mg^{2+} in TPP-dependent riboswitch". *FEBS letters* 579.12 (2005), pp. 2583–2588.

[195] Arati Ramesh and Wade C Winkler. "Magnesium-sensing riboswitches in bacteria". *RNA biology* 7.1 (2010), pp. 77–83.

[196] Sandro Bottaro, Pavel Banáš, Jiří Šponer, and Giovanni Bussi. "Free Energy Landscape of GAGA and UUCG RNA Tetraloops". *The journal of physical chemistry letters* 7.20 (2016), pp. 4032–4038.