Scuola Internazionale Superiore di Studi Avanzati

Doctoral Thesis

Conformational changes in the adenine riboswitch

Author:
Francesco Di Palma

Supervisor:
Prof. Giovanni Bussi

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in

Physics and Chemistry of Biological Systems

XXVI Doctorate Series
October 16^{th}, 2014

Declaration of Authorship

I, Francesco Di Palma, declare that this thesis titled, 'Conformational changes in the adenine riboswitch' and the work presented in it are my own. I confirm that:

- This work was done wholly while in candidature for a research degree at this Istitution.
- 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: _____

"Per correr miglior acque alza le vele
omai la navicella del mio ingegno,
che lascia dietro a sé mar sì crudele;"

Dante Alighieri, *Divina Commedia, Purgatorio, Canto I (1-3)*

SISSA

Abstract

Physics Area
Molecular and Statistical BioPhysics

Doctor of Philosophy

Conformational changes in the adenine riboswitch

by Francesco Di Palma

Riboswitches are *cis*-acting genetic control elements that have been found in the untraslated region of some mRNAs in bacteria and plants. Riboswitches are known to regulate the genetic expression by means of conformational changes triggered by highly specific interactions of the aptamer with the sensed metabolite. The non-coding sequence in the mRNA of *add* gene from *V. vulnificus* contains an adenine responsive riboswitch. Classical molecular dynamics simulations of its aptamer have been performed, both in presence and absence of its physiological ligand starting from the experimental crystal structure. We first use steered MD to induce the opening of the P1 stem and investigate its stability. Our results show that the ligand directly stabilizes the P1 stem by means of stacking interactions quantitatively consistent with thermodynamic data. Then, using both umbrella sampling and a combination of metadynamics and hamiltonian replica exchange, we show that the formation of L2-L3 kissing complex cooperates with ligand binding and we quantify the ligand-induced stabilization. In this context also the influence given by either the monovalent cations or divalent cations was evaluated. Conformational changes at pairings detailed level are characterized using a recently introduced technique that is able to distinguish and classify each interaction (i.e. Watson-Crick base pair, non-canonical bp, stacking). Results are compatible with known experimental measurements and shed a new light on the ligand-dependent folding mechanism of the adenine riboswitch.

Contents

For her countless gifts, to my ninety three granny

Chapter 1

Introducing the *add* adenine riboswitch

1.1 Preface

In 1956 Francis Crick wrote: «The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred back from protein to either protein or nucleic acid »[1, 2]. This has also been simpler, but less precisely, described as «DNA makes RNA and RNA makes protein». However, it is something that decades after decades until now has been updated, because originally it only implied DNA replication, transcription and translation, mainly lacking of reverse transcription, RNA replication and post-traslational modifications (added in the 1970 paper by Crick [2]), ignoring all the different existing types of gene regulation that have enriched our knowledge about molecular biology. Living organisms not only have to deal with their life-cycle, but they must be able to "feel" the environment around them and sense the surrounding stimuli, then converting these input signals into useful cellular responses. Most of these responses are transcription factors mediated through DNA binding and coordinating in such way the activity of RNA polymerase or of proteins that induce allosteric effects on their regulatory targets. In the early 1970s, researchers discovered that also regions of mRNA transcripts have a regulatory role in the expression of their gene products [3]. By the early 1990s, several new regulatory mechanisms focused on the action of RNA (often referred to as riboregulation) had been found out [4–10]. Since then, many diverse RNA-based regulatory mechanisms have been discovered, including one that regulates interference and epigenetic regulation by long, noncoding RNA in eukaryotes [11, 12]. These emerging themes make the extent to which RNA plays a role in the normal cellular homeostasis regulation as well as in

driving physiological responses to changes in environment or extracellular signals, one of the most significant discoveries of the post-genomic era [13]. One common form of riboregulation is the use of ribonucleic acid sequences encoded within mRNA that directly affect the expression of genes encoded in the full transcript. These regulatory elements are known as riboswitches.

1.2 What is a riboswitch?

A riboswitch is a RNAs that directly monitor a physiological signal, in the absence of any other cellular factor, and exhibit structural shifts in response to that [14]. Since their first description in 2002 [15, 16], riboswitches have come into the limelight in RNA biology as an important and widespread cis-acting gene regulatory mechanism found in numerous bacteria, also evolutionarily distant, in archaea, plants, fungi, and algae. Riboswitches are structured RNA domains located in the untranslated region (UTR) of the transcript mRNA they regulate by means of allosteric structural changes induced by the binding of the sensed ligand [17]. Indeed, they change their conformation in response to specific metabolite binding [18, 19] and they have been proposed as modern descendants of an ancient sensory and regulatory system in the RNA world [20]. The discovery of riboswitches revealed that the ability of RNA to form precision binding pockets for small molecules is harnessed by modern cells to sense metabolites. At variance with the DNA-binding regulatory proteins, the RNA-binding proteins and the small regulatory RNAs processed and helped by proteins to regulate in trans the activity of other RNA transcripts, the riboswitches are able to carry out their function, binding their sensed ligand, as a stand-alone regulator in the absence of any other protein factor. Initially, several riboswitches were found by examining literature that describes the involvement of small molecule metabolites as regulators of gene control. For some of them, researchers reported the identification of a particular metabolite as a regulator of gene expression, although a regulatory protein that might respond to changes in metabolite concentration notably remained undiscovered. In each of these cases, gene regulation was later found to be riboswitch mediated, meaning that the mRNA under control directly bound its regulatory metabolite in the absence of proteins. They selectively respond to fundamental metabolites such as coenzymes, nucleobases and their derivatives, amino acids, ions, and tRNAs [21], to control a broad range of genes, including those involved in the metabolism or in the uptake of the ligand itself since the targeted ligand is often synthesized or processed by the gene(s) located downstream of the riboswitch. A classification of riboswitches based either on the sensed ligand or the organization of the binding pocket can be found in the literature [20, 22]. To regulate their target gene, riboswitch can either act on transcription, on translation, or, more rarely, as interfering,

antisense or self-splicing RNAs [20]. To understand the importance of the riboswitches in biology, it must be considered that riboswitch-mediated gene control is prevalent in bacteria, for example it regulates ≈ 2% of all genes in the bacterial species *Bacillus subtilis* [23]. Many pathogenic bacteria use riboswitches to control essential metabolic pathways and structure-based ligand design approaches are currently used to explore the possibility of developing novel antibacterial agents using them as potential drug targets [24–30]. Although most of the riboswitches have been found exclusively in bacterial species, thiamine-pyrophosphate-sensing riboswitches have also been identified in plants and fungi [31–33], and other riboswitches may also be present in eukaryotes [34, 35]. Although few examples of riboswitches have been identified in eukaryotes to date, the large amount of non-coding RNAs that are transcribed in these cells provides enormous potential for metabolite-RNA interactions. In bacteria they reside at the mRNA 5'-UTR and have a general common architecture [36]. Namely, they are composed of an aptamer region, recognizing and binding the effector ligand, and an expression platform that directly interacts with the gene expression apparatus (the RNA polymerase in the case of a regulation of transcription or the ribosome if we are dealing with a translational regulator). In this way, there is a transduction of the ligand-induced conformational "switch" into a modulation of gene expression. Indeed, upon ligand binding a structural shift occurs between the ON and the OFF state, stabilizing one conformation to the detriment of the other. Usually the switch results in the mutually exclusive formation of an anti-terminator stem or an intrinsic transcription terminator stem. Otherwise the formation of the complex solves the competition between a translation inhibitor structure and another one that allows the translation initiation. Thus, this means that there is a switching sequence that changing its conformation drives the riboswitch alternative folding.

1.3 Purine-sensing riboswitches

Among more than 20 natural aptamer classes [20], purine-sensing riboswitches have the peculiarity to recognize the targeted purine by utilizing a conserved pyrimidine [37, 38]. This family of riboswitches emerge as important model systems for exploring various aspects of RNA structure and function [39] because of their structural simplicity and relatively small size. The known purine-sensing riboswitches can be assigned to one of four riboswitch classes [17, 23, 40, 41]. However, it is possible that there are more purine-sensing riboswitch classes that have yet to be discovered. It should also be noted that there are other riboswitch classes that target purine-containing compounds, but are not included in the purine-sensing riboswitch group. These motifs include the coenzyme B_{12} riboswitch class [16], the various S-adenosylmethionine riboswitch classes [42–45] and

the S-adenosylhomocysteine riboswitch class [46]. However, the purine moiety constitutes only a portion of these ligands, whereas the four described here that are labelled as purine-sensing riboswitches all recognize a purine, modified purine or a purine deoxyribonucleoside. Also, members of the purine-sensing riboswitch classes all appear to bind their respective metabolites, at least in part by making a Watson-Crick (WC) base-pair (bp) with their target ligand. For example, guanine was shown to repress expression of the *xpt* operon, which consists of genes that transport and metabolize purines in *B. subtilis* [47]. However, a protein factor that could sense guanine and control operon expression had not been identified. From the comparison between the nucleotide sequences of the *xpt-pbuX* 5'-UTR with regions upstream of other genes involved in the purine metabolism, a highly conserved region was identified. This motif forms the aptamer for a class of riboswitches that selectively binds to guanine [23] with nM binding affinity. Subsequently, in various bacterial species numerous guanine riboswitch representatives were identified, always located upstream of an operon or a gene involved in purine synthesis and transport. Soon after the guanine riboswitch class was identified, a subset of this class was found to preferentially recognize adenine [48], and one example has been shown to bind adenine with a K_d of $\approx$ 300 nM. Representatives of adenine-sensing riboswitches reside in the mRNAs encoding adenine metabolism and transport proteins (e.g. *add*, a translational initiator as schematically explained in Figure 1.1, and *pbuE*, a transcriptional anti-terminator). Concerning the sequence and the secondary structure adenine riboswitch aptamers closely resemble the guanine one, having a conserved nucleotide content and and similar set of secondary interactions. However, a highly-conserved nucleotide within the aptamer core is changed from a cytidine residue in guanine riboswitches to a uridine residue in adenine riboswitches. This nucleotide is critical in targeting the ligand, making a WC bp with the ligand. Indeed, a single nucleotide change between C and U residues was found to be the only necessary base substitution to change ligand specificity between guanine and adenine [17]. However, these nucleotides are not the only riboswitch residues involved in ligand recognition. Many of the conserved residues in the aptamer junctions make up the conserved core region, which creates the tight binding pocket for the ligand, and at least other four residues within this region are involved in making direct ligand contacts (Figure 1.2A-B).

The precise interactions used by guanine and adenine riboswitch aptamers to selectively bind their ligands were confirmed and expanded through the use of X-ray crystallization and NMR analyses. For example, X-ray crystallographic studies of the guanine riboswitch upstream of *xpt-pbuX* in *B. subtilis* [50, 51] and the adenine riboswitch from the *add* gene of *V. vulnificus* [51] bound to their respective ligands revealed the details of purine binding by these RNAs showing that the 3D structures of the two aptamers

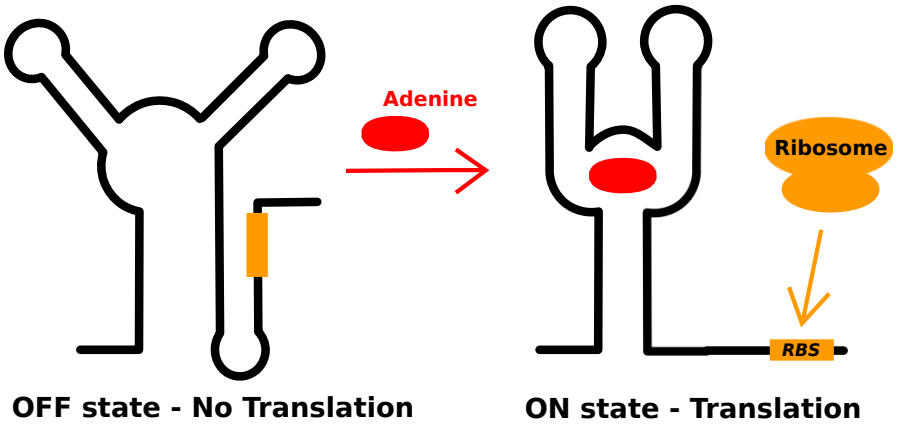

FIGURE 1.1: Cartoon showing the OFF (left) and ON (right) states of the A-riboswitch. When ligand is not present the ribosome binding site (orange, RBS) is paired with a portion of the aptamer and translation is blocked. When ligand is present the RBS is free to interact with the ribosome and translation can be initiated. Adapted from Di Palma et al. [49]

are virtually identical (Figure 1.2A-B). The global architecture of these aptamers explains some of the nucleotide sequences and secondary structures that are characteristic of guanine and adenine riboswitches. For example, the P2 and P3 stems (Figures 1.3A and 1.3B) adopt a parallel alignment and are held in place by interactions between the L2 and L3 loops. When the metabolite is bound, the overall structure of the aptamer is somewhat similar to that of a tuning fork [50]. The structural requirements place length restrictions on the P2 and P3 stems, each of which are usually formed from 7 bp. Furthermore, the nucleotides in the loops that form the tertiary contacts are highly conserved.

Another purine-riboswitch class that recognizes the $preQ_1$, a natural modified guanine, was more recently identified in *B. subtilis queC* transcript [41]. This molecule is a precursor used for the biosynthesis of queuosine, a hypermodified 7-deazaguanosine nucleoside, that is found in the anticodon wobble position of certain tRNAs [53]. This RNA motif is most often found upstream of genes involved in queuosine biosynthesis and binds $preQ_1$ causing transcription termination. The $preQ_1$ riboswitch class structure does not resemble that of the other known purine-binding motifs, but consists of one or two simple stem-loop regions, with a minimum aptamer length of only 34 nucleotides (Figure 1.2D). This small and simple riboswitch is, nevertheless, selective for its ligand over related analogues , such as guanine and hypoxanthine, with binding affinity in the order of approximately few tens nM. Thus, although the secondary structure model of the $preQ_1$ riboswitch differs from that of the other purine riboswitches, recognition of the ligand is still likely to be achieved via WC base-pairing to a specific cytidine residue within the riboswitch aptamer.

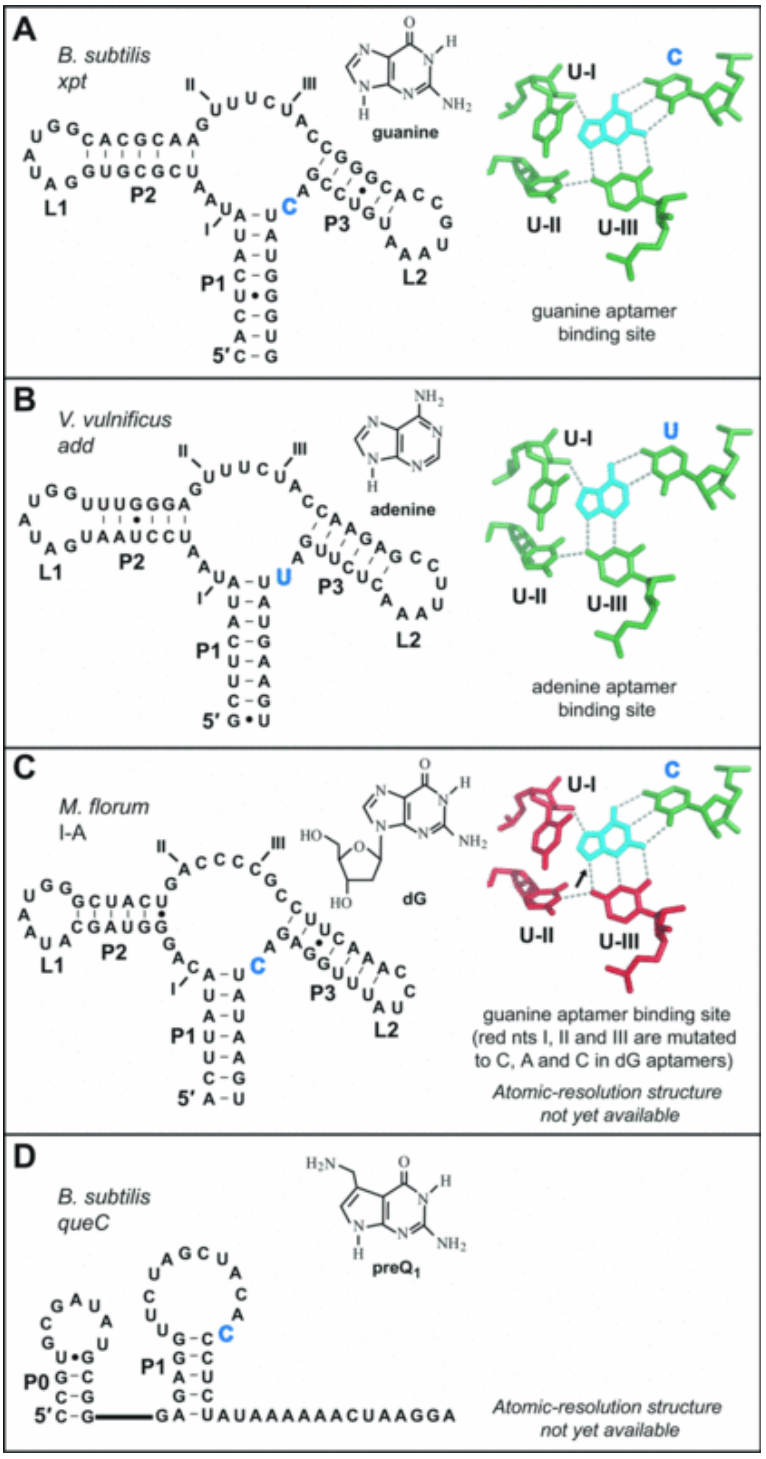

FIGURE 1.2: Secondary structures and partial 3D structure views of purine-sensing riboswitch aptamers. A) *xpt*; B) *add*; C) 2'-deoxyguanosine-sensing riboswitch aptamer from *Mesoplasma florum*; D) *queC*. Adapted from Kim and Breaker [37].

1.4 The *add* adenine riboswitch

One of the most characterized members of the purine family is the adenine sensing riboswitch (A-riboswitch) [48], found in the mRNA 5'-untraslated region, it *cis*-regulates

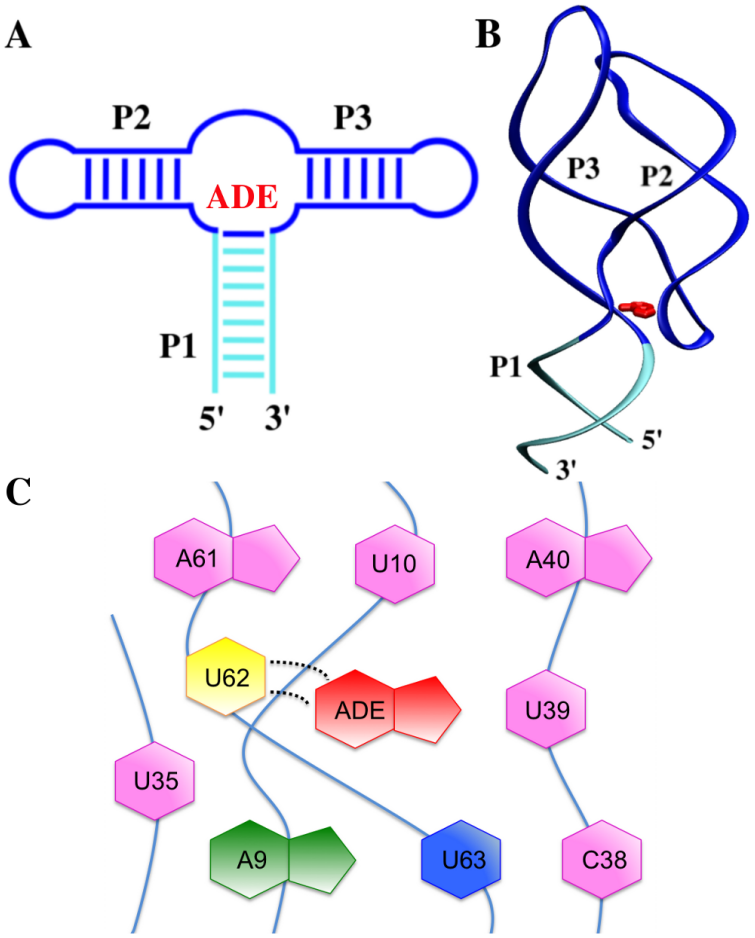

FIGURE 1.3: Adenine riboswitch aptamer and binding site. A) Secondary structure elements and B) 3-dimensional structure with bound adenine. C) Cartoon representation of the binding site; the two dotted lines represent the hydrogen bonds of the WC pairing between the U62 and the ligand. All the other bases are directly linked with the ligand forming a very stabile hydrogen-bond network. Adapted from Di Palma et al. [52]

the adenosine deaminase (*add*) gene in *Vibrio vulnificus* acting rho-independently at the translational level [48]. The A-riboswitch is a translational regulator and contributes to regulate the amount of adenine present in the system. It acts using a negative feedback mechanism [37]. In the absence of the adenine the Shine-Dalgarno (SD) sequence of the *add* gene is paired with a complementary sequence (anti-SD) and is not available for ribosome binding. Conversely, when the ligand is present it stabilizes an aptamer conformation in which the anti-SD becomes an integral part of the aptamer terminal helix, so that the SD sequence is unpaired as well as the starting codon, and translation can be initiated. In practice, the A-riboswitch regulatory activity depends on the availability of the ligand: in the presence of adenine the riboswitch is in the ON state, and the protein synthesis is permitted, whereas in the absence of the ligand the riboswitch folds into the OFF state blocking the translation initiation (Figure 1.4).

If a riboswitch controls gene expression by regulating transcription termination, then the

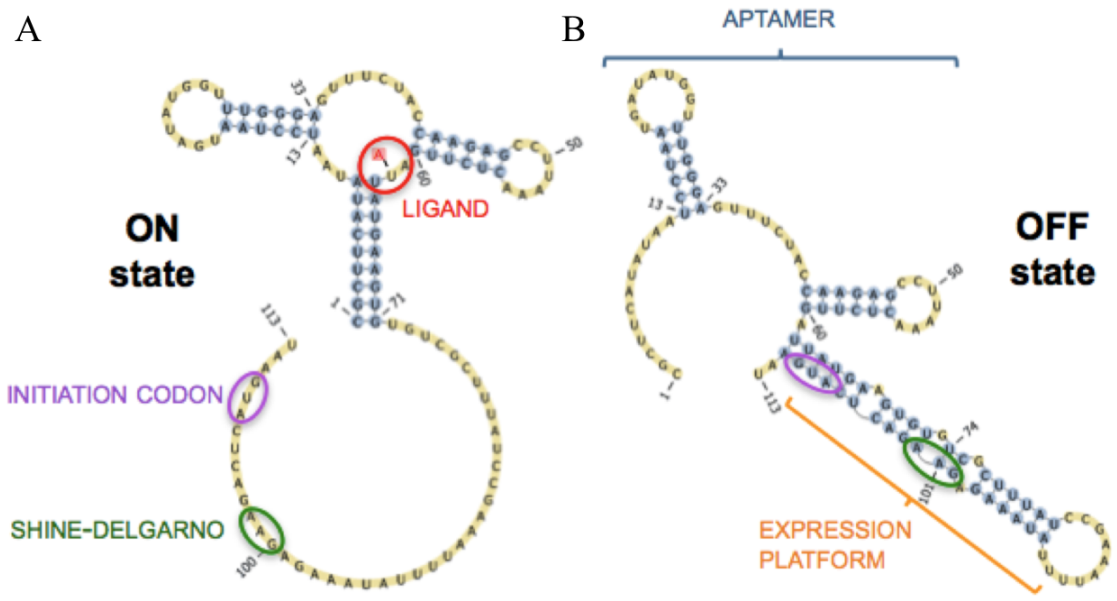

FIGURE 1.4: Secondary structure representation of the *add* riboswitch in the ON (A) and OFF (B) states. The ligand, the initiation codon and the Shine-Dalgarno sequence are labeled. Adapted from Di Palma et al. [52]

decision on gene activation or repression will be controlled by either kinetic or thermodynamic parameters (or a combination of both) [54–57] (Figure 1.5). A riboswitch aptamer that reaches equilibrium before the genetic decision can be described as a thermodynamically driven riboswitch. In other words, such riboswitches should exhibit K_d values that reflect the concentration of the metabolite needed to trigger riboswitch function in cells. However, if a riboswitch aptamer has insufficient time to reach binding equilibrium with its ligand, the concentration of ligand needed to trigger the riboswitch is dependent on the kinetics of the gene control process. The *add* riboswitch is likely to be completely thermodynamically controlled [57]. This means that adenine binding by this riboswitch can occur even after transcription of the expression platform, implying that the lifetime of transcription is not a factor in determining riboswitch character. In contrast, the adenine riboswitch in the 5' UTR of the *B. subtilis pbuE* gene appears to have the ability to be driven either kinetically or thermodynamically depending on various factors, such as the transcriptional time scale and the lifetime of the RNA-ligand complex [55, 56]. The speed of the RNA polymerase, as well as transcriptional pausing at pause sites along the template must be taken into consideration, because any stalling by the RNA polymerase may allow the riboswitch system to reach equilibrium by the time the genetic decision must be made. Thus it is possible for different riboswitches within the same class to operate differently under different control regimes.

The X-ray crystal structure of the *add* riboswitch aptamer domain was solved in 2004 and was found to have a "tuning-fork"-like three dimensional structure wherein the adenine binding occurs in a pocket formed at the junction of three stems (P1, P2, P3), two

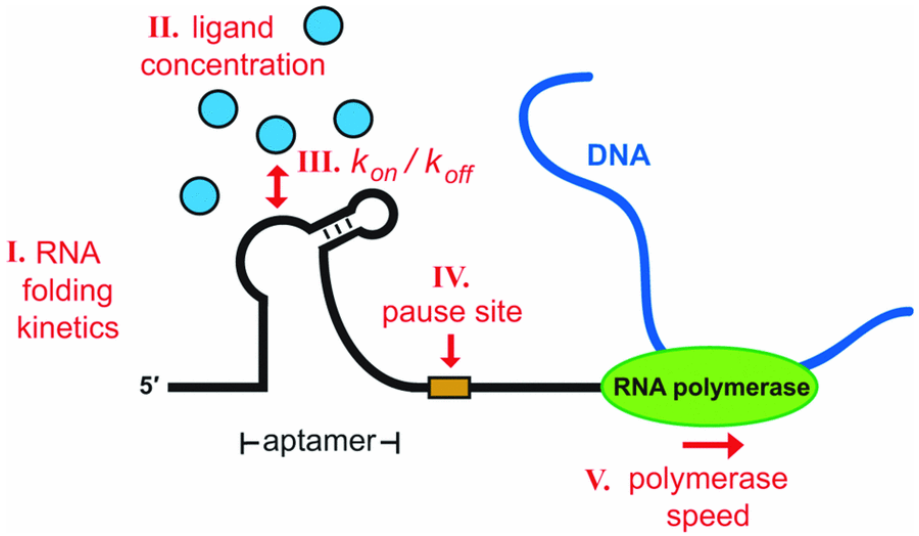

FIGURE 1.5: Factors affecting kinetic and thermodynamic control of riboswitch systems. A model depicting five different variables (I through V) that contribute to determining whether a riboswitch is kinetically or thermodynamically controlled. Adapted from Kim and Breaker [37].

of which are hairpins interacting via kissing loops (L2-L3) [51]. Thus, the ligand is completely encapsulated into the this three-way junction structure (Figure 1.3C). The specificity for adenine is ensured by canonical WC bp established between a uracil in conserved position and the ligand [56, 58]. There are three structurally important regions 1.6: the binding pocket, the P1-stem, that is formed in the ON-state and it is disrupted in the OFF-state, and the loop-loop tertiary interaction between L2 and L3, usually called "kissing loops". The latter includes two inter-loop WC bp [59, 60].

The ligand-dependent structural mechanism inducing the switch between the ON- and the OFF-state in the A-riboswitch mostly remains to be elucidated. It has been proposed that P1 is stabilized by the ligand [50] and that this could be a common feature in many riboswitch classes [61]. Moreover, it has been suggested that the ligand stabilizes also the interaction of the distal kissing complex [62]. At the same time a stable kissing interaction seems to contribute to the ligand binding energy stabilizing the complex in a cooperative fashion [39, 57, 63]. The role of the ligand in the structural organization of the aptamer has been investigated using structure-based fluorescence spectroscopy [57], real-time multidimensional NMR techniques [63] showing a joint interplay between the different secondary structure elements and the ligand, and single-molecule spectroscopy experiments [64] providing an insightful overview on the folding dynamics [64]. These investigations however lack both the atomistic details needed for an accurate structural characterization of the process as extensively discussed by [65] and the distinct energetic contributions associated to ligand binding. To this aim, molecular dynamics (MD) at atomistic resolution [66], where all the atoms of a system are evolved with physics based equations, could provide important insights on RNA conformational changes. It is not

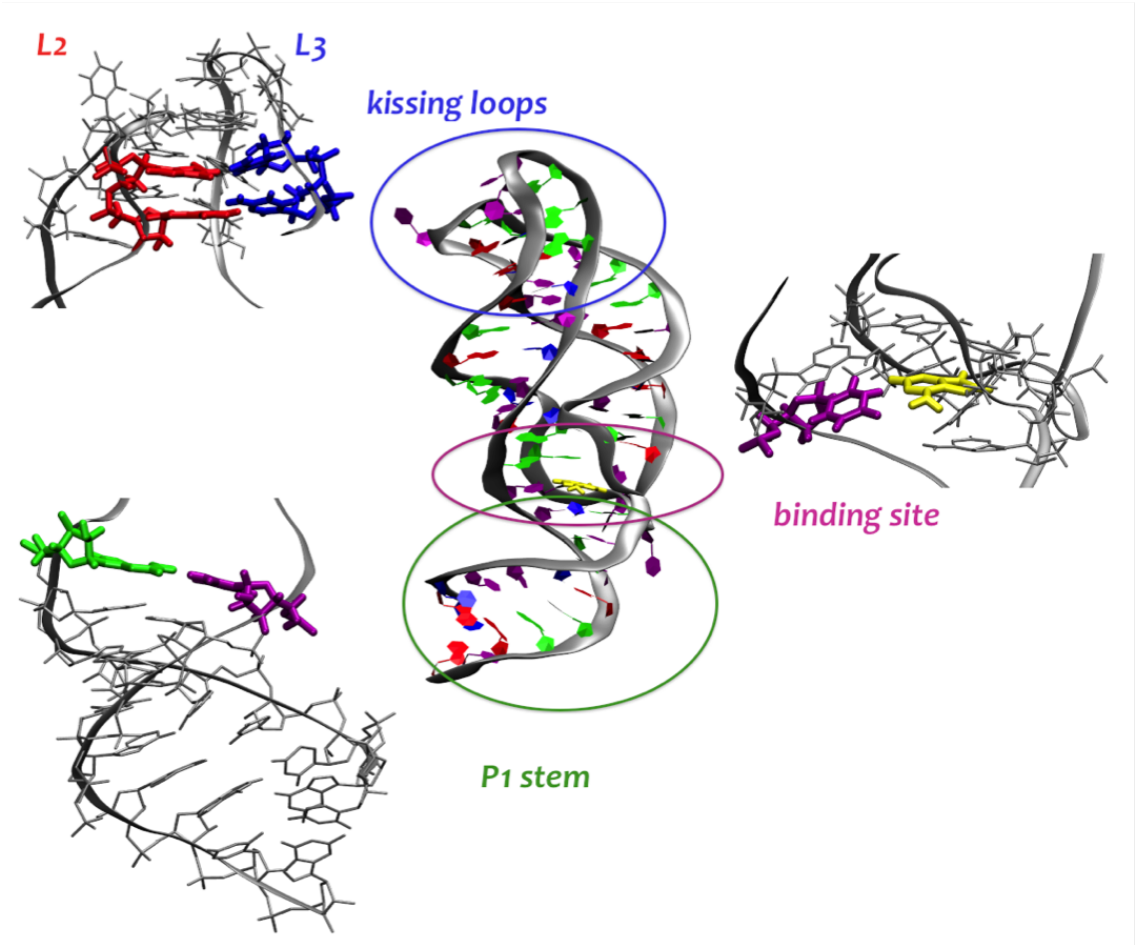

FIGURE 1.6: 3D structure of the riboswitch aptamer bound with adenine (yellow) at nucleotides resolution (G in red, C in blue, A in green, U in purple). Zoom on the three key regions at atomic level.

a static picture of the system in that specific moment in the particular experimental conditions, but it can be considered as a powerful microscope that shows the evolution of the system in time and reveals details not otherwise appreciable. Although *in silico* techniques have been used to investigate the ligand role [67–72] obtaining an accurate description of the system from a structural point of view, a quantitative estimation of the energetic contributions associated to ligand binding, in particular regarding the role of direct P1-ligand interactions and the possible long-range influence of well-structured binding site on the loop-loop interaction, has not yet been provided. Indeed, in a few cases a computational approach has been employed to provide a thermodynamic characterization of the system [52, 67, 71].

1.5 RNA and ions: a too tight relationship

RNA folding into stable tertiary structures is remarkably sensitive to the concentration and type of cations present. Metal ions are required to stabilize RNA tertiary structure and to begin the folding process. An understanding of the physical basis of ion-RNA interactions is therefore a prerequisite for a quantitative accounting of RNA stability. The thermodynamic stability of a folded RNA is strongly intertwined with the counterions and the free energy of this interaction must be accounted for in any realistic RNA simulations. Accurate simulations of RNAs and the understanding of how they reach their thermodynamic lower states are nowadays hot topics, but atomistic simulations that are able to answer these questions remain very difficult to be performed. The high negative charge of an RNA molecule works against its folding into a compact structure. Positive ions promote folding by reducing the repulsion between RNA phosphates, but some ions are much more effective than others at this task. A "dramatic" example is the ability of sub-millimolar Mg^{2+} concentrations to stabilize RNA tertiary structures that are otherwise only marginally stable in the presence of monovalent cation at more than one thousand times that concentration [73]. Since each residue contains an anionic phosphodiester group, the principles of charge neutralization and electrostatic condensation dictate that cations must be closely associated with the polyanionic RNA molecule. In principle, these can be any cationic species, but in general, the condensation layer consists of the abundant surrounding monovalent and divalent ions. In vivo, Mg^{2+} and K^{+} are believed to dominate in this role. Charge neutralization becomes particularly important during the process of RNA folding, as the negatively charged backbones from two or more regions of the primary sequence most come close together in space. Indeed, the phosphate groups are drawn together, further intensifying the negative potential. Consequently, metal ions interact more strongly with folded RNA than unfolded RNA, explaining why the association of metal ions stabilizes RNA secondary and tertiary structure. Without cations to screen these charges, the repulsive forces generated in the close-packed structure would overwhelm the energetically favorable interactions that dictate the proper 3D structure. This is a key step in the assembly of RNA tertiary structure, a contraction or "collapse" of the secondary structure into compact conformations [74]. And this collapse transition depends exclusively on metal ions that screen the negative charge on the phosphate groups (Figure 1.7). Gel filtration, native gel electrophoresis and small-angle X-ray scattering indicate that RNAs form compact but loosely folded intermediates, which undergo further conformational rearrangements before adopting the final tertiary structure [75]. In more details, a common representation of the RNA folding problem in terms of three distinct states. Two main transitions can be described as follows: An unfolded RNA completely lacking base-base hydrogen bonding first forms an intermediate state with only secondary structure; additional tertiary interactions form

the native folded RNA in a second step. RNA secondary structures tend to be so stable that most folding studies examine only this last transition. The ensemble of secondary structures in the intermediate state is usually ill-defined and frequently includes base pairings that must be broken to achieve the folded structure.

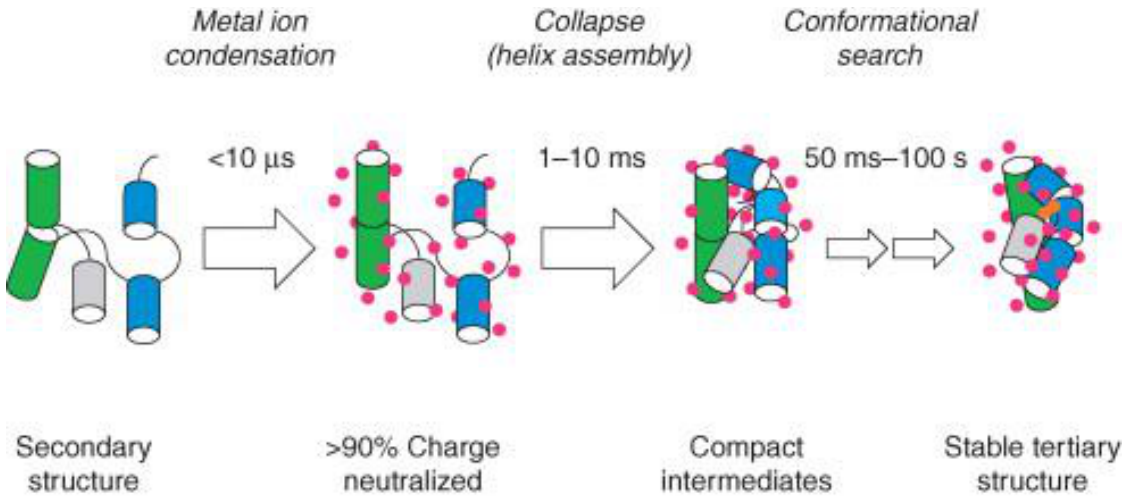

FIGURE 1.7: Metal ion-induced folding of RNA. Association of positive ions (red spheres) with the RNA neutralizes $> 90\%$ of the phosphate charge, inducing collapse of the RNA into more compact conformations. Additional conformational rearrangements lead to the native tertiary structure. Adapted from Woodson [75].

Since there is a formal charge of -1 for every residue, RNAs carry around a sufficient number of metal ions in a condensation layer to neutralize the charge. Studies measuring the number of Mg(II) ions bound to different RNAs confirmed this expectation [76, 77]. The relevant questions concern the way ions influence the equilibrium between intermediates and folded state: To what degree do added ions stabilize the folded RNA relative to the previous state, and why are some ions more effective than others? These are difficult questions to answer at a molecular level. The main reason for this is the different environments that ions associated with an RNA may experience in terms of the extent of direct ion-RNA contacts, the influence of water on the location of the ion, and the degree to which the hydration of the ion is perturbed. Each of these environments requires a different approach in evaluating the energetics of the interaction. Three are the different kinds of ion environments, each of which is dominated by different energetic factors: diffuse ions, water-positioned ion and chelated ions. The right diagram in Figure 1.8 shows an ion completely solvated by water, an example of diffuse ion. The only energetic consideration for these ions is coulombic interaction between ionic charges and the RNA electrostatic field; there are no contacts of the ion with the RNA surface, and the hydration of the ion is not perturbed. These ions have been captured by the RNA electrostatic field but, for the most part, are not confined to a precise location. Spectroscopic studies and thermodynamic data show that the majority of metal ions interact

non-specifically with the RNA's electrostatic field [78]. Understanding the electrostatic forces that drive RNA folding is extremely challenging, because folding is coupled to interactions between the metal ions themselves, with solvent and slightly with the RNA. In this case, each individual counterion is very weakly bound and in rapid exchange with more freely diffusing ions. The middle panel of Figure 1.8 shows a water-positioned ion. These type of ions have a single water layer between themselves and the RNA surface, but they are close enough to the RNA that steric packing and hydrogen bonding characteristics of the hydrating water influence the positioning of the ion. Energetic factors that must be considered here are coulombic interactions between the ion and RNA, perturbation of the ion and RNA hydration layers, and the degree to which water hydrogen bond donors are satisfied. Chelated ions instead, in the left panel of Figure 1.8, make at least two direct contacts with electronegative groups of the RNA. In this case, the free energy required to partially dehydrate both the ion and the corresponding RNA site becomes a major energetic factor. This is, of course, added to the electrostatic interaction between the ion and RNA. How different metal ions enable RNAs to fold depends on the electrostatic potential of the RNA and correlated fluctuations in the positions of the ions themselves [79, 80]. Theoretical models, fluorescence spectroscopy, small angle scattering and structural biology reveal that metal ions alter the RNA dynamics and folding transition states.

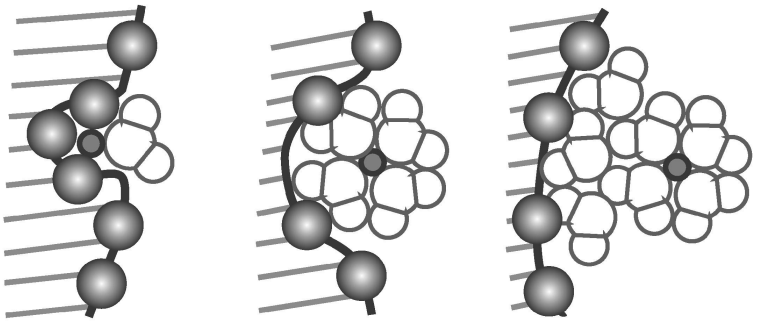

FIGURE 1.8: Environments that might be seen by an RNA-associated ion differ in the extent of direct ion-RNA contacts (left), water-mediated ion-RNA contacts (middle), and hydration (right). The ion is represented by a small circle; water molecules are outlined, and shaded spheres represent RNA backbone phosphates. Adapted from Draper et al. [79].

The availability of accurate ion parameters in our FF is therefore the minimal requirement to perform a reliable MD simulation, especially if the aim of our *in silico* investigation is to determine the different effect in terms of stability given by monovalent and divalent cations to the RNA molecule of interest, the *add* aptamer, and in particular to the formation of its kissing loop interaction. In this framework, a new set of parameters for magnesium has been recently published [81]. This newly developed set has better reproduced the Mg^{2+}-water interaction structural and experimental kinetic data than

existing models showing also an improved description of the Mg^{2+}-phosphate binding. Thus, these parameters make a quantitative comparison between K^+ and Mg^{2+} influence on the adenine riboswitch feasible using advanced sampling MD techniques.

1.6 Outline and aims of this PhD project

Two are the main framework of this thesis that reports the results obtained during a four-year PhD: From the research point of view, the structural and thermodynamics investigation through computational techniques of the *add* adenine sensing riboswitch found in the 5'-UTR of the adenosine deaminase mRNA in *V. vulnificus*. From the didactic point of view, the study and the application of several simulation methods using a quite big and structured RNA domain (compared to the vast majority of the RNA systems studied with molecular dynamics techniques reported in the literature) in order to step-by-step approach biological issues with increasing level of difficulty using sampling techniques with gradually increased complexity, accuracy and efficiency. Riboswitches are known to regulate the genetic expression by means of conformational changes triggered by highly specific interactions of the aptamer with the sensed metabolite. Thus, to understand its folding process it is necessary to deeply typify the behavior of all the different secondary and tertiary structure that are formed in the presence and in the absence of its physiological ligand. This characterization at atomistic resolution allow us to consider also the effects of the different ionic environment around these interactions. At first classical molecular dynamics simulations of the add aptamer in the apo and holo forms have been performed, starting from the experimental crystal structure (ligand-bound). We use steered MD to induce the opening of the P1 stem and investigate its stability. Our results show that the ligand directly stabilizes the P1 stem by means of stacking interactions quantitatively consistent with thermodynamic data [52, 82]. Then, using both umbrella sampling and a combination of metadynamics and hamiltonian replica exchange, we show that the formation of L2-L3 kissing complex cooperates with ligand binding and we quantify the ligand-induced stabilization on this distal tertiary interaction [49, 83]. This long range effect is probably due to a "flow of stability" that from the binding pocket with adenine-bound reaches the two loops through the corresponding stems. Furthermore, in this context the influence given by either the monovalent cations or divalent cations was evaluated, respectively simulating the aptamer in a H_2O-K^+ and H_2O-Mg^{2+} solution [83]. Also in these cases we have found a good agreement with experiments recovering qualitatively the behavior of the system in presence and absence of adenine [49], moreover obtaining reliable thermodynamics data from our simulations [83]. RNA conformational changes at the level of individual base pairing are characterized using a recently introduced technique that is able to distinguish and classify each

interaction (i.e. Watson-Crick base pair, non-canonical bp, stacking) [84]. All the results are compatible with known experimental measurements and shed a new light on the ligand-dependent folding mechanism of the adenine riboswitch giving hints at atomistic details on its several conformational changes. Most of the methods and results presented in this Thesis have been adapted from published or submitted works of our group. In details, some of the methods discussed in Chapter 2 have been partly adapted from Di Palma et al. [52, 82] and Bottaro et al. [84]; the results presented in Chapter 3 from Di Palma et al. [52, 82]. Part of the results presented in Chapter 4 will be published in a paper which is currently in preparation [83] and another part has been already submitted [49].

Chapter 2

Methods and Theory: *in silico* approaches to the adenine riboswitch

2.1 Introduction

The dynamics of tertiary interactions, the terminal helix pairing, and the organization of the binding pocket of the *add* adenine riboswitch are structural changes that one needs to characterize for an in-deep understanding of the behavior of this RNA domain. However, the motion picture that can be inferred from biological assays (e.g. In-line probing, partial RNase T1 cleavage, SHAPE assays) and biopsysics experiments, like real-time multidimensional NMR, single-molecule spectroscopy using optical tweezers and X-ray crystallography, lacks the atomistic details about the multiple conformations sampled during the process. Although *in silico* techniques have been used to investigate the ligand role [67–71], a quantitative estimation of the energetic contributions associated to ligand binding, in particular regarding the role of direct P1-ligand interactions and the contribution of the bound ligand to the stabilization of the loop-loop interaction, has not yet been provided.

To this aim, molecular dynamics (MD) at atomistic resolution [66], where all the atoms of a system are evolved with physics based equations, could provide important insights on RNA conformational changes. In principle, one could imagine to perform a long MD simulation where the riboswitch is initialized in its holo form (ligand bound) and the ligand is subsequently removed. On the appropriate time scale, one could expect the riboswitch to spontaneously convert into the correct apo form (ligand unbound). Unless using a special-purpose molecular simulation machine [85], this is not possible since

the typical timescales required for these conformational changes are on the order of the millisecond, whereas MD is still limited to the microsecond time scale. Yet, the prediction of thermodynamic information such as the ligand-induced stabilization of the holo form is still very difficult with straightforward MD. Several techniques have been developed in the last decades to tackle this issue and to allow changes that should happen on a long time scale to be observed in practical simulation times [86, 87]. These techniques can answer to relevant biological questions: Which are and how strong are the interactions responsible for the ligand-induced stabilization of the terminal helix? Is it possible that the presence of the ligand into the binding site influence the distal kissing interaction, or viceversa?

In this Chapter we explain the theory that rules the *in silico* techniques that we used to carry out the simulations enhancing the sampling of conformational changes in the riboswitch aptamer of our interest: steered MD [88], Umbrella Sampling [89], Metadynamics [90] and Hamiltonian Replica Exchange methods [91]. Furthermore, the analysis methods employed to rationalize the data and provide the structural and thermodynamics properties resulting from our simulations are also described (i.e. Jarzynski equality [92, 93] , reweighing scheme [94], weighted-histogram analysis method (WHAM) [95], error estimate by bootstrapping, baRNAba [84]).

2.2 What is Molecular Dynamics?

One of the principal tools in the theoretical study of biological molecules are the MD simulations. This computational method calculates the time dependent behavior of a molecular system. In practice, the purpose of MD is to compute the positions and velocities of a set of interacting atoms at the present time, quantities that depend also on their values in the previous step. The output of a MD simulation is a trajectory of all the sampled conformations of our set of atoms. The molecular dynamics method was first introduced by Alder and Wainwright in the late 1950's [96, 97] to study the interactions of hard spheres. Many important insights concerning the behavior of simple liquids emerged from their studies. The next major advance was in 1964, when Rahman carried out the first simulation using a realistic potential for liquid argon [98]. The first molecular dynamics simulation of a realistic system was done by Rahman and Stillinger in their simulation of liquid water in 1974 [99]. The first protein simulations appeared in 1977 with the simulation of the bovine pancreatic trypsin inhibitor [100]. Today in the literature, one routinely finds molecular dynamics simulations of solvated proteins, protein-DNA complexes, RNA as well as lipid systems addressing a variety of issues including the thermodynamics of ligand binding and the folding of small proteins. The

number of simulation techniques has greatly expanded. Molecular dynamics simulations generate information at the microscopic level, including atomic positions and velocities. The conversion of this microscopic information to macroscopic observables such as pressure, energy, heat capacities, etc., requires statistical mechanics. Statistical mechanics is fundamental to the study of biological systems by molecular dynamics simulation. In a molecular dynamics simulation, one often wishes to explore the macroscopic properties of a system through microscopic simulations, for example, to calculate changes in the binding free energy of a particular drug candidate, or to examine the energetics and mechanisms of conformational change. The connection between microscopic simulations and macroscopic properties is made via statistical mechanics which provides the rigorous mathematical expressions that relate macroscopic properties to the distribution and motion of the atoms and molecules of the N-body system; molecular dynamics simulations provide the means to solve the equation of motion of the particles and evaluate these mathematical formulas. With molecular dynamics simulations, one can study both thermodynamic properties and/or time dependent (kinetic) phenomenon. Statistical mechanics is the branch of physical sciences that studies macroscopic systems from a molecular point of view. The goal is to understand and to predict macroscopic phenomena from the properties of individual molecules making up the system. The system could range from a collection of solvent molecules to a solvated protein-DNA or -RNA complex (e.g. a ribosome). Few definitions have to be introduced to start this short dissertation on the theory behind a MD simulation. The thermodynamic state of a system is usually defined by a small set of parameters, for example, the temperature, T, the pressure, P, and the number of particles, N. The mechanical or microscopic state of a system is defined by the atomic positions, q, and momenta, p; these can also be considered as coordinates in a multidimensional space called phase space. For a system of N particles, this space has 6N dimensions. A single point in phase space describes the state of the system. An ensemble is a collection of points in phase space satisfying the conditions of a particular thermodynamic state. A molecular dynamics simulations generates a sequence of points in phase space as a function of time; these points belong to the same ensemble, and they correspond to the different conformations of the system and their respective momenta. The molecular dynamics simulation method is based on Newton's second law or the equation of motion, $F = m * a$, where F is the force exerted on the particle, m is its mass and a is its acceleration. From a knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system. Integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. From this trajectory, the average values of properties can be determined. The method is deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any

time in the future or the past. Molecular dynamics simulations can be time consuming and computationally expensive. However, computers are getting faster and cheaper. Simulations of solvated proteins or nucleic acids are calculated up to the microsecond time scale, however, simulations into the millisecond regime have been reported. If we want to study a particular system by means of a MD simulation, the requirements of a good program to perform the computation and a powerful computer are necessary but not sufficient conditions. In addition we need a good model to represent the interatomic forces acting between the atoms composing the system. Ideally this can be done from first principles, solving the electronic structure for a particular configuration of the nuclei, and then calculating the resulting forces on each atom. Since the pioneering work of Car and Parrinello [101] the development of *ab initio* MD simulations has grown steadily and today the use of the density functional theory [102, 103] allows to treat systems of a reasonable size (several hundreds of atoms) and to achieve time scales of the order of hundreds of ps, so this is the preferred solution to deal with many problems of interest. However quite often the spatial and/or time-scales needed are prohibitively expensive for such *ab initio* methods. In such cases we are obliged to use a higher level of approximation and make recourse to empirical force field (FF) based methods. If on one hand this allow an improvement of several order of magnitude on the time-scale of our simulations, on the other hand the quality of a FF needs to be assessed experimentally. But what is a FF in practice?, and which are the main terms entering into the description of a standard FF? A force field is a mathematical expression describing the dependence of the energy of a system on the coordinates of its particles. It consists of an analytical form of the interatomic potential energy, $U(r_1, r_2, \cdots, r_N)$, and a set of parameters entering into this form. The parameters are typically obtained either from *ab initio* or semi-empirical quantum mechanical calculations or by fitting to experimental data such as neutron, X-ray and electron diffraction, NMR, infrared, Raman and neutron spectroscopy, etc. Molecules are simply defined as a set of atoms that is held together by simple elastic (harmonic) forces and the FF replaces the true potential with a simplified model valid in the region being simulated. Ideally it must be simple enough to be evaluated quickly, but sufficiently detailed to reproduce the properties of interest of the system studied. There are many force fields available in the literature, having different degrees of complexity, and oriented to treat different kinds of systems. However a typical expression for a FF may look like this:

$$
\begin{aligned}
V = & \sum_{i}^{n_{bonds}} b_i \left(r_i - r_{i,eq}\right)^2 \\
& + \sum_{i}^{n_{angles}} a_i \left(\theta_i - \theta_{i,eq}\right)^2 \\
& + \sum_{i}^{n_{dihedrals}} \frac{V_i}{2}\left[1 + \cos(n\phi_i - \phi_{i,eq})\right] \\
& + \sum_{i}^{n_{imp-tors}} k_i \left(\xi - \xi_{i,eq}\right)^2 \\
& + \sum_{i,j\neq i}^{non-bonded\ pairs(i,j)} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \\
& + \sum_{i,j\neq i}^{non-bonded\ pairs(i,j)} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}
\end{aligned}
\tag{2.1}
$$

where the first four terms refer to intramolecular or local contributions to the total energy (bond stretching, angle bending, and dihedral and improper torsions) or simply the bonded-terms, and the last two terms serve to describe the repulsive and Van der Waals interactions by means of a Lennard-Jones potential, and the Coulombic interactions, representing the non-bonded terms. Each atom of the molecule of our interest has its own FF parameters and the combination of them describe the Hamiltonian of the system.

2.3 Steered Molecular Dynamics

Steered MD, firstly proposed by Grubmuller et al. [88], can be considered as an *in silico* representation of mechanical manipulation experiments such as pulling with AFM or optical tweezers. However, whereas mechanical manipulation techniques can only be used to pull the distance between two atoms (typically, the end-to-end distance of a polymer), steered MD can be used to pull any conceivable function of the microscopic coordinates. Two main protocols are employed, namely constant force or constant velocity [104]. In the first one, a force is directly applied to one or more atoms, and the extension or the displacement is monitored throughout the simulation. We focus here on the second one, where a moving harmonic potential is used to pull the system along a predefined CV $s(x)$. Here x represents the microscopic coordinates of the system. The bias potential thus reads

$$
V(x) = \frac{k}{2}(s(x) - s_0 - vt)^2 \tag{2.2}
$$

Here k is the spring stiffness, s_0 , the initial position of the restraint and v its velocity. This procedure allows the system to be forced to move along a CV with a known schedule. For example, one can pull a torsional angle to enforce an isomerization process to happen in a prefixed amount of time. This makes the method particularly convenient when a single CV has been identified to be relevant for the process under investigation and limited computational resources are available.

2.4 Jarzynski equality

Once a CV and a pulling protocol able to induce the desired conformational change has been determined, it is necessary to interpret the resulting trajectories. We notice here that typical simulations are in the nanosecond to microsecond time scale. Thus, conformational changes are artificially induced in a time that is orders of magnitude faster than the natural time required for a spontaneous process. Special analysis techniques are required to recover the correct thermodynamic properties from a simulation that is very far from thermodynamic equilibrium [105].

If pulling is done in a quasi-static limit, the work performed corresponds to the free-energy change. In the more realistic case of finite velocity the work provides an upper estimate of the free-energy change. Since steered MD is a stochastic process, pulling should be performed several times and multiple trajectories should be averaged. The Jarzynski equality [92] looks as follow:

$$e^{-\beta \Delta F} = \langle e^{-\beta W} \rangle \tag{2.3}$$

In the left-hand side ΔF is what we want to know, the difference in free energies between two equilibria with parameter values κ_f and κ_i. In the right-hand side an average over all possible paths that take the system in equilibrium for a certain parameter value κ_i in its initial Hamiltonian to a state where that parameter is changed into κ_f.

Accordingly, the free-energy change can be obtained from an exponential average of the work

$$e^{-\beta \Delta F} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} e^{-\beta W_i} \tag{2.4}$$

Here N is the total number of trajectories, W_i the work performed on i-th trajectory, $\beta = 1/k_B T$ the inverse of the thermal energy and k_B is the Boltzmann constant. This

identity is exact only in the limit of infinite number of realizations. Averages computed with a finite number of trajectories are known to overestimate the free-energy change as well [106]. However, this expression is very useful since it can correct for the presence of unphysical trajectories in the ensemble. Those are indeed expected to exhibit a large performed work and thus contribute to the exponential average with a negligible weight.

We remark that the value of the work can also be used as a guideline for the choice of the CV: if two different CVs are capable to induce the same conformational change, the one for which the work is lower is the one that allows a transition to happen closer to equilibrium. With such a CV the number of trajectories required to converge the results is expected to be lower.

2.5 Reweighting scheme

The Jarzynski relationship allows one to pull a CV and compute the free-energy landscape projected on that CV. We have recently shown [94] that the scheme can be modified so as to allow the calculation of the free-energy landscape projected on another, a posteriori chosen CV. In practice, using this method one can, for instance, enforce an easy-to-pull distance and then estimate the free-energy difference of the process by looking at the rupture or formation of interactions among neighboring residues or at changes in the number of surrounding water molecules so as to quantify solvation.

The method is based on the combination of an identity first suggested by Jarzynski [93] and of WHAM (see also the dedicated section below) [95]. It is also related to a method introduced by Hummer and Szabo [107]. The algorithm allows weights to be computed for the configuration at time t from i-th trajectory as

$$w_i(t) = \frac{e^{-\beta(W_i(t)-F(t))}}{\int e^{-\beta(V(x_i(t),t')-F(t'))}dt'} \tag{2.5}$$

Here $W_i(t)$ is the work performed up to time t on the i-th replica and $F(t)$ is a term that corresponds to the free energy of the restrained system at time t and can be obtained solving

$$e^{-\beta F(t)} = \sum_i \int w_i(t')e^{-\beta V(x_i(t'),t)}dt' \tag{2.6}$$

These two equations must be solved self-consistently to obtain the weights $w_i(t)$ of each of the configurations saved along the MD simulations that can then be used to compute the free energy with respect to any a posteriori chosen CV $s'(x)$ as

$$F(s') = -k_B T \log \sum_i \int w_i(t) \delta(s' - s'(x)) dt \tag{2.7}$$

Here the δ function selects all the configurations for which the value of the new CV is exactly s'. In combination with the maximum likelihood approach introduced by Minh and Adib [108], this method has also been generalized to bi-directional pulling [109].

2.6 Error estimate

To compute the error in the estimated free energies one can use a bootstrapping procedure that takes advantage of the fact that several independent estimates of the same value are computed. Given N the number of trajectories to be analyzed, the simplest approach consists in re-sampling randomly N trajectories with replacement from this ensemble. Weighting factors can then be re-computed using only these trajectories in the self-consistent procedure, potentially including duplicate ones. The standard deviation of the results provides an estimate of the error. With this procedure the error can be computed for weights w or directly for any desired observable quantity O. Namely, one can compute the free-energy difference between two relevant metastable conformations for every bootstrap sample and compute the error from their standard deviation.

We notice that the combination of the trajectories here is a non-linear procedure, so that the average of the bootstrap samples is not necessary identical to the result obtained including all the trajectories. We thus here define the error as

$$\varepsilon = \sqrt{\frac{1}{M-1} \sum_{m=1}^{M} (O_m - O_{tot})^2} \tag{2.8}$$

Here M is the number of bootstrap samples, O_m the observable computed from the m-th bootstrap sample, and O_{tot} the observable computed using all the available trajectories.

2.7 Umbrella Sampling and WHAM

Umbrella sampling is the standard way of using non-Boltzmann sampling to overcome free energy barriers. When Torrie and Vallau in 1977 came out with their work [89], umbrella sampling used a function $w(x)$ that weights hard-to-sample configurations, equivalent to adding a bias potential of the form

$$\Delta V(x) = -k_B T \ln w(x) \tag{2.9}$$

w is found by trial-and-error such that easy to sample configurations on the unbiased potential remain easy to sample; w acts like an "umbrella" covering also the hard-to-sample regions of configuration space. w is an explicit function of the CVs, $w(x) = W[s(x)]$. Obtain a single umbrella sampling simulation that takes the system far from its initial point is not straightforward because coming up with the umbrella potential that would enable exploration of CV space is quiet difficult, especially for complex and big systems. It is however possible and advantageous to combine results from several independent trajectories, each with its own potential that that covers a small volume of CV space that overlaps with nearby volumes of the "neighbor" trajectory. The most popular way to combine the statistics of such a set of independent umbrella sampling runs is the WHAM. To compute statistics of CV space using WHAM at first it is necessary to choose the points in CV space that define the little local "windows" to be sampled and the relative bias potential used to restrain the sampling. Not knowing how the free energy changes in CV space makes the first task somewhat challenging, since more densely packed windows are preferred in regions where the free energy changes rapidly. Without any *a priori* knowledge about the free-energy surface as a function of the chosen CV it is thus difficult to apply straightforwardly an optimal protocol; since the simulations are independent, more bias can be added later if it is necessary to obtain convergent free-energy calculation. A convenient choice to apply the US bias potential is a simple harmonic spring that restrains each trajectory to the choosen reference point s_i in CV space:

$$\Delta V_i(x) = \frac{k}{2}|s(x) - s_i|^2 \tag{2.10}$$

The points s_i and the value of k must be chosen such that $s(x)$ from any single trajectory makes excursions into both its nearest neighbors windows in CV space in order to have an overlap in the sampled CV space between adjacent probability distributions. From the different windows, a histogram is realized for each restrained trajectory producing the biased statistics on $s(x)$. Now the WHAM is deputed to reconstruct the total unbiased statistics $P_0(s)$ summing the single window histograms:

$$P_0(s) \propto \left(\frac{\sum_i N_i(s)}{\sum_i e^{-\beta[\Delta V_i(s) - f_i]}} \right) \tag{2.11}$$

where the corrective term f is treated as a constant for each window, then solving it for specific values such that the single unbiased probability $P_0(s)$ results to be continuous across the overlapping regions

$$e^{-\beta f_i} = \int ds\ e^{-\beta \Delta V_i(s)} P_0(s) \tag{2.12}$$

Thus, this last two formulas has to be iteratively solved in order to obtain an estimate of the free-energy:

$$F(s) = -k_B T log P_0(s) \tag{2.13}$$

2.8 baRNAba

To define the number of stacking interactions and the number of base-pair contacts a local coordinate system was constructed in the centre of each six-membered rings, with the x axis pointing towards the C2 atom and the z axis orthogonal to the ring plane (Figure 2.1A). The relative position and orientation between two nucleotides j and k is based on the vector r_{jk}, i.e. the position of ring center k relative to the coordinate system constructed on base j. Following this definition, r has an intuitive interpretation in terms of base-stacking and base-pairing interactions, indeed it well describe the the interactions between two bases. Conveniently it is expressed in cylindrical coordinates ρ, θ and z as shown in Figure 2.1B. Note that r is invariant for rotations around the axis connecting the 6-membered rings.

The criteria for determining the canonical WC base pairs are:

1) the base pair must be AU or GC;
2) The relative position of the bases is compatible with the geometry of a WC interaction.

The latter condition is considered satisfied when the product of the Gaussian function $\mathcal{N}(r_{jk}; \mu, \sigma) \times \mathcal{N}(r_{kj}; \mu, \sigma) > 10^{-8}$. Mean μ and covariance σ were obtained from the empirical distribution of WC pairs in the crystal structure of the large ribosomal subunit [110]. The criteria for determining the non-canonical base pairs are:

1) the ellipsoidal distance $\mathcal{D}_{jk} \equiv \sqrt{x_{jk}^2/25 + y_{jk}^2/25 + z_{jk}^2/9} < \sqrt{2.5}$, and $\mathcal{D}_{kj} < \sqrt{2.5}$;
2) $|z_{jk}|$ and $|z_{kj}| < 2$ Å;
3) it is not a WC pair.

The criteria for determining the stacking base pairs are:

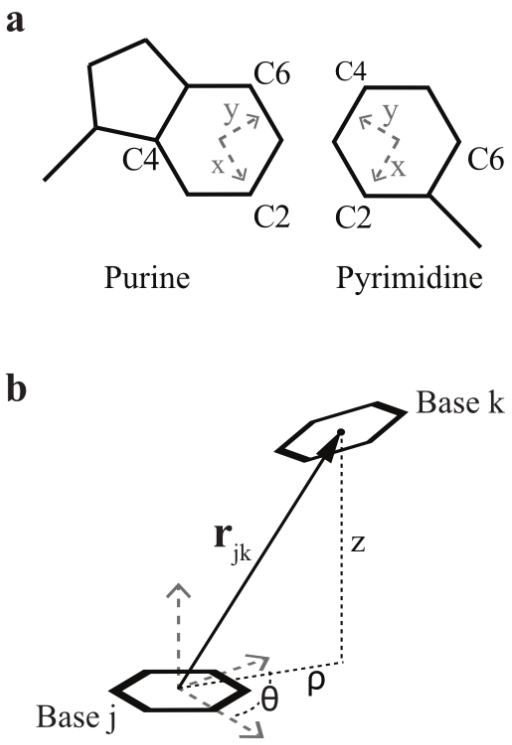

FIGURE 2.1: (a) Definition of the local coordinate system for purines and pyrimidines. (b) The vector $\mathbf{r}_{jk}$ describes the position of base ring k in the coordinate system constructed on base ring j and is expressed in cylindrical coordinates ρ, θ, z. Adapted from Bottaro et al. [84]

1) the ellipsoidal distance $\mathcal{D}_{jk} < \sqrt{2.5}$, and $\mathcal{D}_{kj} < \sqrt{2.5}$;
2) $|z_{jk}|$ and $|z_{kj}| > 2$ Å;
3) $x_{jk}^2 + y_{jk}^2 < 5$ Å and $x_{kj}^2 + y_{kj}^2 < 5$ Å.

This procedure yields similar results compared to the MC-annotate software [111] and was shown to be useful for characterizing both structural and dynamical properties of RNA molecules [84]. The software used to perform this structural analysis is available online (http://github.com/srnas/barnaba).

2.9 Metadynamics

Steered MD is suitable for small conformational changes. Sometime the process under investigation is so complex that a single descriptor is not enough to enforce the transition. In this case, metadynamics (MetaD) is a viable alternative to steered MD. In MetaD a history dependent bias potential is built so as to disfavor already visited conformations in the CV space. The bias potential can then be used to estimate the underlying free-energy landscape. espandere Indeed, MetaD is a powerfull algorithm not only for accelerating

rare events in systems having high free-energy barriers that otherwise are difficult to override, but also to reconstruct the thermodynamics of the system of our interest along the chosen CV. In the algorithm the normal evolution of the system is biased by potential contructed on-the-fly during the simulation as a sum of Gaussians centered on the already visited points in CV space corresponding to the sampled conformations, so as to follow the behavior of the CV. Thus, the potential is updated by adding a new Gaussian with a certain deposition rate (frequency). This means that at time t, the "metadynamics" potential acting on the system $S(x)$ is given by

$$V_G[S(x),t] = h \sum_{t'=\tau_G,2\tau_G,\ldots}^{t'<t} \exp\left(-\frac{|S[(x)t] - S[x(t')]|^2}{2w^2}\right) \tag{2.14}$$

where h and w are respectively the height and the width of the Gaussian and t' is time interval between successive potential updates. Larger are the added Gaussians, faster will be the free energy surface exploration, on the contrary smaller are them, longer time the simulation will take to fully visit the landscape even if a more accurate reconstruction will be obtained. MetaD can be used to "escape free energy minima" or to exhaustively explore a predefined region in the CV space. In the first case the task can be easily accomplished with the previous formulation, whereas for the second aim we need something more that makes the algorithm more efficient and less time-demanding to reach convergence. The basic assumption of MetaD is that the external potential after a sufficiently long time provides an estimate of the underlying free energy. Namely the simulation time the algorithm takes to fill each minima of the profile, until the motion of the CV becomes diffusive, will be the sufficient time to obtain a converged free energy profile. In a single run it is very long and difficult for V_G to converge, oscillating instead around the best possible estimate. A possible way to avoid this problem and decrease the error is to use the Well-tempered MetaD. It has an optimized schedule to decrease over simulation time the deposition rate of bias by modulating the Gaussian height, such that also the difference between the MetaD estimate and the true free-energy, related to the bias growth rate, decrease [112]. In practice the WTMetaD is implemented by rescaling the Gaussian height as follows

$$h = h_0 \tau G e^{-\frac{V_G(S,t)}{k_B \Delta T}} \tag{2.15}$$

Here, h_0 is the previous Gaussian height and ΔT is an input parameter with the dimension of a temperature that reflects the biased trajectory propensity to make excursion away from the free-energy minimum from the point in CV space in which the simulation is at the current step. This algorithm improvement, from the efficiency and free-energy

convergence point of views, lead to a bias deposition rate decrease proportional to $\frac{1}{t}$ and to a progressively closer dynamics of the microscopic variables to the thermodynamic equilibrium. Using the WTMetaD the bias potential does not converge to the negative of the free-energy but to a fraction of it, not compensating the free energy surface

$$V_G[S(x),t] = \Delta T \ln\left(1 - \frac{hN(S(x),t)}{\Delta T}\right) \tag{2.16}$$

$$P(S(x),t) \to e^{-\frac{F(S(x))+V_G(S(x),t)}{T}} \to e^{-\frac{F(S(x))}{T+\Delta T}} \tag{2.17}$$

$$-\frac{T+\Delta T}{\Delta T} V_G(S(x),t) = F(S(x)) \tag{2.18}$$

This results in sampling the CVs at a higher effective temperature $T + \Delta T$. This parameter can be tuned to decide the size of the region of the landscape we would explore, if $\Delta T > 0$ we have the equilibrium MD, whereas if $\Delta T > \infty$ we recover the standard MetaD. The introduction of this history-dependent potential modify the probability distribution of the biased CVs, but it can be easily reconstructed. It is also possible to recover equilibrium Boltzmann probability distribution for other unbiased degrees of freedom [113]. One of the good features of MeatD is that it is relatively easy to detect errors and missing CVs. MetaD is a powerful tool as it allows to embed the a priori knowledge about the conformational change in the selection of the CVs. However, sometime the procedure of looking for good CVs is frustrating. Moreover, in complex cases, a list of a few CVs able to completely describe a conformational change just does not exist. A possible solution is to use approaches where a large number of CVs can be simultaneously or independently biased, such as bias-exchange MetaD [114] or temperature accelerated MD [115, 116].

2.10 Hamiltonian Replica Exchange

An alternative is to combine MetaD with the REST2 replica exchange variants [117]. A combination like this one, but with parallel tempering [91, 118], has been first proposed in Bussi et al. [119] greatly improving the performance of both the techniques. In particular, combining them the high free-energy regions, which are not accessible with conventional parallel tempering, are sampled and in such a way it is possible to characterize the most populated configurations in the relevant free-energy basins from both the structural and thermodynamic point of views. At variance with this inspiring

work, we perform multiple WTMetaD simulations using different Hamiltonians (instead of different temperatures), periodically allowing exchange of configurations according to a replica exchange scheme. Indeed, the difference between the replicas is not restricted to be a change in temperature like in parallel tempering. Any control parameter can be changed, and even the expression of the Hamiltonian can be modified [91]. Indeed, in the Hamiltonian replica-exchange (HREx) methods every replica is simulated according to a different Hamiltonian but at the same temperature. Two are the main advantages choosing this method: A smaller dependence of the results on the choice of the rescaled Hamiltonians than, for example, the dependence of US efficiency on the choice of the CVs; A smaller number of replicas than parallel tempering is required, signature of an increased efficiency. The acceptance for a coordinate swap between two replicas, r_i and r_j, reads:

$$\alpha = min\left(1, \frac{e^{-\left(\frac{U_i(r_j)+U_j(r_i)}{k_B T}\right)}}{e^{-\left(\frac{U_i(r_i)+U_j(r_j)}{k_B T}\right)}}\right) \tag{2.19}$$

where U is the potential energy of a replica defined as the sum of the force-field potential, thus including all the effects of the rescaled parameters (see ahead).

In the HREx, using the same temperature for all the replicas, $P(r_1) \times \cdots \times P(r_N) \propto e^{-\frac{U_1(r_1)}{k_B T} - \ldots - \frac{U_N(r_N)}{k_B T}}$. This means that the ensemble probability only depends on $\frac{U}{k_B T}$ and a double temperature is completely equivalent to a halved energy. Hamiltonians scaling however can take advantage of the fact that the total energy of the system is an extensive property. Thus, one can limit the scaling to the portion of the system which is considered to be interesting and which has the relevant bottlenecks. From the technical point of view the system has to be splitted into two regions, the one in which we would accelerate the CVs space exploration that can be named $\mathfrak{H}$ ("hot"), and the rest of the molecule in the $\mathfrak{C}$ region ("cold"). Then, a parametrized Hamiltonian $\mathcal{H}$ has to be defined $\mathcal{H}(\lambda), 0 \leq \lambda \leq 1$, where 0 means that force field terms contributing to the energy barriers become infinite and 1 leaves $\mathcal{H}$ without any reparametrization. Processing the different topologies of the different replicas the proposed implementation [120] scales:

The charge of atoms in $\mathfrak{H}$ by a factor $\sqrt{\lambda}$;
The $\varepsilon(L-J)$ of atoms in $\mathfrak{H}$ by a factor λ;
The proper and improper dihedral potentials for which both the 1^{st} and 4^{th} atoms are in $\mathfrak{H}$ by a factor λ;
The proper and improper dihedral potentials for which either the 1^{st} or the 4^{th} atom are in $\mathfrak{H}$ by a factor $\sqrt{\lambda}$.

The reparametrization results in some crucial effects on the interactions between the atoms of the system:

Inside $\mathfrak{H}$ they are kept at an effective temperature $\frac{T}{\lambda}$;
Between the two regions they are kept at an effective intermediate temperature $\frac{T}{\sqrt{\lambda}}$;
All the interactions inside $\mathfrak{C}$ are kept at temperature T.

The free-energy landscape is filled in parallel in every replica, and in the CVs space thanks to the MetaD the dynamics of the system becomes diffusive with a smaller computational time expense. This greatly improves the capability of HREx to explore low probability regions, leading to a more reliable estimate of the height of the relevant free-energy barriers. Furthermore, it allows sampling of the degrees of freedom not explicitly included in the CVs, improving thus WTMetaD accuracy.

Considering the combination of the two techniques the acceptance for a coordinates swap between two replicas has to take in account also the "metadynamics" bias. Thus, to be accepted an exchange has to satisfy this condition:

$$\alpha^{(HREx-WTMetaD)} = min\left(1, \frac{e^{-\left(\frac{U_i(r_j)+V_{Gi}(r_j)}{k_BT}+\frac{U_j(r_i)+V_{Gj}(r_i)}{k_BT}\right)}}{e^{-\left(\frac{U_i(r_i)+V_{Gi}(r_i)}{k_BT}+\frac{U_j(r_j)+V_{Gi}(r_j)}{k_BT}\right)}}\right) \tag{2.20}$$

Chapter 3

Aptamer terminal helix ligand-induced stabilization

3.1 Introduction

It has been proposed that P1 is stabilized by the ligand [50] and that this could be a common feature in many riboswitch classes [61]. The role of ligand binding in the structural organization of the aptamer has been investigated with single-molecule spectroscopy providing an insightful overview on the folding dynamics [64], yet lacking the critical atomistic details needed for an accurate structural characterization of the process as extensively discussed by [65]. Although *in silico* techniques have been used to investigate the ligand role [67–71], a quantitative estimation of the energetic contributions associated to ligand binding, in particular regarding the role of direct P1-ligand interactions, has not yet been provided. In this context, state-of-the-art free-energy methods combined with atomistic simulations can bridge the gap providing an unparalleled perspective on the mechanism and dynamics of the biomolecular process of interest [121]. To tackle this issue we used SMD simulations to study the P1 stem formation of the adenine riboswitch aptamer in the presence and in the absence of the cognate ligand. Similarly to other methods, steered MD acts on an a priori chosen collective variable (CV). We present and compare two applications where the steering is applied on CVs of growing complexity. First, the full terminal helix is disrupted pulling its termini and looking at the effect of the ligand. This study can provide a qualitative insight on the ligand role. Second, we analyze the ligand-induced stabilization of the first base pair (bp) of the terminal stem, the A9-U63 bp, which directly stacks with adenine and was used as a proxy for the P1 stability. We also describe how data from steered MD can be post-processed to provide quantitative thermodynamic information. In particular how

applying the Jarzyinski equality and combining the output of the pulling simulations with a reweighting scheme that has been recently developed in our group [94], allow us to obtain a quantitative measurement of the energy of the system. Our non-equilibrium simulations provide measurements of the stability of the A9-U63 bp and quantify the direct ligand-dependent stabilization of the pairing. In the following our results are presented and compared with melting and single-molecule experiments. A structural model for the conformational switch emerging from the combination of our results and previous experimental data is also discussed.

3.2 System Setup and Protocols

Several techniques can be used to accelerate MD. In some of them a priori information about the process under investigation is required. These methods require the definition of descriptors named collective variables (CVs), which are arbitrary functions of the microscopic coordinates. A bias potential is then added to modify the observed probability distribution of these CVs. If these CVs include the degrees of freedom that are responsible for the conformational change under investigation, a bias can be built that favors the transition state and thus greatly enhances the transition rate. CVs should thus not only be able to distinguish reactants from products but also to properly drive the system through the transition. As an example, a natural CV to describe a binding process could be the distance between two molecules. However, if the binding involves some important rearrangement it will be difficult to accelerate it by just biasing this CV. The comparison of several CVs and the choice of a relevant one is itself an active component of the investigation process that helps understanding the dynamics of the system of interest. Methods where a bias potential acts on a preselected CV have their root in the umbrella sampling method [89]. Some examples are steered MD [88], adaptive biasing force [122], and metadynamics [90]. In the following sections we describe the actual protocol used to prepare and analyze our simulations. We carried out the simulations of the aptamer domain of the *add* A-riboswitch in different forms, namely the entire aptamer (PDB id 1Y26) [51] has been simulated in the presence (Holo) and in the absence (Apo) of the cognate ligand, the adenine; the Apo form was generated by adenine removal from the ligand-bound crystal structure: this deletion is justified by the fact that the Apo and Holo form have been shown experimentally to share an overall similar secondary structure [123]. This is at variance with e.g. the *pbuE* adenine riboswitch in which the two structures are different. The generation of the Apo form by simply removing the ligand has been adopted also in other recent works [71] [ALTRE CITAZIONI]. Additionally, to better estimate the ligand-induced stabilization, we also simulated a truncated aptamer ($\triangle$1-8/64-71), both in the Apo and Holo forms. The first

eight bp of the P1 stem (i.e. whole stem from C1 to U8 and from A64 to G71, except for A9-U63 bp) were cut in both systems. This deletion is not biologically meaningless, because it has been shown experimentally that a series of aptamer variants with shorter P1 helix are still able to bind the ligand [124]. The deletion reduced the noise during the pulling allowing to focus the calculation on the influence of the ligand on the A9-U63 pairing.Long unbiased molecular dynamics (MD) for all the four systems were performed to test the stability of the aptamer in different conditions. In the truncated systems the terminal bp was restrained in its initial configuration to mimic the presence of the rest of the stem. Furthermore, the full-length systems were pulled from the terminal bases to disrupt the entire P1 stem (Figure 3.1) thus allowing its different stability between in the Holo and in the Apo forms to be qualitatively inferred. At last, to quantify this difference, SMD simulations of both the △1-8/64-71 systems were done enforcing the breaking of the A9-U63 bp that directly stacks with the ligand (Figure 3.2).

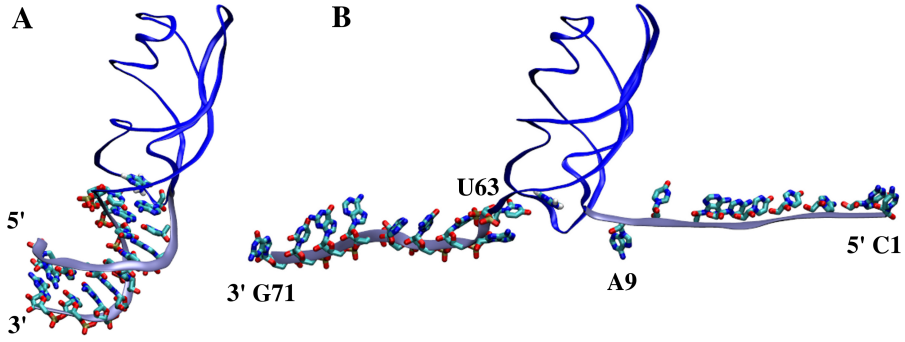

FIGURE 3.1: Initial (A) and final (B) configuration of the SMD simulation opening the P1 stem here shown for the Holo form. The backbone of the aptamer is in blue except for the P1 stem, in light blue. The ligand and the 18 bases forming the helix are shown. The P1 stem is formed in A and disrupted in B. Adapted from Di Palma et al. [52]

We remind that this is not the case when the expression platform is present, since the riboswitch is expected to undergo a conformational change towards the ON state in the presence of the ligand.

Molecular dynamics simulations were performed using the Amber99 force field [125] refined with the *parmbsc0* corrections [126] and the more recent refinement on the χ torsional parameters [127]. Adenine was parametrized using the general Amber force field (gaff) [128]. Partial atomic charges were assigned using the restricted electrostatic potential fit method [129] based on an electronic structure calculation at the HF/6-31G* level of theory performed with Gaussian03 [130]. Bond-lengths were constrained with LINCS [131] and the electrostatic interactions were calculated using the particle-mesh Ewald method [132]. For both forms, the following protocol was used to prepare the systems (Table 3.1 for details) for MD simulations: steepest descent minimization (200 steps) starting from the X-ray structure. Solvation with ≈13000 TIP3P water molecules

[133] and NaCl at 0.15 M concentration (plus extra Na^+ counter-ions to neutralize the charges of the systems) in a hexagonal prism (lattice vectors in nm [(10,0,0), (0,7,0), $(0,\frac{7}{2},\frac{7\cdot\sqrt{3}}{2})$]) that was created orienting the major length of the aptamer along the X axis. Steepest descent minimization (200 steps) for ions and solvent; the systems were thermalized at 300 K, initially for 200 ps with frozen solute positions and then for 5 ns in NPT ensemble (1 atm) with stochastic velocity rescaling [134] and Berendsen barostat [135]; to maintain the systems oriented along the largest lattice vector (X) a restraint was imposed with a force constant of $4{\cdot}10^3$ $(kJ/mol)/nm^2$ on the Y and Z components of the distance between phosphate atoms of A52 and G71. Each system was simulated for 48 ns in NVT ensemble to assess the stability of the aptamer.

TABLE 3.1: Specifications for the simulated systems about the P1 stem investigations

System	Total atoms	Water molecules	aptamer atoms	Ions (Na+Cl)
Holo form	41628	13078	2257	122 (96+26)
Apo form	41676	13099	2257	122 (96+26)
Holo (P1-SMD)	120654	39364	2257	290 (180+110)
Apo (P1-SMD)	120588	39347	2257	290 (180+110)
Holo Δ1-8/64-71	41101	13078	1746	106 (80+26)
Apo Δ1-8/64-71	41149	13099	1746	106 (80+26)

3.2.1 Steered Molecular Dynamics

A first set of steered MD simulations aimed at inducing the opening of the whole P1 stem was performed in a larger rhombic dodecahedral simulation box, solvating again the system with ≈39500 water molecules in a larger rhombic dodecahedral box with distance between periodic images equal to 12 nm, adding ions to maintain the same ionic strengh (P1-SMD systems in Table 3.1). After another equilibration, a steered MD was performed using as a CV the distance between the centers of mass of the terminal nucleotides (C1 and G71). By incrementing this distance the rupture of the P1 stem is enforced. In our protocol we pull it from 1.05 to 12.3 nm. The latter value is sufficient to completely unfold the 9 bps of the helix. Two different velocities were used to evaluate the influence of the pulling velocity, namely 0.5625 nm/ns (fast, corresponding to a 20 ns simulation) and 0.225 nm/ns (slow, corresponding to a 50 ns simulation). In both cases, the spring constant was chosen based on CV fluctuations in a free simulation and set to $7.6{\cdot}10^4$ $(kJ/mol)/nm^2$. A second set of steered MD simulations aimed at inducing the rupture of the A9-U63 pair was performed in the smaller hexagonal prism simulation box. The first eight bps of the P1 stem (i.e. from C1 to U8 and from A64 to G71) were cut in both systems so as to reduce the noise during the pulling and to allow the calculation to be focused on the influence of the ligand on the A9-U63 pairing. Water molecules were allowed to relax filling the space left by the 16 removed bases through

an additional 1 ns equilibration in which the positions of aptamer atoms were frozen followed by 5 ns of unrestrained NPT simulation. Then the systems were simulated for 48 ns in the NVT ensemble restraining the terminal bases in the initial state to avoid any spontaneous flipping. A moving harmonic restraint on the root-mean-square deviation (RMSD) from the reference crystal structure was used to induce the A9-U63 pairing rupture. This bp is the one directly interacting with the ligand when it is present in the binding site. The RMSD was chosen as a CV because it can unambiguously discriminate between the native conformation (low RMSD) and any other possible conformation. The steered CV was pulled at constant velocity of 0.175 nm/ns from 0 to 0.35 nm of RMSD in 2 ns setting the spring constant to $3.9{\cdot}10^4$ $(kJ/mol)/nm^2$. Such a pulling scheme induced the complete unbinding of the A9-U63 bp in the presence and the absence of the ligand (see Figure 3.2). Initial conformations were taken from a restrained ensemble at equilibrium. We here chose equally spaced starting frames from a preliminary 8.192 ns simulation where the RMSD was restrained at 0. For the two system 512 independent SMD simulations were performed, corresponding to an aggregate time of approximately 1μs each.

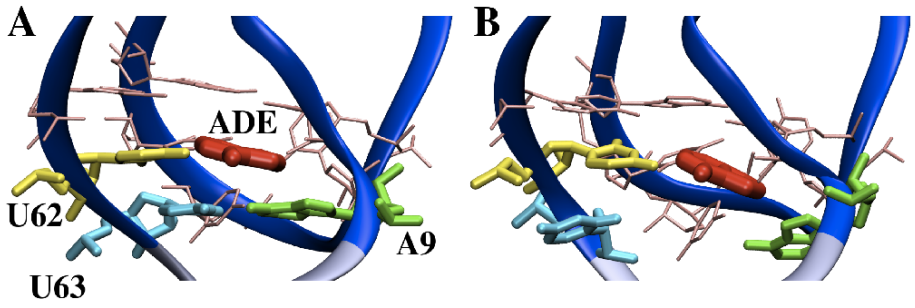

FIGURE 3.2: Representative structures of the Holo binding pocket at the beginning (A, RMSD = 0 nm) and at the end (B, RMSD = 0.35 nm) of the SMD. The portion of the P1 stem removed in our simulations is in light blue. Bases forming the binding pocket are labeled, ligand is shown in red. A9-U63 pair is formed in A and disrupted in B. Adapted from Di Palma et al. [52]

3.2.2 Sample input files

For the pulling simulations we used the PLUMED plugin [136, 137]. We provide here sample input files for the latest PLUMED version (2.1). This is a sample input file for the pulling simulations where the entire P1 stem is disrupted by pulling the distance between the terminal nucleotides:

```
# make the molecule whole across periodic boundary condition
WHOLEMOLECULES ENTITY0=1-2257
# compute center of mass of nt C1
c1: COM ATOMS= 1-29
```

```
# compute center of mass of nt G71
c2: COM ATOMS= 2223-2257
# compute the distance between the two nt
d: DISTANCE ATOMS=c1, c2 NOPBC
# apply a moving restraint (fast pulling)
r: ...
MOVINGRESTRAINT
ARG=d KAPPA0=7630 AT0=1.05 AT1=12.30
STEP0=0 STEP1=10000000
# the line above should be replaced with this for slow pulling
# STEP0=0 STEP1=25000000
...
# print the results for subsequent analysis
# namely: distance, center of the restraint, performed work
PRINT ARG=d, r.d_cntr,r.d_work FILE=COLVAR
```

This is a sample input file for the pulling simulations where the A9-U63 bp is disrupted by pulling its RMSD from the native structure:

```
# make the molecule whole across periodic boundary condition
WHOLEMOLECULES ENTITY0=1-1746
# compute the RMSD from the native structure
# ref.pdb only contains atoms from the relevant nucleotides
rmsd: RMSD REFERENCE=ref.pdb TYPE=OPTIMAL
# apply a moving restraint (fast pulling)
r: ...
MOVINGRESTRAINT
ARG=d KAPPA0=39150 AT0=0 AT1=0.35
STEP0=0 STEP1=1000000
...
# additionally, monitor hydrogen bonds count
hb: ...
COORDINATION GROUPA=17,17,20,23 GROUPB=1734,1738,1735,1735
D_0 0.35 R_0 0.0000001 NN 6 MM 12 PAIR
...
# print the results for subsequent analysis
PRINT ARG=d,rmsd.d_cntr,rmsd.d_kappa,rmsd.d_work,hb FILE=COLVAR
```

To choose the starting conformation for multiple steered MD simulations one should perform a restraint simulation replacing the MOVINGRESTRAINT command above

with RESTRAINT:

RESTRAINT ARG=d KAPPA=39150 AT=0

Starting points should be then selected as equally spaced from the resulting trajectory.

3.2.3 Structural analysis

The P1 unfolding in the presence and in the absence of the ligand was analyzed from the structural point of view. To monitor the behavior of the 9 bps forming the stem we computed the RMSD of each bp from its native conformation during the simulation. The value of this RMSD is expected to start from zero and to increment when the specific bp is broken. We also analyzed the trajectories with our recently developed approach (baRNAba) [84] where a local coordinate system is constructed in the center of the six-membered ring of each nucleobase and pairings are identified based on the relative position of the two bases. This procedure allows pairs to be annotated using the Leontis-Westhof nomenclature [138].

3.3 Results

We carried out the simulations of the aptamer domain of the *add* A-riboswitch in different forms, namely the entire aptamer (PDB id 1Y26) [51] has been simulated in the presence (Holo) and in the absence (Apo) of the cognate ligand, the adenine; additionally, to better estimate the ligand-induced stabilization, we also simulated a truncated aptamer (△1-8/64-71), both in the Apo and Holo forms. Long unbiased molecular dynamics (MD) for all the four systems were performed to test the stability of the aptamer in different conditions. In the truncated systems the terminal bp was restrained in its initial configuration to mimic the presence of the rest of the stem. Furthermore, the full-length systems were pulled from the terminal bases to disrupt the entire P1 stem (Figure 3.1) thus allowing its different stability between in the Holo and in the Apo forms to be qualitatively inferred. At last, to quantify this difference, SMD simulations of both the △1-8/64-71 systems were done enforcing the breaking of the A9-U63 bp that directly stacks with the ligand (Figure 3.2). The stability of both the Apo and Holo systems was evaluated monitoring the root mean square deviation (RMSD) from the native structure along 48 ns MD runs (Figure 3.3A-B). Ligand removal did not affect the overall stability of the Apo aptamer in this time-scale, and secondary and tertiary structures were substantially unchanged.

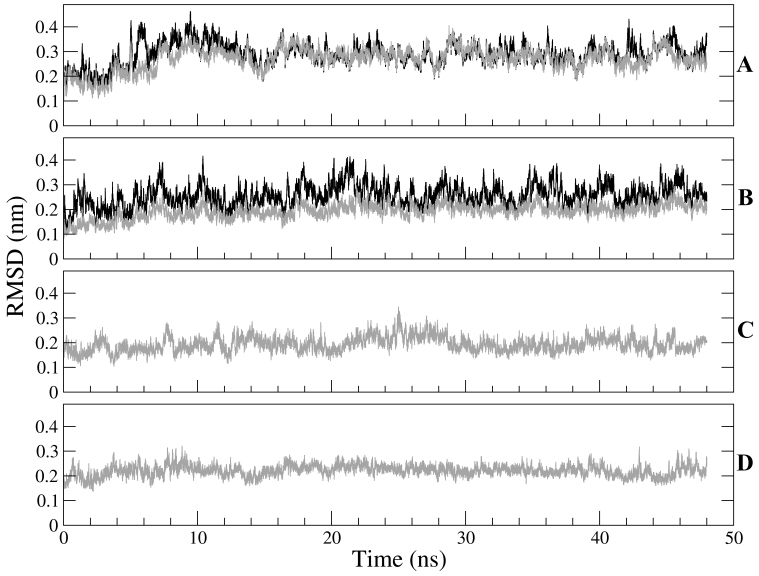

FIGURE 3.3: Root mean square deviation (RMSD) from native structure. A) Holo form during 48 ns equilibration, computed on the whole aptamer (black) and on the bases from 9 to 63 (grey). B) Same as A) done on the Apo form (whole aptamer, black; bases from 9 to 63, grey). The difference between black and grey profiles in both panels, indicates that the P1 stem is less stable than the rest of the aptamer. C) △1-8/64-71 Holo RMSD along the NVT 48 ns equilibration. D) Same as C) for the △1-8/64-71 Apo form. Adapted from Di Palma et al. [52]

3.3.1 Pulling the P1 stem

The analysis of the trajectories obtained by pulling the P1 stem showed that the secondary and tertiary structure elements of the rest of the aptamer were not affected by the opening of the helix. Indeed, few residues have shown larger fluctuations then the other bases. These nucleotides are not paired and exposed to the solvent. We first analyzed the rupture of the P1 stem pulling on the distance between the terminal bases. This was done at two different pulling rates, namely $v = 0.5625nm/ns$ (hereafter called "fast") and $v = 0.225nm/ns$ (named as "slow") for both the apo and the holo form. Here one can qualitatively observe that the unzipping mechanism for the P1 stem is the one discussed recent work of our group [94]. The mechanism for strand separation in the RNA double helix was elucidated, suggesting that the strand separation mechanism occurs by a stepwise process in which the probability of unbinding of the base at the 5' terminus is significantly higher than that at the 3' terminus. This is exactly what have occured during the P1 stem pulling, both in the fast and the slow simulations and regardless to the presence of the ligand. The bases at the 5' end are systematically opened, loosing their interaction with their stacking base, before the ones at the 3' end, without any sequence specificity/dependance. At the end of the simulation the entire P1 stem and part of the binding pocket are disrupted.

The rupture of the individual bps is shown in Figures 3.4 and 3.5 by monitoring the RMSD from native and the occupation of the bp respectively. The bp close to the ligand (A9-U63) was broken at a later stage in the holo form when compared with the apo form. Table 3.2 reports the time corresponding to the definitive disruption of each bp for all the cases. The largest discrepancies between holo and apo can be observed in the slow simulation for the A9-U63 pair confirming that this pair is stabilized by the ligand presence.

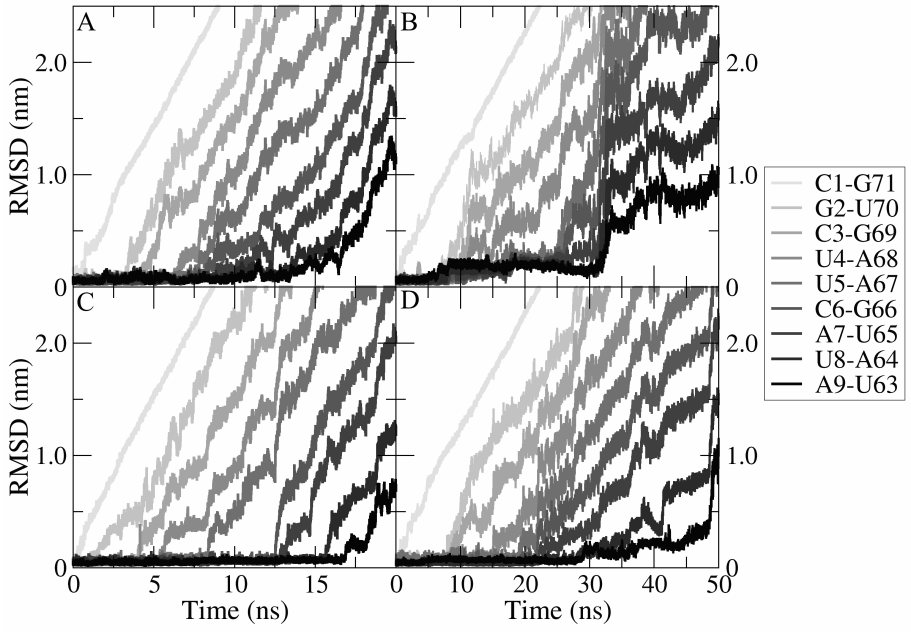

FIGURE 3.4: RMSD from the native conformation of each base pair of the P1 stem (A, apo fast, B apo slow, C holo fast, D holo slow). In the pulling simulations the 9 bp forming the P1 stem were unpaired. We here monitored the RMSD of each bp (grey-scale) from their native conformation and they can be considered disrupted at RMSD > 0.5. In both fast and slow runs, A9-U63 bp (in black) was disrupted at a later stage in the holo form. Adapted from Di Palma et al. [82]

TABLE 3.2: Time corresponding to the definitive disruption of P1 stem base pairs during pulling

bp/ns	fast apo	fast holo	slow apo	slow holo
C1 - G71	0.3	0.3	0.7	1.1
G2 - U70	3.5	2.2	9.0	9.0
C3 - G69	4.5	4.2	9.9	10.0
U4 - A68	6.6	6.0	11.2	16.9
U5 - A67	7.8	8.4	24.5	19.0
C6 - G66	8.7	12.5	25.7	23.0
A7 - U65	10.6	13.2	31	27.5
U8 - A64	14.5	15.7	31.5	35.5
A9 - U63	16.6	18.2	32	44.5

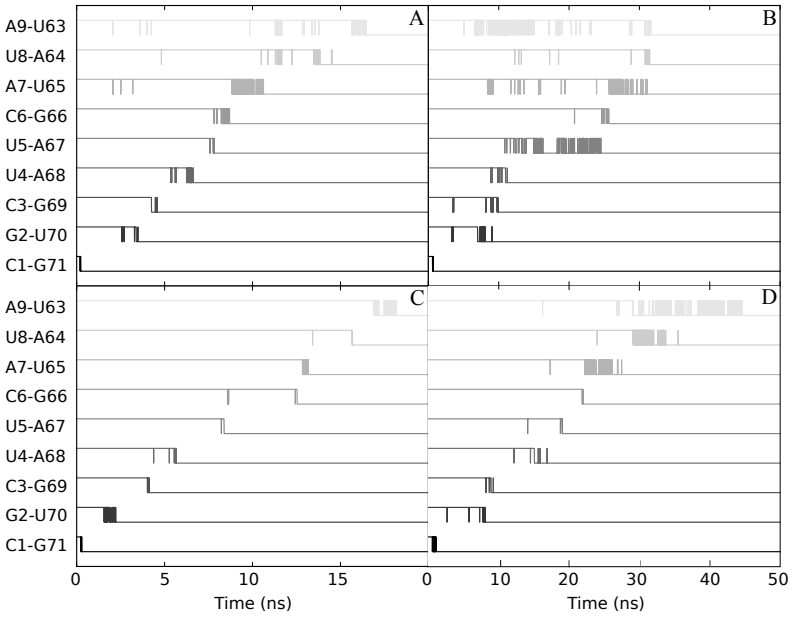

FIGURE 3.5: Monitoring of the P1 stem base-pair ruptures. Individual base pairing of the 9 bp of the P1 stem (grey-scale) along the four simulations has been identified using baRNAba (A, apo fast, B apo slow, C holo fast, D holo slow). Refer to Table 3.2 for more details about the times at which the base pairs are definitively unpaired. Adapted from Di Palma et al. [82]

In Figure 3.6 the work performed to open the P1 stem, with and without the ligand, is shown for fast and slow runs. It can be appreciated that the total work is larger in the fast runs. We remind here that the work provides in general an over estimation of the free-energy change, the difference being the dissipated work. The closer the process is to equilibrium the lower the dissipated work. The slow pulling simulations are thus here closer to equilibrium than the fast ones. However, the pulling velocities employed in our calculations are several orders of magnitude faster than the ones typically employed in mechanical manipulation experiments and the process is highly out of equilibrium also for our slow pulling. In this respect, the values of the work exerted, on the order of hundreds of kJ/mol, have no quantitative meaning and cannot be used to determine reliably the free-energy change.

Apo and holo forms are also compared in Figure 3.6. For fast pulling simulations the apo and holo profiles are hardly distinguishable. At this pulling rate artifacts caused by the highly out-of-equilibrium pulling procedure likely hide the relatively small differences between the apo and the holo system. However, at a slow pulling rate an interestingly informative behavior is observed. The plots of the work exerted on the system in the apo and holo forms are not significantly distinguishable until around the 30^{th} nanosecond of pulling. After this point the two plots clearly diverge and the holo system appears to

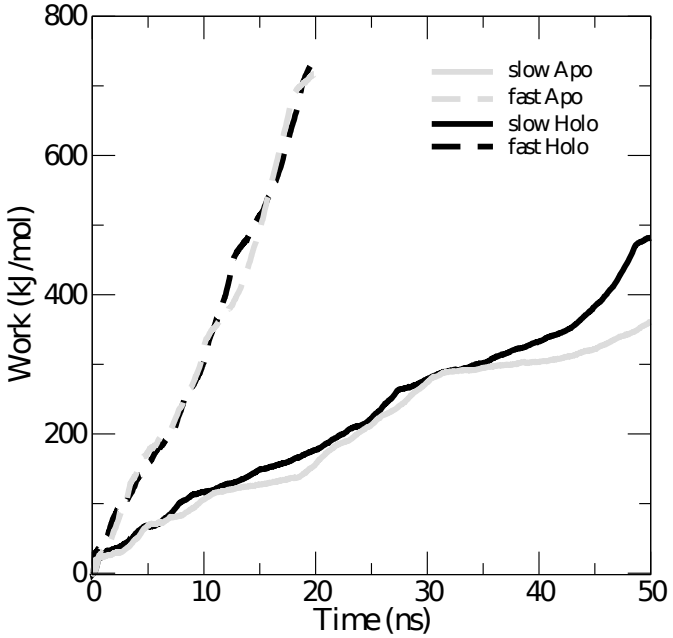

FIGURE 3.6: Mechanical work performed during the unfolding process of whole P1 stem as a function of the simulation time. Profiles for both fast (dashed) and slow (solid) runs are shown. Results for apo and holo form are in grey and black, respectively. Dissipation is larger in the faster simulations. Adapted from Di Palma et al. [82]

be the most stable one. Interestingly, the 30^{th} nanosecond of pulling corresponds to the beginning of the displacement of the bp (A9-U63) that directly interacts with the ligand as shown in Figures 3.4 and 3.5. From this qualitative data, one may speculate that the effect of the ligand stabilization is poorly transmitted along the P1-helix beyond the first interacting bp (A9-U63). This behavior is compatible with the picture in which the ligand stabilizes the P1 stem [61]. It was however difficult to extract quantitative information on the ligand-P1 interaction from these simulations because the rupture is a stochastic event and extensive sampling would be required. Moreover, as pointed out in a recent paper [65], the end-to-end distance could be a non-optimal CV for pulling experiments or simulations since local bp formation plays an important role in global stem folding.

The overall result of these simulations is that the presence of the ligand is stabilizing the P1 stem by means of interactions with its first bp (A9-U63). However, at this level this is only a qualitative indication. The rupture of an RNA helix is a rather complex process. Pulling on the distance between the termini might be effective in experimental settings or in coarse-grain models (Lin and Thirumalai, 2008) when the slow pulling speed allows the process to remain close to equilibrium. On the contrary, in atomistic MD it is definitely unlikely to induce such a complex process in a meaningful manner and at a reasonable time scale. As a matter of fact, due to the complexity of the process

and the inadequate representation of its determinants by one CV, even using steered MD simulations on the time scale of the unbiased transition the process would likely happen out of equilibrium [139].

3.3.2 Pulling the A9-U63 basepair

The pulling of the entire P1 stem suggested a possible role of the ligand in stabilizing the adjacent A9-U63 bp. We here analyze and discuss the results obtained by enforcing the rupture of such a bp directly. We remind that these simulations were performed in an *ad hoc* modified system where the rest of the stem P1 was removed. The quantitative analysis of the P1-ligand interaction was better obtained from the simulation of both the △1-8/64-71 systems. We verified that, when the P1 stem is replaced with the A9-U63 bp restrained to be in canonical WC pairing, the aptamer remains stable (Figure 3.3C-D). Remarkably, fluorescence experiments have shown that the aptamer can also fold and bind adenine when large fractions of the P1 stem are removed [124]. Moreover, the weak interactions between the P1 stem and the rest of the aptamer, but for the A9-U63 bp, allowed the first 8 bps to be truncated. This validates the possibility of using the two structures, △1-8/64-71 Holo and △1-8/64-71 Apo, to investigate the direct P1 stabilization given by the adenine. In the following we focus on the SMD simulations performed on these truncated forms. Typical initial and final conformations from the SMD are shown in Figure 3.1.

3.3.2.1 Analysis of work profiles

The unbinding event of the A9-U63 bp is described as a function of the value of the steered RMSD in Figure 3.7. The initial value corresponds to the configuration with the WC pairing formed, whereas at the final value (0.35 nm) the pairing is completely broken. Even if the ensembles of 512 work profiles for the two forms are broadly spread and superposed, the free-energy profiles computed using the Jarzynski equality [92] as the exponential average of the two sets of data are clearly distinguishable (Figure 3.7A).

Qualitatively it is worth highlighting that during the breaking of the A9-U63 bp the Apo form (red line) is always lower in free energy than the Holo form (blue line). It follows that the breaking of the monitored bp in the Apo form was unambiguously more probable than in the Holo one (Figure 3.7A). However, such an approach was still a long way off from quantitatively accounting for the energetic stabilization of the A9-U63 bp related to the presence of adenine in the binding site. Within this framework there was no way to automatically detect when the nucleobases reached the unbound configuration. It was thus difficult to avoid systematic errors in the comparison of the two systems.

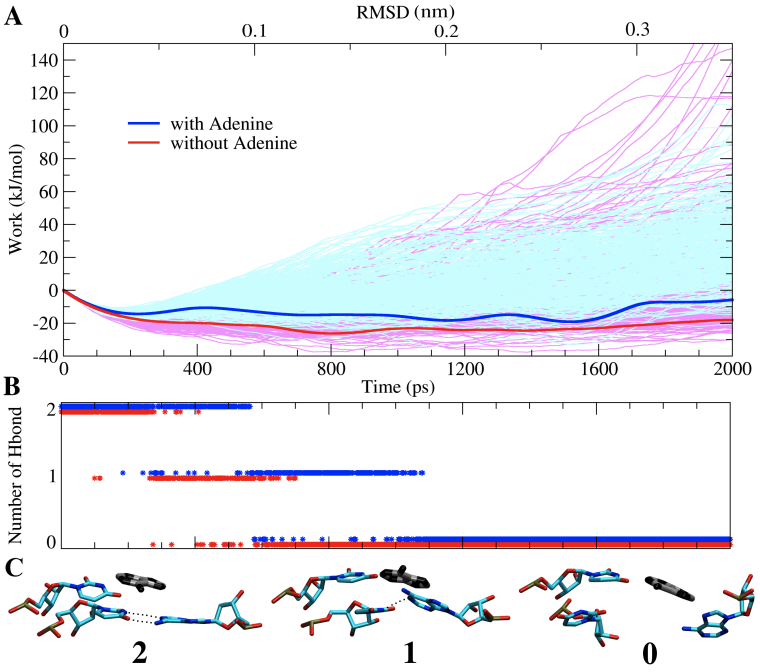

FIGURE 3.7: Unbinding process of the A9-U63 bp. (A) Mechanical work performed as a function of the value of the steered RMSD, or equivalently of time, for 512 simulations for Apo (pink) and Holo form (light-blue). The respective free energies resulting from the Jarzynski equality are shown in thicker red and blue lines. The initial free-energy decrease related to the entropy gain induced by the restraint movement has no consequence on the final result. (B) Hydrogen bonds occurrence for both the systems Apo (red) and Holo (blue). (C) Snapshots of the Holo form (ligand in black) with 2, 1 or 0 hydrogen bond (dotted lines) formed between A9-U63. Adapted from Di Palma et al. [52]

Furthermore, few low-work realizations occurred during the unpairing in the presence of adenine. In these 6 low-work realizations (outliers) the number of hydrogen bonds was non-zero at large RMSD values and the structural analysis of the trajectories revealed the transient formation of a cis Hoogsteen/Sugar non-canonical bp at large RMSD values [140] (highlighted in Figure 3.8). Similar transient conformations were not observed when pulling the apo form. We remind that using the Jarzynski equality is very sensitive to low work realizations. Thus, due to the exponential nature of the Jarzynski average, these low-work realizations dominated the free energy profile for the Holo form further compromising the possibility of a quantitative comparison with the Apo form.

We now focus especially on the free-energy change computed using the Jarzynski equality (shown also in Figure 3.9), for both apo and holo forms. After an initial decrease, which is related to the entropy gain when larger values of the RMSD are allowed, the profiles stabilize to values that are visually different in the apo and holo forms. These data qualitatively confirms that the ligand is stabilizing the A9-U63 bp. For the holo form, we show the results both including and excluding the 6 outliers highlighted in Figure 3.9. The inclusion of the outliers does not change the qualitative pictures. However, it

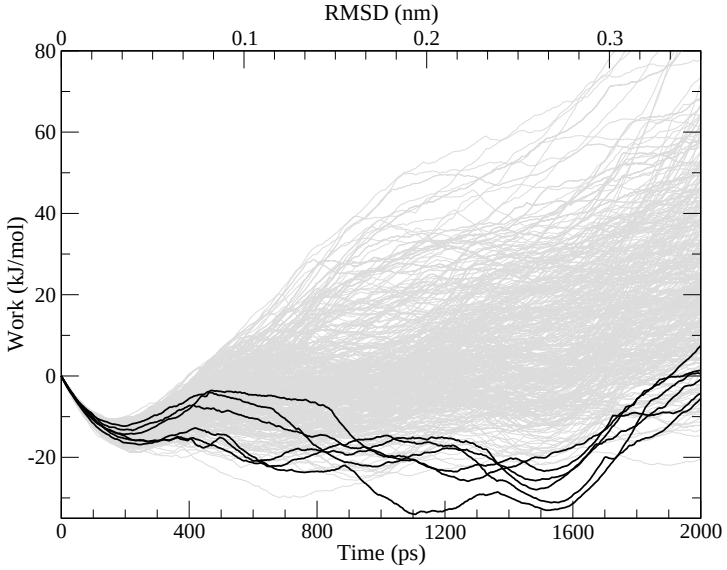

FIGURE 3.8: Mechanical work performed during the rupture of the A9-U63 base pair for the holo form (grey). Work is shown as a function of time (lower scale) and of the position of the restraint (upper scale). 512 simulations are overlaid. Outlier simulations for which a non-canonical pairing is observed at large RMSD are highlighted in black. Adapted from Di Palma et al. [82]

is not trivial to quantify in an objective manner the difference because the two profiles present several ripples.

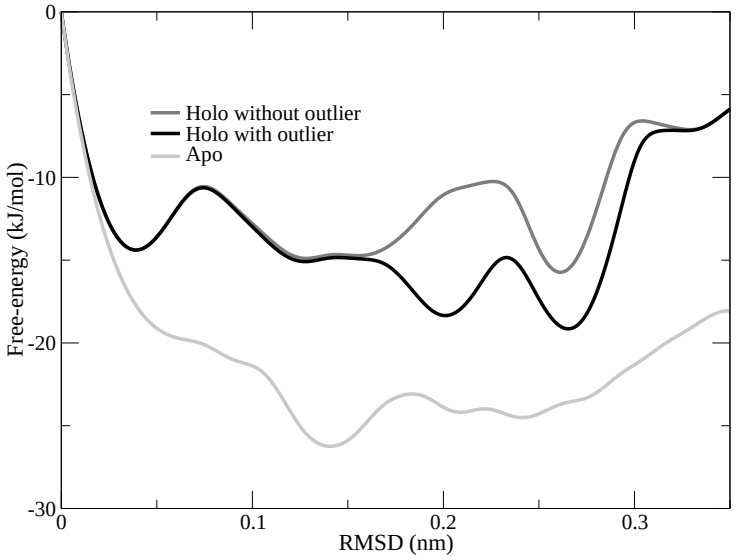

FIGURE 3.9: Free-energy profiles of the A9-U63 bp unpairing process resulting from Jarzynski equality as a function of the value of the RMSD from native. The profiles computed using all the 512 simulations for holo (black) and apo (light grey) form are shown. In dark-grey here we show the profile for the the holo form obtained excluding the 6 outliers. Adapted from Di Palma et al. [82]

3.3.2.2 Energetics of hydrogen bond breaking

We thus analyzed the trajectories in terms of number of hydrogen bonds formed between A9 and U63, a discrete variable that more strictly reported on the breaking of the pairing (Fig. 3.7B-C). In this metrics, the bound (1 or 2 hydrogen bonds) and unbound (0 hydrogen bond) ensembles could be clearly and unambiguously identified thus allowing a quantitative comparison between the Apo and the Holo system. Additionally, the configurations from the outlier trajectories could be assigned properly to one or the other ensemble in spite of their atypical RMSD value. Indeed, the behavior of the outliers can be better interpreted by looking at the hydrogen bonds between A9 and U63 along the pulling (Figure 3.10). Here one can appreciate that in the selected outlier the hydrogen bonds survive longer, so that low energy states are still possible at relatively large values of the RMSD. This is an example of how ambiguous the RMSD can be as a CV along which reconstruct the free-energy profile.

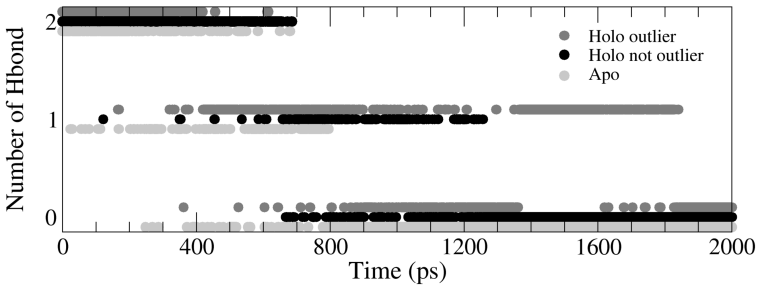

FIGURE 3.10: Hydrogen bonds occurrence between A9 and U63 for typical apo (light grey) and holo simulations (black). The different behavior of one among the outliers of the holo form is shown in dark grey. Adapted from Di Palma et al. [82]

To solve such an ambiguity we directly looked at the rupture of the interactions describing the opening of the bp and then we reconstructed the free-energy profile by reweighting the probability distribution accordingly. We thus used the reweighting scheme (discussed in more details in 2) to compute the free energy as a function of the number of hydrogen bonds between A9 and U63. The Jarzynski relationship allows one to pull a CV and compute the free-energy landscape projected on that CV. Colizzi and Bussi have recently shown that the scheme can be modified so as to allow the calculation of the free-energy landscape projected on another, a posteriori chosen CV [94]. In practice, using this method one can, for instance, enforce an easy-to-pull distance and then estimate the free-energy difference of the process by looking at the rupture or formation of interactions among neighboring residues or at changes in the number of surrounding water molecules so as to quantify solvation.

The differences in free energy (ΔF) between the ensembles, with and without hydrogen bonds, was computed using this just mentioned reweighting scheme. The values and the

associated standard errors were estimated for both systems. In Table **??** we report the free-energy difference between the state where at least one hydrogen bond is formed and the state where there are no hydrogen bonds. For the Apo form $\Delta F = +2.1 \pm 1.2\ kJ/mol$ suggesting that the bp could spontaneously break in the absence of adenine. For the Holo form $\Delta F = -1.3 \pm 1.5\ kJ/mol$ implying that the presence of the ligand and its pairing with U63 stabilized the stacked bp. The $\Delta\Delta F$ between the two forms is equal to $-3.4 \pm 2\ kJ/mol$. This value quantifies the thermodynamic stabilization to the formation of the base pair which directly interacts with adenine in the P1 stem. The values of ΔF reported in Table 3.3 quantify the stability of the A9-U63 bp in presence and absence of the ligand. These absolute numbers should be interpreted with care because of the intrinsic over-estimation of free-energy changes when the Jarzynski equation is applied to a finite sample of trajectories. Additionally, the number of possible conformations with no hydrogen bonds is very large and its entropic contribution could be underestimated here. However, the apo and holo results can be expected to be affected by similar systematic errors so that their difference should reliably estimated ($\Delta\Delta F$). The obtained result is compatible with the stabilization provided by an additional AU bp as measured from thermodynamic experiments [141]. Table 3.3 also reports for the influence of the outliers in the final results. The difference is small and does not affect our conclusion. However, it is larger than the reported error. We remark that low-work realization have a large impact in all analysis procedures based on the Jarzynski equality, and that they should be included. We also notice that by removing these trajectories the apparent error evaluated by the bootstrap procedure is artificially decreased.

TABLE 3.3: ΔF required for the disruption of the A9-U63 base pairs

Apo	Holo	Holo w/o outliers
2.1 ± 1.2	-1.3 ± 1.5	-3.1 ± 0.9

In Figure 3.11 we show how the error estimation on the reported free-energy changes depends on the number of bootstrap iterations. As it can be seen, the error estimate is quite robust and a few bootstrap iterations are enough to converge its calculation. We underline here that this estimate only includes statistical errors that can be detected by looking at the differences between the 512 realizations. As a consequence if an alternative low-work pathway is never sampled in any of the steered MD simulations then the result can be biased. These errors can be expected to affect in the same manner simulations on similar systems made with the same protocol. Thus, we believe that the method can be reliably used to estimate differential stabilities.

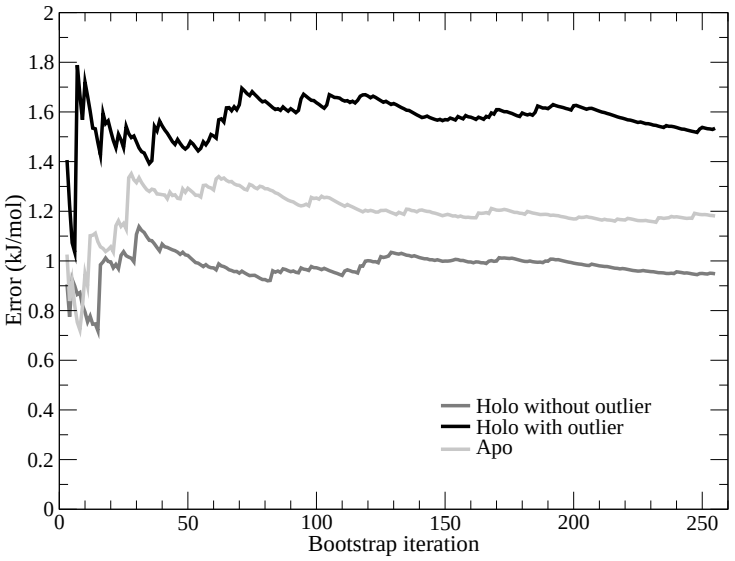

FIGURE 3.11: Error in the free-energy difference between states with no hydrogen bonds and states with at least one hydrogen bond, as a function of the bootstrap iteration. Data for holo (black) and apo (light grey) are shown. The error for the holo form obtained excluding the 6 outliers is shown in dark grey. Adapted from Di Palma et al. [82]

3.4 Discussion

Our simulations at atomistic detail provide for the first time the free-energy contribution of ligand stacking to the formation of the P1 stem in a riboswitch. In particular, the presented *in silico* approach allows the energetics involved in the aptamer stabilization upon ligand binding to be dissected in detail. Below we compare our results with single-molecule manipulation, both *in vitro* and *in silico*, and thermodynamic data from dsRNA melting experiments. We also provide a model for ligand-modulated co-transcriptional folding of the *add* riboswitch.

3.4.1 Comparison with experiments

Our results are in nice agreement with thermodynamic data based on dsRNA melting experiments [141, 142]. The comparison between our simulations and those experiments can be straightforwardly achieved by considering the pairing between U62 and the sensed adenine as an additional terminal bp of the P1 stem. The direct stabilization of the P1 stem due to the cognate-ligand binding should be thus equivalent to that given by adding one further AU bp to the P1 helix. Using the most recent nearest neighbor energy parameters for the comparison of RNA secondary structures [141, 142], the free-energy difference between the sequence of the P1 stem with and without the additional AU

base pair, $\begin{matrix}\text{5'-CGCUUCAUA}\mathit{A}\text{-3'}\\ \text{3'-GUGAAGUAU}\mathit{U}\text{-5'}\end{matrix}$ and $\begin{matrix}\text{5'-CGCUUCAUA-3'}\\ \text{3'-GUGAAGUAU-5'}\end{matrix}$, can be computed [143, 144] as $\Delta\Delta F = -3.7\ kJ/mol$, consistently with our results.

Our free-energy estimates complement previously reported investigations in which the role of the ligand in the folding process of the A-riboswitch has always been referred to the whole aptamer [64, 67] and never specifically to the P1 stem. Using a one-bead-per-nucleotide coarse-grained model, the $\Delta\Delta F$ has been computed as approximately $-15\ kJ/mol$ [67]. Notably also this calculation has been done using a shortened P1 stem, possibly affecting the ΔF estimation. Single-molecule force spectroscopy experiments have been also performed to characterize the folding pathway of the aptamer, reporting a much larger value of $\Delta\Delta F(33.5 \pm 9 kJ/mol)$ for the ligand-induced stabilization of the entire aptamer. In those experiments, the effect of the ligand can only be seen in the very last transition along the folding pathway. The last metastable conformation along the folding pathway is not affected by the ligand. This conformation has been interpreted as one where the stems P2 and P3 are folded and the binding pocket is already organized as in the fully folded conformation. We remark that it is very difficult to access to these molecular details in force spectroscopy experiments, and that often one has to infer the conformation from macroscopic information such as the polymer length. Therefore, the difference between the values reported in our work and the values measured from single molecule experiments, could be ascribed to the fact that the last transition along the folding pathway also requires a substantial pocket reorganization, possibly coupled with the formation of the L2-L3 kissing complex. This reorganization is ligand dependent and could take into account for the additional coupling between ligand binding and P1 stem folding. Furthermore, in both these works the separated contributions of the P1-ligand stacking, of the interaction between the ligand and the junctions J1-2, J2-3 and J3-1, and of the interaction between loops L2-L3 could not be discerned.

3.4.2 Possible methodological improvements

Steered MD is one of the most straightforward enhanced sampling techniques. It is grounded on the solid theoretical framework of the Jarzynski equality, and allows one to induce a conformational transition in a prefixed time. It thus can be used also when limited computational resources are available. Moreover, it is easy to analyze the work distribution to assess the method and possibly optimize the choice of the CV. However, steered MD is limited to the pulling of a single CV at a time, and requires one to guess the range of relevant values for the CV. An alternative is umbrella sampling [89], often used in a multiple-restraint approach in combination with weighted-histogram analysis method (Kumar et al, 1992). Umbrella sampling can be used with 1 or 2 CVs, but care

must be taken since sometime it is not straightforward to recognize when CVs are not sufficient to describe the reaction (see e.g. Di Palma, Bottaro, and Bussi, submitted). For more complicated conformational changes, where one CV is not enough to describe the reaction, metadynamics [90] could be a better alternative. When used in its well-tempered formulation [112], metadynamics do not require the boundary of the relevant region in the CV space to be prefixed. If a large number of CV is required one can use the bias exchange formalism [114]. Alternatively, when finding effective CV is too difficult, metadynamics can be combined with complementary techniques such as parallel tempering [145].

Chapter 4

Kissing loop interaction dynamics and thermodynamics

4.1 Introduction

The open framework on the investigations about the role of the ligand in the structural organization of the aptamer [57, 63, 64] lacks both the atomistic details and the distinct energetic contributions associated to ligand binding. It has been suggested experimentally that the ligand stabilizes the interaction of the distal kissing complex [62]. At the same time a stable kissing interaction seems to contribute to the ligand binding energy stabilizing the complex in a cooperative fashion [39, 57, 63]. Also *in silico* techniques have been used obtaining a more accurate description of the system from a structural point of view [68–70, 72]. In a few cases a computational approach has been employed to provide a thermodynamic characterization of the system [52, 67, 71]. In particular, Allnér *et al.* [71] computed the free-energy profile corresponding to the formation of the kissing complex using molecular dynamics (MD) simulations in explicit solvent, both in the presence and in the absence of the ligand, using the CHARMM36 force field [146, 147]. MD does not require experimental inputs and can in principle be used in a predictive fashion. However, accuracy of atomistic force fields is still debated and it is thus very important to compare results obtained employing different sets of parameters. In this part of the project we use atomistic MD with the latest variant of the Amber force field [127] in combination with enhanced sampling techniques [87] to provide a more detailed perspective on the formation of the loop-loop interaction. It is a tertiary contact named kissing stem-loop also called loop-loop pseudoknots that occurs when the unpaired nucleotides in one hairpin loop, base pair with the unpaired nucleotides in another hairpin loop. In this particular case we have two guanine (G25 and G26) of

L2 that interact through canonical Watson-Crick base pairs with two cytosine (C48 and C49) of L3. Furthermore there are other interloop stackings and non-canonical interactions that stabilize this complex, something present also in other loop-loop interactions (i.e. having a different sequence) (REF Garcia). The combined approach allows this contribution to be dissected from the other ligand-aptamer interactions and the impact of the ligand on the stability of the loop-loop interaction to be quantified. At first, we reproduced exactly the same protocol that has been used by Allnér *et al.* [71] in order to perform a fair comparison between the two force-fields on this particular system. More then this reference work, we extended the US simulations to obtain converging results. The effects of the initialization protocol on the US windows, depending on the parameters used and the directionality of the pulling simulations performed to generate the starting structure of each of them, are also discussed in detail. Secondly, a more complex approach was employed taking advantage of the combination of metadynamics and Hamiltonian replica-exchange that allow us to manage two important degrees of freedom of the system to study more deeply and accurately the thermodynamics of the kissing loop interaction using 16 replicas. The use of these two advanced sampling techniques produced simulation for more than 6 microseconds on four systems. Indeed, we tried not only to compute the free energy difference between the Apo and the Holo form, but also to detect and to understand the different influence of monovalent and divalent cations on the loop-loop tertiary interaction. This additional effect, like the tight network of hydrogen bonds between the loops nucleotides, make the choice of the suitable CVs to bias more difficult, thus the combination of the two methods is quite challenging due to the complexity of the system under investigation.

4.2 Systems setup and protocols

In this chapter will be illustrated how umbrella sampling (US) simulations [89] can be applied to the study of the thermodynamics of the kissing loop in the presence and in the absence of the cognate ligand. Furthermore, we also show how tackling this problem with a more complex combination of simulation techniques, metadynamics and Hamiltonian replica-exchange, improved the obtained results both from the structural and thermodynamics point of view. These simulations produced a lot of data and the analysis of these long trajectories allowed also a preliminary description of the influence of K^+ and Mg^{2+} on the adenine aptamer domain. To set up the US protocol with multiple restraints [95] we enforced the breaking/formation of the loop-loop interaction steering the distance between the two loops and using the resulting structures as starting conformations for each simulated window. All the simulation parameters are discussed

in detail in the following subsections. As well as, in the proper section the 4 systems simulated with metadynamics and Hamiltonian replica-exchange are also described.

4.2.1 Systems set up

The US simulations reported hereafter were performed on two systems: the *add* aptamer domain complexed with adenine (Holo form) and without adenine bound (Apo form). In both cases we used the X-ray structure solved by Serganov *et al.* [PDB:1Y26] [51]. The ligand was removed to simulate the unbound state. MD simulations were performed using the Amber99 force field [125] refined with the parmbsc0 α/γ corrections [126] and the latest χ torsional parameters [127]. The general Amber force field [128] was used to parametrize the ligand. Partial atomic charges were assigned using the restricted electrostatic potential fit method [129] based on an electronic structure calculation at the HF/6-31G* level of theory performed with Gaussian03 [130]. The electrostatic interactions were calculated using the particle-mesh Ewald method [132] and bond-lengths were constrained with LINCS [131]. The systems were set-up following exactly the protocol described in Allnér *et al.* [71]: aptamers were solvated in a rhombic dodecahedron having 8 nm as box vector lenghts, with a Mg^{2+}-H_2O solution using approximately 11000 TIP3 molecules [133], and a recent parametrization for divalent cations [81]. The 5 crystallographic Mg^{2+} were initially kept at their respective position, whereas the additional 30 ions added to neutralize the system ($[Mg^{2+}]$ = 0.18 M) were randomly placed. A steepest descent minimization (150 steps) was performed followed by 200 ps of MD at constant temperature (298 K, using stochastic velocity rescaling [134]) and pressure (1 atm, using the Berendsen barostat [135]) with positional restraints on both RNA and ions so as to equilibrate water. This procedure was repeated first removing the constraints on the ions and then removing all the remaining constraints. Finally, 12 ns unrestrained simulations at constant volume were performed for each system. More than these two systems described above, the Hamiltonian replica-exchange simulations with metadynamics were also performed on the Apo and the Holo form in a K^+-H_2O solution containing 70 monovalent cations. These other systems were set up with the same protocol as before, positioning 5 K^+ in place of the crystallographic Mg^{2+}.

4.2.2 Umbrella sampling protocols

In order to compute the thermodynamic stability of the loop-loop interaction we employed US simulations with multiple harmonic restraints. The distance between the center of mass (CoM) of the backbone atoms of the two loops (L2: bases 20-26; L3:

bases 48-54; Figure 4.1B) was used as a collective variable (CV). We will refer hereafter to this distance as L. 44 uniformly spaced reference values were taken in the range spanning from 12.5 to 34 Å, and restraints with stiffness $k = 20$ $(kcal/mol)/\text{Å}^2$ were employed. In the production phase of the US simulations each of the 44 windows was run for 5 ns.

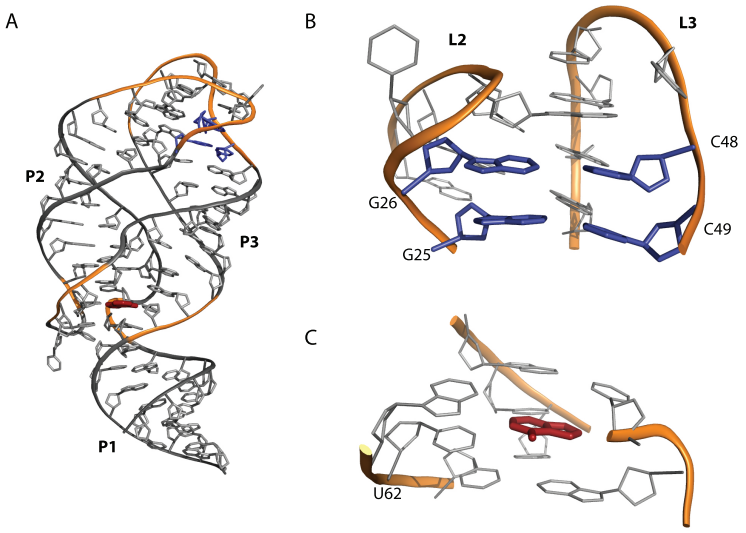

FIGURE 4.1: A) Three dimensional representation of the aptamer with the adenine bound. The stems are shown in grey and labeled. The backbone of the loops and of the junctions is shown in orange. B) Close-up on the loop-loop (L2 and L3) interaction with focus on the Watson-Crick base pairs (G25-C49, G26-C48, in blue). C) Close-up on the ligand binding site with the adenine (red) paired with U62. Adapted from Di Palma et al. [49]

A very important issue in US simulation is the generation of the starting conformations. We here performed two independent US sampling simulations, using starting conformations generated with two different protocols (hereafter referred to as *forward* and *backward*). To generate the starting points for the forward US simulations we employed the same protocol used by Allnér *et al.* [71]. Namely, we ran a series of 44 short (0.25 ns) simulations with a stiffer restraint ($k = 40$ $(kcal/mol)/\text{Å}^2$) keeping the CV at the 44 reference values, where each simulation was initialized from the last frame of the previous one. In this way, before each US window starting structure was sampled, we let the system equilibrate. The reference values were iterated allowing an increasing distance between the loops. The backward US simulation was initialized with an equivalent procedure but iterating the restraints in the opposite order, i.e. starting from the structure with undocked loops. In principle, if US simulations are converged, the result should be independent of the initialization procedure. In order to assess the backward procedure for the US method, we repeated it for the Apo form using a softer restraint ($k = 20$ $(kcal/mol)/\text{Å}^2$). We performed two additional simulations:

A Starting from the final snapshot of the forward procedure explained above (with the stiffer restraint) we perform a backward procedure with the softer restraint, then using the snapshots obtained with this alternative protocol to perform another US simulation.

B We repeated both the forward and the backward procedures using the softer restraint

Here after there are the sample input files for the latest PLUMED version (2.1): First, both the forward and backward initialization procedure to sample the starting conformation of each of the 44 US windows; Second, an example of US simulation.

```
# make the molecule whole across periodic boundary condition
WHOLEMOLECULES ENTITY0=1-2257
# compute center of mass of loop 2
com1: COM ATOMS=596-607,623-641,656-674,686-704,719-737,749-767,783-811,817-823
# compute center of mass of loop 3
com2: COM ATOMS=1496-1507,1520-1538,1551-1569,1581-1599,1611-1629,1644-1662,1677-1695,1710-1716
# compute the distance between the CoM of the two loops
ll: DISTANCE ATOMS=com1,com2
# apply a step by step moving restraint letting the system relax at each step before the starting configuration for the US window is sampled
MOVINGRESTRAINT ...
ARG=ll
STEP0=0 AT0=1.25 KAPPA0=16750
STEP1=124999 AT1=1.25
STEP2=125000 AT2=1.3
STEP3=249999 AT3=1.3
STEP4=250000 AT4=1.35
STEP5=374999 AT5=1.35
STEP6=375000 AT6=1.4
STEP7=499999 AT7=1.4
STEP8=500000 AT8=1.45
STEP9=624999 AT9=1.45
# always the same for the restraint from 1.5 to 3.15 nm [...]
STEP78=4875000 AT78=3.2
STEP79=4999999 AT79=3.2
STEP80=5000000 AT80=3.25
STEP81=5124999 AT81=3.25
```

```
STEP82=5125000 AT82=3.3
STEP83=5249999 AT83=3.3
STEP84=5250000 AT84=3.35
STEP85=5374999 AT85=3.35
STEP86=5375000 AT86=3.4
STEP87=5499999 AT87=3.4
LABEL=m
... MOVINGRESTRAINT
# print the results for subsequent analysis
# namely: distance, center of the restraint, performed work
PRINT ARG=ll, m.d_cntr,m.d_work FILE=COLVAR
```

setting KAPPA0=8375 instead of KAPPA0=16750, a softer restraint is applied to the CV, as was described above for the Apo form additional simulations.

The backward US simulation was initialized with an equivalent procedure but iterating the restraints in the opposite order

```
MOVINGRESTRAINT ...
ARG=ll
STEP0=0 AT0=3.4 KAPPA0=16750
STEP1=124999 AT1=3.4
STEP2=125000 AT2=3.35
STEP3=249999 AT3=3.35
STEP4=250000 AT4=3.3
STEP5=374999 AT5=3.3
STEP6=375000 AT6=3.25
STEP7=499999 AT7=3.25
STEP8=500000 AT8=3.20
STEP9=624999 AT9=3.20
# always the same for the restraint from 3.15 to 1.5 nm [...]
STEP78=4875000 AT78=1.45
STEP79=4999999 AT79=1.45
STEP80=5000000 AT80=1.4
STEP81=5124999 AT81=1.4
STEP82=5125000 AT82=1.35
STEP83=5249999 AT83=1.35
STEP84=5250000 AT84=1.30
STEP85=5374999 AT85=1.30
```

STEP86=5375000 AT86=1.25
STEP87=5499999 AT87=1.25
LABEL=m
... MOVINGRESTRAINT

To perform an US simulation this is the Plumed input file:

```
# make the molecule whole across periodic boundary condition
WHOLEMOLECULES ENTITY0=1-2257
# compute center of mass of loop 2
com1: COM ATOMS=596-607,623-641,656-674,686-704,719-737,749-767,783-811,817-823
# compute center of mass of loop 3
com2: COM ATOMS=1496-1507,1520-1538,1551-1569,1581-1599,1611-1629,1644-1662,1677-1695,1710-1716
# compute the distance between the CoM of the two loops
ll: DISTANCE ATOMS=com1,com2
# count the number of native inter-loop interactions
c: COORDINATION GROUPA=774,775,778,808,809,812,711,715,648,682,685,815,812
GROUPB=1545,1548,1550,1514,1517,1519,1567,1544,1699,1666,1669,1702,1705 R_0=0.35
PAIR
# apply the harmonic restraint centered at a specific CV position
res: RESTRAINT ARG=ll AT=2.85 KAPPA=8375
# print the results for the analysis
# namely: loop-loop distance, applied bias, number of native contacts
PRINT ARG=ll,res.bias,c FILE=COLVAR
```

4.2.3 Analysis methods

The data were analyzed using the last 4 ns of each window. The potential of mean force (PMF) profiles were constructed using the weighted histogram analysis method (WHAM) [95] implemented by Grossfield [148] taking the CV values distribution resulting from the US simulations. This implementation of WHAM allows to compute errors with a bootstrapping procedure that assumes uncorrelated samples. To avoid artifacts due to possible correlations we instead adopted a blocking procedure. Namely, we split the final 4 ns of each trajectory in four blocks of 1 ns each and performed the WHAM calculation using only a single block from each simulations. The four resulting profiles are aligned at their CV starting value (12.1 Å for the forward profiles, 34.4 Å for the backward profiles) and error at each point is computed as the standard deviation among the four profiles divided by $\sqrt{4}$.

To define the number of stacking interactions and the number of base-pair contacts a new local coordinate system for purines and pyrimidines was defined (as described in 2.7). It is a minimalist representation for RNA molecules consisting in one oriented bead per nucleotide. The vector $\mathbf{r}_{jk}$ describes the position of each base ring k in the coordinate system constructed on base ring j and is expressed in cylindrical coordinates (Figure 2.1). Thus, this means that the relative position and orientation between two nucleobases. is described by this vector. There are several criteria for determining the canonical WC bp, the non-canonical bp and the stacking interaction (see Chapter 2.7); this is necessary because it is well known that the strength and nature of pairing and stacking interactions depend on the base-base distance, on the angle θ as well as on other angular parameters (e.g., twist, roll, tilt) in a non-trivial manner [138, 149]. This procedure was shown to be useful for characterizing both structural and dynamical properties of RNA molecules [84]. The software used to perform this structural analysis is available online (http://github.com/srnas/barnaba).

4.2.4 Metadynamics and HREx protocol

To obtain a more accurate thermodynamic characterization and a wide conformational sampling of the aptamer apical tertiary interaction, such as to study the effects on this RNA domain given by different ions both from the energetic point of view and the structural one, we chose to bias our simulations through WTmetaD running more than one replica at the same time with Hamiltonian replica exchange. This choice was done for several reasons, mainly to efficiently accelerate the sampling and to faster override the barriers of the free energy landscape obtaining better results in a shorter time with a less computational time-consuming approach thanks to the high parallelization of the used programs. The biased CVs for the MetaD algorithm are the distance between the center of mass (CoM) of the backbone atoms of the two loops (L2: bases 20- 26; L3: bases 48-54) and the number of native interloop contacts between the loops bases (the atoms involved in each pairing are listed in the sample input file ahead), taking the PDB structure as reference, thus including the six hydrogen bonds constituting the two peculiar WC base pairs. A different width of the Gaussians was used for the two CV due to their different fluctuations computed during the long equilibration runs (described in the System set up subsection). The $\mathfrak{H}$ region in which the Hamiltonians are reparametrized includes all the atoms of the loops, whereas the $\mathfrak{C}$ region comprehends all the other nucleotides of the aptamer (1-19, 27-47 and 55-71, plus the ligand in the Holo form). We set λ according to the geometric distribution between 1 and 0.5 to scale the force field parameters of the 16 replicas we have simulated, running each of them for 100 ns. Thus, this means that using the Hamiltonians reparametrized accordingly to these λ values we simulated the

region $\mathfrak{H}$ of the systems at an effective temperature between 300 and 600 K, even if in practice the temperature is set at 300 K for all the simulations.

For each replica a plumed input file as to be provided, the following line show a practical example:

```
# make the molecule whole across periodic boundary condition
WHOLEMOLECULES ENTITY0=1-2257
# compute center of mass of loop 2
com1: COM ATOMS=596-607,623-641,656-674,686-704,719-737,749-767,783-811,817-823
# compute center of mass of loop 3
com2: COM ATOMS=1496-1507,1520-1538,1551-1569,1581-1599,1611-1629,1644-1662,1677-1695,1710-1716
# compute the distance between the CoM of the two loops
ll: DISTANCE ATOMS=com1,com2
# count the number of native inter-loop interactions
c: COORDINATION GROUPA=774,775,778,808,809,812,711,715,648,682,685,815,812
GROUPB=1545,1548,1550,1514,1517,1519,1567,1544,1699,1666,1669,1702,1705 R_0=0.35
PAIR
# used to perform metadynamics on two CVs: the loop-loop distance and the native inter-loop contacts. Two different widths of the Gaussian hills were used for the two CVs.
m: METAD ARG=ll,c SIGMA=0.0128,0.167 HEIGHT=0.8 PACE=500 BIASFACTOR=19
TEMP=300
# printed data for the analysis: loop-loop distance, number of native contacts, metadynamics bias
PRINT ARG=ll,c,m.bias FILE=COLVAR
```

Each of the 16 runs give as output also the HILLS file, which is used both to restart the calculation and to reconstruct the free energy as a function of the CVs. It contains the Gaussian height re-scaled using the bias factor step by step.

4.3 Results

4.3.1 Evidences from Umbrella Sampling

4.3.1.1 Forward process

The analysis of the Holo forward and Apo forward US trajectories allowed the PMF for the disruption of the kissing complex to be computed. The resulting profiles are plotted

in Figure 4.2 for both Holo and Apo systems. The PMF shows a minimum at $L \approx$ 12.5 Å, corresponding to the initial structure. The free energy change upon disruption of the kissing complex for the Holo structure is ΔG = -52 ± 2 kcal/mol. For the Apo structure the stability of the complex is largely reduced to ΔG = -35 ± 3 kcal/mol. The stabilization of the kissing complex provided by the ligand can thus be estimated as $\Delta\Delta G$ = -17 ± 3 kcal/mol.

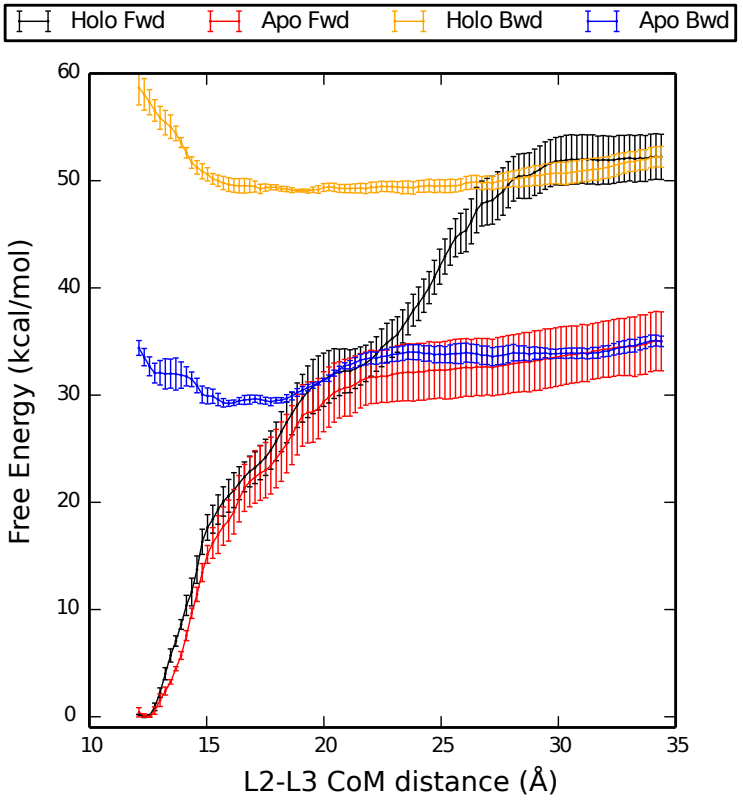

FIGURE 4.2: Potential of mean force as a function of the distance between the centers of mass of the L2 and L3 loops. Results for Holo and Apo forms are shown as obtained from two independent umbrella sampling simulations using different protocols to obtain the initial structures (forward, Fwd, and backward, Bwd, see main text for definition). Fwd and Bwd profiles are aligned at the maximum distance. Adapted from Di Palma et al. [49]

To understand which are the interactions that are relevant for the kissing complex formation we analyze inter-loop pairings and inter- and intra-loop stacking interactions for each of the restrained simulations (Figure 4.3). For both the Holo and the Apo forms, at a $L \gtrapprox 16$ Å, only the two inter-loop WC base pairs (G25-C49, G26-C48) peculiar of the loop-loop interaction are still formed. On the contrary, all the non-canonical base pairs are disrupted. In the Apo structure the inter-loop WC pairings were irreversibly lost at $L > 23$ Å, whereas in the Holo structure they are at least partially maintained until $L \approx 30$ Å. We also analyzed the rupture of stacking interactions, distinguishing intra-loop and inter-loop contacts. Inter-loop stacking behaves in a manner qualitatively

similar to the inter-loop WC pairings, going to zero at a distance $L \approx 23$ Å (Apo) and 30 Å (Holo). On the contrary, the intra-loop stacking interactions are still present when the kissing loop is disrupted, indicating that the internal structure of the two loops is preserved during undocking. It can be observed that in the Holo simulation the number of intra-strand stacking slightly decreases (≈ 5) for 29 Å $\lesssim L \lesssim$ 30 Å because of the distortion in the structure induced by the one of the two inter-loop WC pairings. After this residual interaction is lost all the intra-loop stacking contacts are recovered.

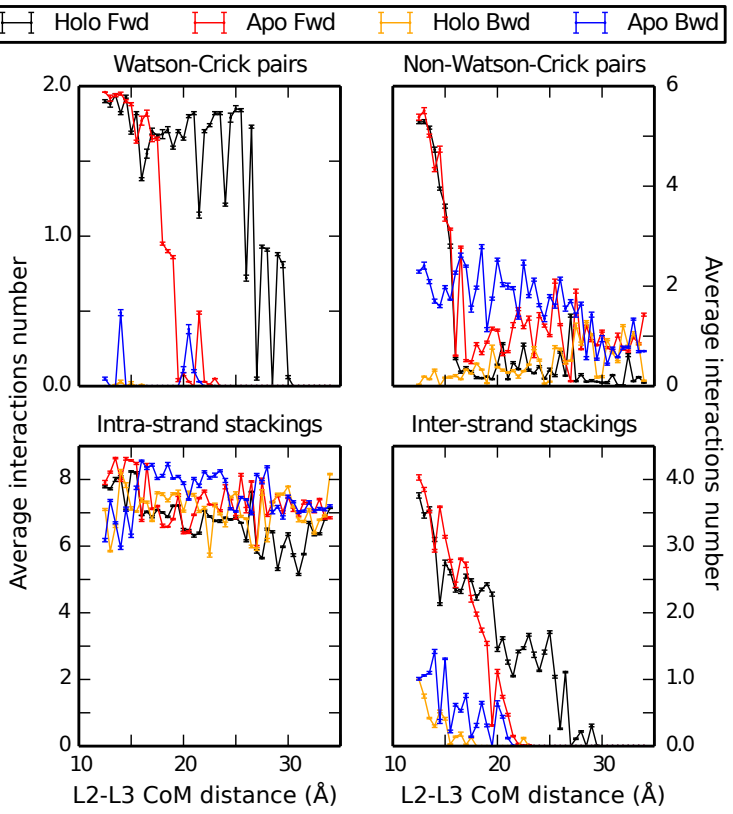

FIGURE 4.3: Average count of inter and intra-loop interactions from umbrella sampling simulations. Results are shown for both Apo and Holo forms, using both forward (Fwd) and backward (Bwd) protocols (see main text for definition). Watson-Crick and non-Watson-Crick pairings as well as intra and inter-loop stackings are shown as indicated. Adapted from Di Palma et al. [49]

4.3.1.2 Backward process

In order to better assess the convergence of the free energy landscape for kissing complex formation, we also reconstructed the PMF profiles of the Holo and Apo structure from US simulations initialized with the backward process (Figure 4.2). The forward and backward profiles were aligned at $L = 34$ Å, since at that distance the starting structure of the backward process is equal to the final structure of the forward one. The free-energy change upon docking is estimated taking the difference between the minimum

value of the PMF (Holo: $L \approx 19$ Å; Apo: $L \approx 16$ Å) and its value for the undocked structure (L = 34 Å): for the Holo $\Delta G = +3.2 \pm 0.9$ kcal/mol, for the Apo $\Delta G = +5.9 \pm 0.6$ kcal/mol. Albeit positive, these numbers are too small and not compatible with the ones found in the forward process. This is a clear signature of hysteresis in the pulling procedure that strongly biases the initial starting points of the US simulation. The reason for this discrepancy can be better understood by performing a structural analysis of the interactions on the different US windows. As it can be seen in Figure 4.3, in the backward process the native WC base pairs are not reformed. In general, a few contacts are formed between the two loops but they are not enough to stabilize the kissing complex. To be sure that this is a systematic effect we also tried a few alternative settings for backward simulations. Their structural analysis is shown in Figure 4.4, where it can be appreciated that only the simulations with protocol A (in light blue) were able to correctly re-form the native WC pairs. Although the restraint stiffness could affect the result, we believe that here the differences are mostly due to the stochastic nature of MD.

The PMF profile obtained starting from the structure sampled with the initialization protocol A are shown in Figure 4.5. Also in this Apo case the free-energy landscape is incompatible with the one obtained from the forward protocol (compare with Figure 4.2), indicating convergence issues in the US calculation. This problem is due to the fact that without properly reforming the kissing interaction with all the native contacts during the initialization procedure, there is no possibility to sample the low-energy conformations during the US simulations.

4.3.2 Analysis of the HREx simulations combined with MetaD

In the study of the A-Riboswitch the combination of MetaD and HREx can be applied to the characterization of the tertiary interaction between the two apical loops of the aptamer. Thanks to the four simulated systems (i.e. Holo/Mg^{2+}, Apo/Mg^{2+}, Holo/K^{+}, Apo/K^{+}) not only the comparison between the apo and the holo form in the presence of Mg^{2+} is possible, but we can also compare the effects of K^{+} on the aptamer in the presence and in the absence of the ligand. Using the *driver* tool contained into the PLUMED package it is possible to post-process an existing trajectory, so as to calculate the probability density as a function of the biased CVs, but also as a function of *a posteriori* choosen variables, using the kernel density estimation. Analyzing by mean of this utility the second half of our simulations, in particular the replica 0 having not rescaled parameters in the Hamiltonian, we have computed the free-energy profiles, directly correlated with the probability density, of the two biased CVs (i.e. L2-L3 CoM distance, number of native inter loop contacts) and of the aptamer radius of gyration,

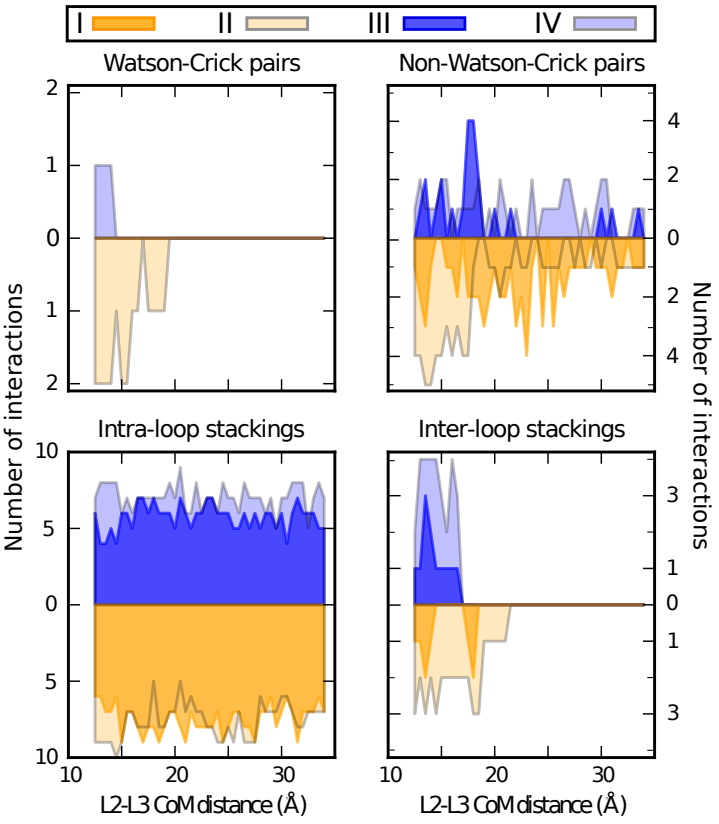

FIGURE 4.4: Count of inter and intra-loop interactions in the 44 starting snapshots resulted from the different Apo US simulation initialization procedures. Results are shown for the 4 Apo runs, using the backward and forward protocol both with $k = 40(kcal/mol)/Å^2$ (I, II, respectively in orange and pale orange), and the two alternative backward protocols with the softer restraint (III in blue from protocol A, and IV in light blue from protocol B). Watson-Crick and non-Watson-Crick pairings as well as intra and inter-loop stackings are shown in the different panels. Adapted from Di Palma et al. [49]

that should increase when the loop-loop interaction is disrupted. The resulting free-energy profiles projected on the single CVs for the four systems are shown in the Figure 4.6, 4.7 and 4.8. In the first two cases we aligned the curves at the point where the loops are open, respectively at a distance of 30 Å between the CoM of L2 and L3, that are completely solvated, and when there are no inter loop contacts, respectively. In the third case we show the four profiles aligned at their minimum because the range of values that were sampled during our simulations were different.

Considering the Figure 4.6, all the profiles, except the Apo/K^+, have their minimum at a value of the distance between the CoM of the loops very close to the reference value of the crystal structure of the aptamer. In the presence of potassium ions and absence of the ligand, the profile show a not so clear minimum at a slightly higher value (≈15 Å) as a result of the less compact structure in this conditions that influence also the distance between the loops. The Apo and the Holo profile in the presence of the divalent

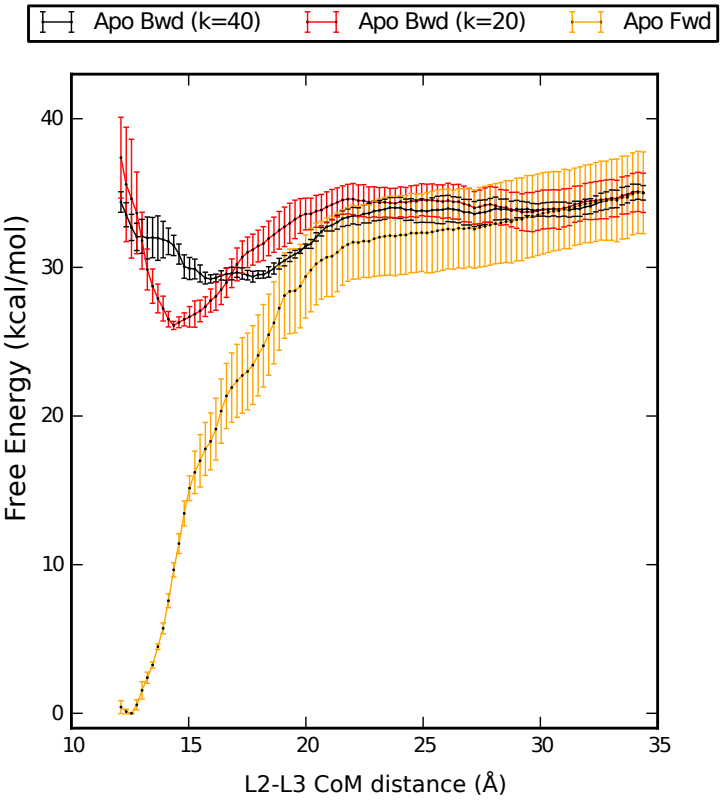

FIGURE 4.5: Potential of mean force obtained with the alternative backward protocol. Results for Apo form are shown in red (Bwd (k = 20)) as obtained from the control umbrella sampling simulations discussed in the Appendix, and aligned with the other two Apo profiles described in the text (Fwd in orange, Bwd (k = 40) in black) for comparison. Adapted from Di Palma et al. [49]

cations have a common behavior with a small shoulder, respectively at ≈17 Å and ≈15 Å, corresponding to the conformation with only the 2 characteristic WC bp still interacting. The Holo/K^+ profile, instead, at ≈ 16 Å, has another minimum comparable in stability with the native structure. Also in this case the two loops are still in contact and not fully solvated due to the presence of the two GC bp. This means that in the presence of monovalent cations and the bound ligand the kissing loop interaction is able to find an alternative and equally stable conformation even if the distance between the CoM of the two loops is larger than the native one. Increasing the LL distance also the free energy of the four profiles increases until a plateau is reached meaning that the tertiary interaction is disrupted. In the Holo/Mg^{2+} system this event happens at a higher distance between the loops CoM (≈ 23 Å) than the other three systems (≈ 19 Å).

Concerning the analysis of the free-energy profiles projected on the second biased CV (Figure 4.7), the number of native inter-loop hydrogen bonds, all the four systems show a more stable structure with 10 interactions than with 12 (number of H-bond in the reference structure). Two the possibile meaning for these feature: First, the different ionic strength conditions between our simulation and the crystallographic experiments,

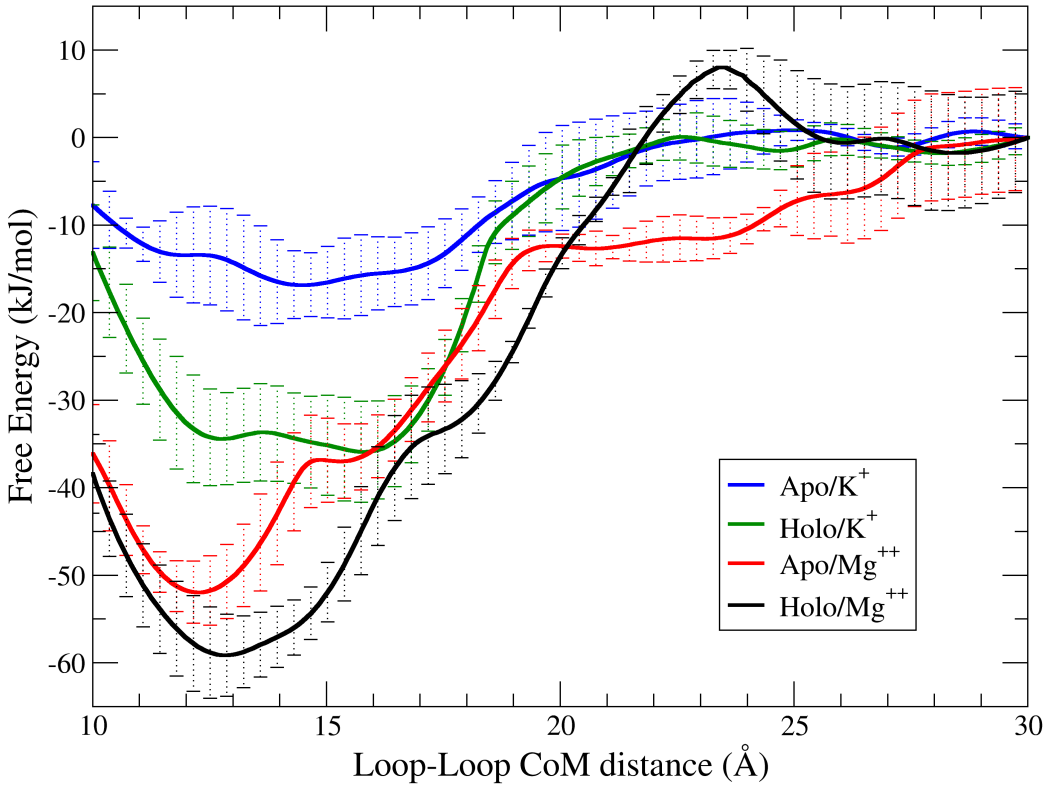

FIGURE 4.6: Free energy projected on the loops center of mass distance. The four profiles are aligned at a loop-loop distance equal 3.0 Åwhere the kissing interaction is broken.

as well as to the constrained and packed nature of the crystal. Second, how the CV is computed in Plumed; it is not discrete but continuos between 0 and 1 through a switching function, $s_{ij} = \frac{1-\left(\frac{r_{ij}-d_0}{r_0}\right)^n}{\left(\frac{r_{ij}-d_0}{r_0}\right)^m}$ where $d_0 = 0$ and $r_0 = 3.5$, to make it differentiable. Generally speaking, in the four profiles we can observe several small barriers describing the disruption/formation of the native H-bonds during the loops docking/undocking process. The two profiles in the presence of K^+ have their deeper minimum when all the H-bonds are broken revealing the unstable nature of the kissing loop interaction in the absence of the divalent cations. Indeed, as demonstrated with UV melting experiments, the docking of the hairpin loops is an intrinsically unfavorable reaction that is strongly favored by magnesium ions [62].

As mentioned in the proper sections of the Chapter 2, the use of WTMetaD and HREx, allows the equilibrium probability distribution of other degrees of freedom than the few explicitly included in the CVs to be recovered. In this framework, an unbiased degrees of freedom that can give hints on the state of the kissing loops interaction is the gyration radius of the whole aptamer. We can consider it as a "reporter" variable because a change in its value it can be only related to a higher distance between the two loops,

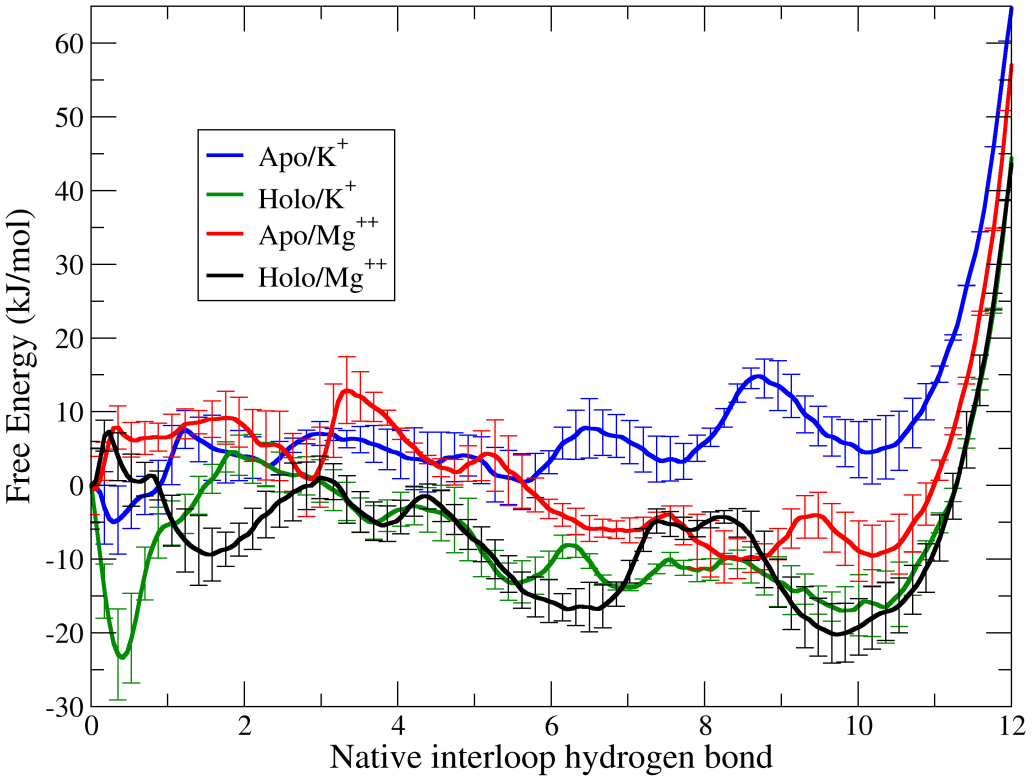

FIGURE 4.7: Free energy projected on the number of native inter-loop hydrogen bonds. The four profiles are aligned at 0 interactions where the loop-loop interaction is disrupted.

and consequently a wider amplitude between the relative stems (P2 and P3) connected by the J2-3. Indeed, the rest of the aptamer, including the P1 stem and the binding site, did not experience any conformational changes that can justify a radius of gyration increase. It is clear in Figure 4.8 that in the presence of divalent cations the apo and holo structure are more compact than in the presence of the monovalent ones, with their minimum free-energy respectively at ≈20 Å and ≈22 Å. Furthermore it is also evident that in the $H_2O - K^+$ solution, the aptamer is allowed to sample a wider range of values so that both the profiles have a second minimum at higher energy representing the structures with the two loops completely solvated at $R_g > 24$ Å. The ΔG between these two minima for the apo is 10±3 kJ/mol and or the holo system is 8±4 kJ/mol. Something different happened in the systems neutralized with magnesium, even if from the structural point of view the two loops lost every interaction the simulations did not sample too much large values of the gyration radius. In theses cases, the two loops did not get as far as in the other two simulations and consequently the linked stems were not open too. This is probably due to the increased screening effect on the negative charges given by the Mg^{2+} than the K^+, thus maintaining the stems closer my means of a reduction of the repulsion between the phosphate groups. Something more can be

said about the profiles obtained from the simulations in the presence of potassium, the apo form seems to reach higher values of R_g. We can speculate this behavior could be due to the absence of the ligand in the binding pocket causing a more flexible J2-3 and less stable P2 and P3 [63] thus inducing the sampling of more extended conformations of this part of the aptamer that our "reporter" variable is able to detect.

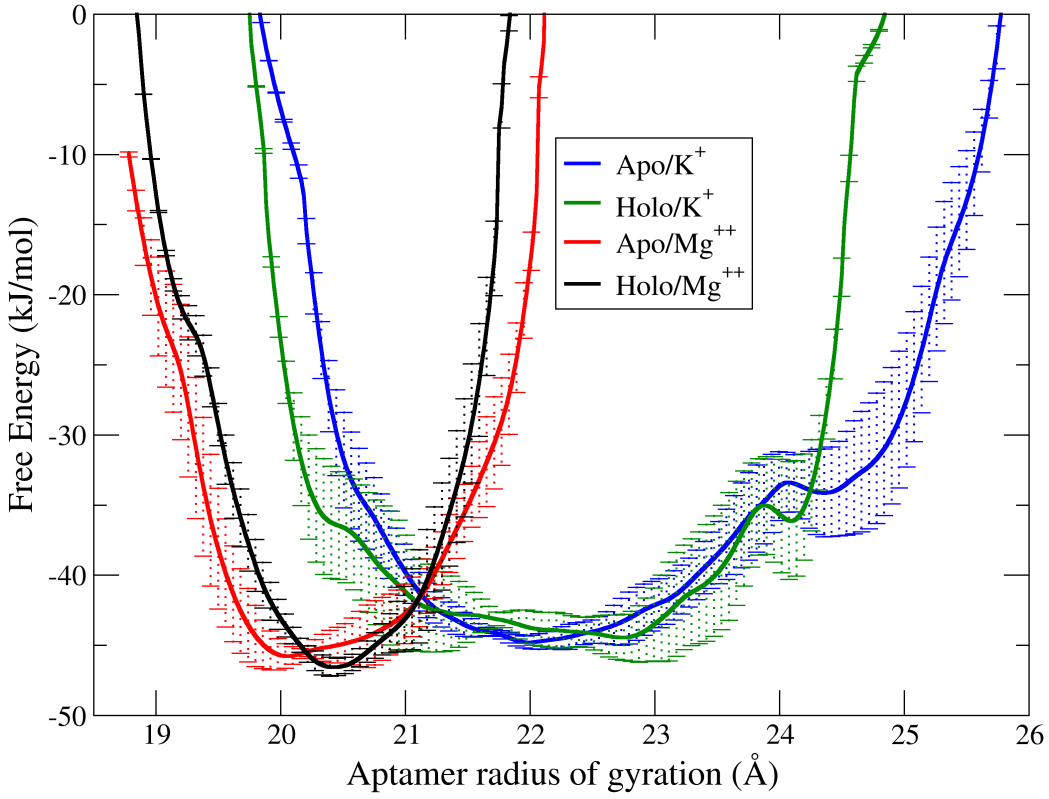

FIGURE 4.8: Free energy projected on the aptamer radius of gyration. The four profiles are aligned at their minimum.

4.4 Discussion

Our US calculations provide quantitative and atomistic details on the mechanism of kissing loop breaking and formation in the *add* riboswitch aptamer domain. The results can be straightforwardly compared with those recently obtained by Allnér *et al.* [71] on the same system using the CHARMM36 force field [146, 147]. In particular the free-energy computed with the forward process has been obtained with an identical protocol so as to allow a fair comparison between the force fields. In our work the estimated stability of the kissing loop complex is ΔG = -52 ± 2 kcal/mol (Holo) and ΔG = -35 ± 3 kcal/mol (Apo), so that upon ligand binding $\Delta\Delta G$=-17 kcal/mol < 0. On the contrary, Allnér *et al.* reported $\Delta\Delta G = +10$ kcal/mol > 0. The sign of $\Delta\Delta G$ indicates whether the ligand binding and the formation of the kissing loop complex are cooperative

(positive) or anticooperative (negative). Results obtained with the two force fields thus interestingly lead to two opposite pictures.

Since 2007, Rieder and collaborators have shown using a structured-based fluorescent spectroscopy approach that the organization of the binding pocket and the tight formation of the L2-L3 hydrogen bonding network occur in a concerted manner, reporting a "long range" induced-fit of the kissing loop interaction by the ligand binding event [57]. Recent experiments probed the differential ligand affinity in aptamers with mutations hindering the formation of the kissing complex [62]. The change in affinity indicates a coperativity between ligand binding and kissing complex formation. This stabilization has been estimated to be $\Delta\Delta G \approx -6$ kcal/mol. This number should be interpreted with caution since it is based on the assumption that the mutated aptamer mimics a ligand-bound state that is accessible to the wild type aptamer [62]. Results obtained with Amber force field are in qualitative agreement with this picture.

We also observe that estimated stability of the kissing loop complex in our calculation, namely ΔG = -52 $\pm$ 2 kcal/mol and ΔG = -35 $\pm$ 3 kcal/mol for holo and apo respectively, is much larger than expected. This could be partly due to the overestimation of stacking interactions in the Amber force field [150]. Results obtained with CHARMM indicate a lower ΔG for both the systems [71], in better agreement with experimental results, even if it is still overestimated. Recently, the thermodynamics of other stand-alone kissing complexes has been also studied using different biophysical techniques [151, 152]. In these two experimental works the stability of the loop-loop interactions of their interest resulted to be -8 $\lesssim \Delta G \lesssim$ -14 kcal/mol. These results shown that different kissing loop complexes with different sequences as well as different sets of intra-loop interactions have different stabilities, that are however defined in a quite small range. However, we would like to point out that the overestimation found in our calculations could also be a consequence of difficult convergence in the US simulations. To test if the US simulation are effectively converged, we tried to recover the profiles from simulations initialized with the backward process, with a procedure inspired by two directional pulling in steered MD [88, 108, 109]. The discrepancy between forward and backward process is an index of high dependence of the PMF on the initialization procedure and poses some questions on the actual convergence of the US simulations. Similarly to steered MD, one can expect that simulations in the forward and backward process are respectively overestimating and underestimating the kissing complex stability. Optimal results can be obtained in steered MD by combining simulations performed with both protocols [108, 109]. We stress here that even if the forward simulations apparently recover the qualitative behavior of the general accepted model, they cannot be trusted for a quantitative estimation of the free-energy change. The fact that the backward process cannot reach the native docked state is a signature of a barrier in an orthogonal degree of freedom that is not properly sampled.

A possible candidate is the barrier related to the desolvation of the loops, required to form the correct interstrand interactions. Additionally, we observe that pulling on the distance between the two loops does not necessarily induce the entropic reduction required upon docking. These issues are expected to affect both forward and backward pulling. Our simulation could not give an estimate of the additional barriers, but we can assume that these issues equally affects the Holo and the Apo systems. Thus, the converged $\Delta\Delta G$ upon ligand binding should be somewhere in between the results from the forward and the backward simulations and we can expect it to be in a wide range between +2.7 kcal/mol and −17 kcal/mol, which is in qualitative agreement with already mentioned experiments [62].

The convergence problem is not related to the US method itself but to the difficulty of describing such a complex docking event using a single distance as a CV. This variable is not sufficient to drive the system through the appropriate transition states. This is likely due to the existence of additional barriers on hidden degrees of freedom (e.g. solvation). We believe that in order to reliably quantify the $\Delta\Delta G$ for this system with US or other biased sampling methods one should employ more complex CVs which are closer to the actual reaction coordinate.

With this aim we have carried out the HREx simulations biasing not only the distance between the CoM of the loops nucleotides but also the inter-loop native contacts with WTMetaD. In our simulation the $\Delta\Delta G$ between apo and holo in the presence of K^+ is -18 ± 6.5 kJ/mol (≈ 5 kcal/mol). From the above mentioned experiments this difference, related to the favorable linkage between the loop docking and organization of the binding pocket around the ligand, results to be $\approx$ 20 kJ/mol. Thus, our results are in nice agreement with the experiments under these conditions. From the analysis of the apo and holo simulations in the presence of Mg^{2+} we estimated $\Delta\Delta G = -8 \pm 5$ kJ/mol (-2 kcal/mol), thus smaller than the experimental result in which this difference has been -27 kJ/mol (-6.5 kcal/mol). An explanation for this difference could be that in our simulations the strong stabilization given by the divalent cations, as clearly shown in Figure 4.6 comparing the Holo/Mg^{2+} and the Holo/K^+ profile or the Apo/Mg^{2+} and the Apo/K^+ profile, was able to mask the effect given by the ligand. The comparison between the $\Delta\Delta G$ obtained by means of our combined approach and the ones estimated with the US simulations confirmed that an in-depth thermodynamics characterization of the kissing loop interaction with more accurate and sophisticated techniques was necessary.

The ΔG between the close and open conformations of the loop-loop interaction in the presence of the magnesium are less overestimated than the ones obtained with the US simulations, for both the holo and apo forms, but however larger in comparison with the

experiments. In the presence of potassium, assuming a two-state equilibrium between the open and close conformation of the loops, the ΔG for the apo form (-17 ± 5 kJ/mol) is larger than the measurements published by Leipply and Draper (≈ -8 kJ/mol) [62]. Instead, for the ligand-bound form these experiments they have not reported the value, whereas in our simulations $\Delta G = -35 \pm 6$ kJ/mol. Another quantity that can be analyzed in our simulations is the Mg^{2+}-RNA interaction free energy (ΔG_{RNA-2+}). It can be obtained making the difference between the minima of the profile in Figure 4.6, considering the holo systems (black and green curves) and the apo ones (red and blue) separately. Our data showed for the holo form $\Delta G_{RNA-2+} = -25 \pm 7$ kJ/mol and $\Delta G_{RNA-2+} = -35 \pm 5$ for the apo form.

In the already mentioned work [62] with which we compared our thermodynamics results, small-angle X-ray scattering (SAXS) has been used to monitor docking of the hairpin loops by changes in the overall dimensions of the aptamer 4.9. The data provided by these experiments allow also a comparison with our results on the radius of gyration, finding a good general quantitative and qualitative agreement as can be seen in Table 4.1. The only effective discrepancy between our simulation and the experiments can be found in the analysis of the holo form with the docked loops. Indeed, for this system in the presence of monovalent cations the X-ray scattering has found a compact structure with $R_g = 20.6$ Å on which the titration with Mg^{2+} have no significant influence, whereas in our simulations in the presence of K^+ the structure is more extended/relaxed than the one in solution with magnesium.

TABLE 4.1: Gyration radius analysis of the aptamer with docked and undocked loops. Comparison between our simulations and SAXS experiments. The values are in Å.

Systems	docked	docked	undocked	undocked
	sim	exp	sim	exp
Apo/K^+	21.5-22.5	24.8	24.5	24.9
Holo/K^+	21.5-22.5	20.6	24.0	24.9
Apo/Mg^{2+}	20.0/20.5	20.8	≈22.0	24.0
Holo/Mg^{2+}	20.5	20.3	≈22.0	22.8

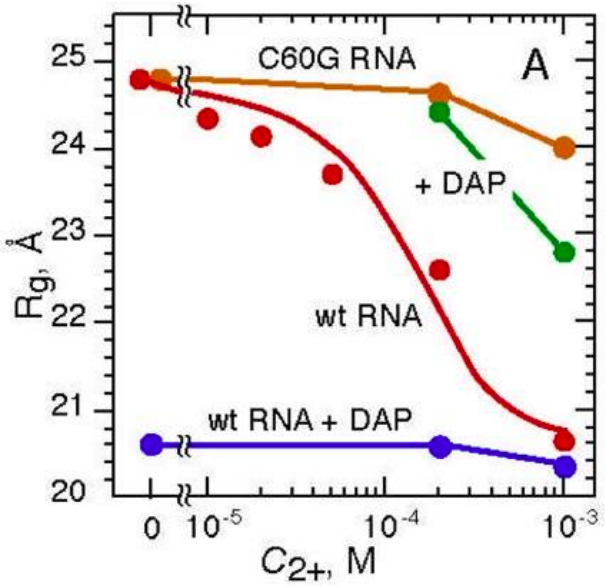

FIGURE 4.9: SAXS studies of different forms of the A-riboswitch aptamer in background buffer of K^+. R_g as a function of bulk Mg^2+ concentration (C_{2+}) for the aptamer with the docked loops in the holo (blue) and the apo form (red), or with the undocked loops in the holo (green) and apo form (orange). The ligand is 2,6-diaminopurine (DAP), where present. Adapted from Leipply and Draper [62].

Chapter 5

Conclusions

5.1 Folding Model

All the different parts of the PhD project here reported provide atomistic details and energetic estimates to the currently accepted model for the folding of the *add* riboswitch upon ligand binding [57, 62, 63]. Altogether, our data and the related experimental works suggest a folding model as depicted in Figure 5.1.

Initially, only the P2 and P3 stems and the corresponding loops (L2, L3, still not interacting each other) are formed and not fully stable (Figure 5.1A). After, the binding pocket begin to pre-organize to recognize the ligand [58, 153, 154] structuring the three junctions and the two loops become prompt to interact (Figure 5.1B).

In the presence of K^+ the formation of the kissing loop interaction is not favorable thus, the aptamer cannot properly fold. Only adding Mg^{2+} (Figure 5.1C_1 and C_2) it is possible for L2 and L3 to come closer and establish an interaction. Adenine binding allows for a complete organization of the aptamer where the three junctions tightly arrange around the ligand (Figure 5.1C_1), stabilizing also the previously formed helices [57]. It has not been clearly established by the experiments if the loop-loop interaction is formed before or after ligand binding [62], but the hypotheses lean towards an adenine concentration-dependent process (as discussed later on). In this context, an alternative pathway in which the loop-loop interaction is formed but not completely stable and the P1 could acquire a partially folded conformation in the absence of adenine is also possible (Figure 5.1C_2) [63].

Finally, the P1 helix becomes fully structured and stabilized by the ligand (Figure 5.1D), to the detriment of the expression platform (see Figure 1.4) [123]. This step is mandatory for translation to be initiated.

We quantified the ligand contribution to the P1 stem formation due to direct interactions to be approximately -4 kJ/mol. Our simulations quantified the cooperative linkage between ligand binding and the formation of the kissing loop complex in -8 kJ/mol. In the matter of the different ions effect, the stabilization of the loop-loop interaction given by Mg^{2+} is ≈-25 kJ/mol in the presence of the ligand, and ≈-35 kJ/mol in its absence.

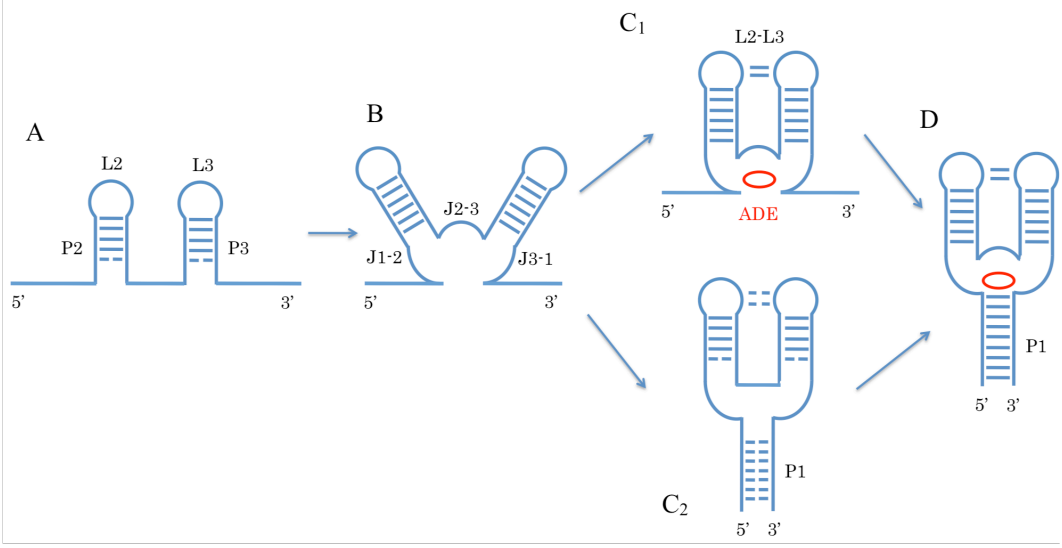

FIGURE 5.1: Schematic representation of the aptamer secondary structure in its folding intermediates in the presence of magnesium with and without the ligand. A) Stems P2 and P3, loops L2 and L3 are folded but not stable. B) The binding pocket begin to preorganize to recognize the ligand structuring the three junctions (J1-2, J2-3, J3-1). Thus, also the 2 loops (L2 and L3) become prompt to interact. C_1) The junctions arrange around the ligand (ADE) and the inter-loop interaction occurs (L2-L3) stabilizing also the stems. C_2) Alternative possible intermediate in which all the 3 stems are not completely and stably folded before ligand binding. Thus, also the junctions and the kissing loops are not stable. D) P1 stem is fully folded and stabilized by the ligand.

Our result on the long range stabilization on the kissing loops given by the ligand and on the ligand-induced P1 stem stabilization is compatible with both the folding pathways (Figure 5.1 panels C_1 and C_2) irrespectively of their relative population and cannot discriminate among them. The relative probability between the two paths can be modulated by the ligand concentration and its binding affinity. Thus, only performing simulations of the ligand binding would make it possibile to distinguish the two states. This in principle would also allow us to estimate a difference in free energy between them. On the one hand, the intermediate shown in Figure 5.1C_1 could be relevant for ligand-RNA binding in an early transcriptional context in which the last 9 nucleobases (i.e those allowing P1 formation) of the aptamer have not yet been synthesized. Indeed, it has been shown that also an aptamer missing a large portion of the P1 stem is able to bind adenine [124], even if with lower binding affinity. On the other hand, the intermediate shown in Figure 5.1C_2 could be populated at low-ligand concentration once the nucleobases allowing P1

formation are synthesized. Later on, after the synthesis of the expression platform, ligand binding could shift the thermodynamic equilibrium towards one of two competing riboswitch conformations (P1 formed and non formed).

5.2 Final remarks

Molecular dynamics at atomistic resolution in combination with enhanced sampling techniques is a powerful tool and can be used to predict the thermodynamics of conformational changes in riboswitches and to generate experimentally testable hypothesis. Ligand-induced stabilization of the P1 stem is crucial for A-riboswitch regulation and function. Here we quantified the direct interaction between adenine and P1 stem and analyzed it at atomistic detail. We have shown how it is possible to use steered MD to quantify the ligand-induced effect on the stabilization of the terminal helix of an adenine riboswitch aptamer domain. Results are in agreement with available thermodynamic data and can help in the structural interpretation of single molecule experiments. Statistical error can be kept under control with a bootstrap analysis. Our results suggest a model for the aptamer folding in which the direct P1-ligand interactions play a minor role on the conformational switch when compared with those related to the ligand-induced aptamer preorganization. Because the structural/functional role of the aptamer terminal helix is a common feature in the "straight junctional" riboswitches [21], we foresee a wider validity in particular for the last step of the model presented herein. Furthermore in this work we addressed the formation of the kissing loop complex in the A-riboswitch aptamer by means of accurate molecular dynamics simulations in explicit solvent combined with enhanced sampling techniques in presence or absence of the cognate ligand, and in the presence of monovalent or divalent cations. Results obtained with the multiple restraints umbrella sampling simulations are qualitatively in agreement and compatible with experiments suggesting that the presence of the ligand stabilizes the kissing loop formation. However, our results also spot some weakness of the umbrella sampling method, calling for calculations performed with more advanced techniques, namely as we did, a combination of the Hamiltonian replica exchange with well tempered metadynamics. Taking advantage of this combined approach we refined our model obtaining more accurate thermodynamics results on the effect of adenine binding on the loop-loop interaction formation. In addition, we estimated the different stabilization given by monovalent and divalent cations to the loops docking also getting good structural data in nice agreement with the experiments.

We can conclude this dissertation saying that this project successfully shed a new light on the ligand-dependent folding mechanism of the *add* adenine sensing riboswitch characterizing at atomistic details and with thermodynamics reliable data the several conformational changes essential for its biological function in the gene expression regulation.

Acknowledgements

The people I would acknowledge are (in no particular order):

- My sup, because when I sent him my wishes for his last birthday he made me notice that it was the first time in 4 years I have called him "boss", thus I started thinking about it..and my answer to his surprise is that during these years I have considered him more as a older brother than a boss. Someone that is able to let me understand every-day working life because he already experienced it few years before me. Someone that is able to entertain my ideas and consider them not as junk only because I am a student and he is the professor. Someone that is able to shake me with the proper words and habits when it is necessary.

- My partner, she probably deserves a nobel prize for putting up with me. Because these last 4 years have been crowd of grim times and barren of lighthearted moments, but these few important ones and what we have previously built together were enough to do not make extinguishable our ardour by the wind of the distance. And now we are already looking forward to meet each other again along our way.

- My "enlarged" family, for misunderstanding all my ideas, but nodding and smiling all the same. For their unconditional support even if they are still trying to fully understand what I have done during my master thesis internship after the final presentation that was in italian. How many years they will take to catch something of my PhD defense in english?

- My friends, and the fews included in my short list surely will be able to recognize his/her name hided in this sentence, for taking my mind off the PhD letting me overflying Frosinone, Roma (or wherever they come from), Milano and Uberlândia even if I was alone in my rented apartment in Trieste.

- My department, my present and former colleagues, because it would be rude to not thank them and no-one has been a passenger of my car without leaving something on the seat.

- "my" Tri(e)ste because this city was able to let me grow as a man by dint of Bora.

List of Figures

List of Tables

Bibliography

[1] F. H. C. Crick. On protein synthesis. In *Symposia of the Society for Experimental Biology*, volume 12, page 138, 1958.

[2] F. H. C. Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, 1970.

[3] E. N. Jackson and C. Yaetofsky. The region between the operator and first structural gene of the tryptophan operon of *Escherichia coli* may have a regulatory function. *Journal of molecular biology*, 76(1):89–101, 1973.

[4] M. Arnaud, M. Débarbouillé, G. Rapoport, M. H. Saier, and J. Reizer. In vitro reconstitution of transcriptional antitermination by the SacT and SacY proteins of Bacillus subtilis. *Journal of Biological Chemistry*, 271(31):18966–18972, 1996.

[5] S. Aymerich and M. Steinmetz. Specificity determinants and structural features in the RNA target of the bacterial antiterminator proteins of the BglG/SacY family. *Proceedings of the National Academy of Sciences*, 89(21):10410–10414, 1992.

[6] F. Houman, M. R. Diaz-Torres, and A. Wright. Transcriptional antitermination in the *bgl* operon of *E. coli* is modulated by a specific RNA binding protein. *Cell*, 62 (6):1153–1163, 1990.

[7] Y. Lu, R. J. Turner, and R. L. Switzer. Function of rna secondary structures in transcriptional attenuation of the bacillus subtilis pyr operon. *Proceedings of the National Academy of Sciences*, 93(25):14462–14467, 1996.

[8] M. Oda, N. Kobayashi, A. Ito, Y. Kurusu, and K. Taira. cis-acting regulatory sequences for antitermination in the transcript of the bacillus subtilis hut operon and histidine-dependent binding of hutp to the transcript containing the regulatory sequences. *Molecular microbiology*, 35(5):1244–1254, 2000.

[9] L. V. Wray and S. H. Fisher. Analysis of bacillus subtilis hut operon expression indicates that histidine-dependent induction is mediated primarily by transcriptional antitermination and that amino acid repression is mediated by two mechanisms:

regulation of transcription initiation and inhibition of histidine transport. *Journal of bacteriology*, 176(17):5466–5473, 1994.

[10] P. Babitzke and C. Yanofsky. Reconstitution of bacillus subtilis trp attenuation in vitro with TRAP, the trp RNA-binding attenuation protein. *Proceedings of the National Academy of Sciences*, 90(1):133–137, 1993.

[11] F. F. Costa. Non-coding rnas: lost in translation? *Gene*, 386(1):1–10, 2007.

[12] J. S. Mattick. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO reports*, 2(11):986–991, 2001.

[13] J. Couzin. HapMap launched with pledges of $100 million. *Science*, 298(5595): 941–942, 2002.

[14] A. L Edwards and R. T Batey. Riboswitches: A common RNA regulatory element. *Nat. Edu.*, 3(9):9, 2010.

[15] A. S. Mironov, I. Gusarov, R. Rafikov, L. E. Lopez, K. Shatalin, R. A. Kreneva, D. A. Perumov, and E. Nudler. Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell*, 111(5):747, 2002.

[16] A. Nahvi, N. Sudarsan, M. S. Ebert, X. Zou, K. L. Brown, and R. R. Breaker. Genetic control by a metabolite binding mrna. *Chem. Biol.*, 9(9):1043–1049, 2002.

[17] M. Mandal and R. R. Breaker. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.*, 5(6):451–463, 2004.

[18] A. Roth and R. R. Breaker. The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.*, 78:305–334, 2009.

[19] A. Serganov and D. J. Patel. Metabolite recognition principles and molecular mechanisms underlying riboswitch function. *Ann. Rev. Biophys.*, 41:343, 2012.

[20] R. R. Breaker. Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.*, 4(2), 2012.

[21] A. Serganov and E. Nudler. A decade of riboswitches. *Cell*, 152(1):17–24, 2013.

[22] R. K. Montange and R. T. Batey. Riboswitches: emerging themes in RNA structure and function. *Annu. Rev. Biophys.*, 37:117–133, 2008.

[23] M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, and R. R. Breaker. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell*, 113(5):577–586, 2003.

[24] P. Daldrop, F. E. Reyes, D. A. Robinson, C. M. Hammond, D. M. Lilley, R. T. Batey, and R. Brenk. Novel ligands for a purine riboswitch discovered by rna-ligand docking. *Chemistry & biology*, 18(3):324–335, 2011.

[25] F. Colizzi, A. M. Lamontagne, D. A. Lafontaine, and G. Bussi. Probing riboswitch binding sites with molecular docking, focused libraries, and in-line probing assays. *Methods in molecular biology*, 1103:141, 2014.

[26] K. D. Warner, P. Homan, K. M. Weeks, A. G. Smith, C. Abell, and A. R. Ferré-D'Amaré. Validating fragment-based drug discovery for biological rnas: Lead fragments bind and remodel the TPP riboswitch specifically. *Chemistry & biology*, 21 (5):591–595, 2014.

[27] J. Sund, C. Lind, and J. Åqvist. Binding site preorganization and ligand discrimination in the purine riboswitch. *The Journal of Physical Chemistry B*, 2014.

[28] K. F. Blount and R. R. Breaker. Riboswitches as antibacterial drug targets. *Nat. Biotechnol.*, 24(12):1558–1564, 2006.

[29] J. Mulhbacher, E. Brouillette, M. Allard, L. C. Fortier, F. Malouin, and D. A. Lafontaine. Novel riboswitch ligand analogs as selective inhibitors of guanine-related metabolic pathways. *PLoS Pathog.*, 6(4):e1000865, 2010.

[30] K. E. Deigan and A. R. Ferré-D'Amaré. Riboswitches: Discovery of drugs that target bacterial gene-regulatory RNAs. *Accounts Chem. Res.*, 44(12):1329–1338, 2011.

[31] T. Kubodera, M. Watanabe, K. Yoshiuchi, N. Yamashita, A. Nishimura, S. Nakai, K. Gomi, and H. Hanamoto. Thiamine-regulated gene expression of *Aspergillus oryzae thiA* requires splicing of the intron containing a riboswitch-like domain in the 5'-UTR. *FEBS letters*, 555(3):516–520, 2003.

[32] J. E. Galagan, S. E. Calvo, C. Cuomo, L-J. Ma, J. R. Wortman, S. Batzoglou, S-I. Lee, M. Baştürkmen, C. C. Spevak, and J. Clutterbuck. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature*, 438 (7071):1105–1115, 2005.

[33] P. Borsuk, A. Przykorska, K. Blachnio, M. Koper, J. M. Pawlowicz, M. Pekala, and P. Weglenski. L-arginine influences the structure and function of arginase mRNA in *Aspergillus nidulans*. *Biological chemistry*, 388(2):135–144, 2007.

[34] N. Sudarsan, J. E. Barrick, and R. R. Breaker. Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA*, 9(6):644–647, 2003.

[35] M. T. Cheah, A. Wachter, N. Sudarsan, and R. R. Breaker. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature*, 447(7143): 497–500, 2007.

[36] W. C. Winkler and R. R. Breaker. Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol.*, 59:487–517, 2005.

[37] JN. Kim and RR. Breaker. Purine sensing by riboswitches. *Biol. Cell*, 100(1):1–11, 2008.

[38] R. T. Batey. Structure and mechanism of purine-binding riboswitches. *Q. Rev. Biophys.*, 1(1):1–37, 2012.

[39] E. B. Porter, J. G. Marcano-Velázquez, and R. T. Batey. The purine riboswitch as a model system for exploring RNA biology and chemistry. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 2014.

[40] J. N. Kim, A. Roth, and R. R. Breaker. Guanine riboswitch variants from mesoplasma florum selectively recognize 2'-deoxyguanosine. *Proceedings of the National Academy of Sciences*, 104(41):16092–16097, 2007.

[41] A. Roth, W. C. Winkler, E. E. Regulski, B. W. K. Lee, J. Lim, I. Jona, J. E. Barrick, A. Ritwik, J. N. Kim, R. Welz, Breaker R. R., and others. A riboswitch selective for the queuosine precursor preQ$_1$ contains an unusually small aptamer domain. *Nature structural & molecular biology*, 14(4):308–317, 2007.

[42] W. C. Winkler, A. Nahvi, N. Sudarsan, J. E. Barrick, and R. R. Breaker. An mrna structure that controls gene expression by binding s-adenosylmethionine. *Nature Structural & Molecular Biology*, 10(9):701–707, 2003.

[43] K. A. Corbino, J. E. Barrick, J. Lim, R. Welz, B. J. Tucker, I. Puskarz, M. Mandal, N. D. Rudnick, and R. R. Breaker. Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in α-proteobacteria. *Genome biology*, 6(8):R70, 2005.

[44] R. T. Fuchs, F. J. Grundy, and T. M. Henkin. The s$_{MK}$ box is a new SAM-binding RNA for translational regulation of SAM synthetase. *Nature structural & molecular biology*, 13(3):226–233, 2006.

[45] Z. Weinberg, J. E. Barrick, Z. Yao, A. Roth, J. N. Kim, J. Gore, J. X. Wang, E. R. Lee, K. F. Block, N. Sudarsan, Neph S. Tompa M. Ruzzo W. L., and Breaker R. R. Identification of 22 candidate structured rnas in bacteria using the CMfinder comparative genomics pipeline. *Nucleic acids research*, 35(14):4809–4819, 2007.

[46] J. X. Wang, E. R. Lee, D. Rivera-Morales, J. Lim, and R. R. Breaker. Riboswitches that sense s-adenosylhomocysteine and activate genes involved in coenzyme recycling. *Molecular cell*, 29(6):691–702, 2008.

[47] L. C. Christiansen, S. Schou, P. Nygaard, and H. H. Saxild. Xanthine metabolism in bacillus subtilis: characterization of the *xpt-pbuX* operon and evidence for purine-and nitrogen-controlled expression of genes involved in xanthine salvage and catabolism. *Journal of bacteriology*, 179(8):2540–2550, 1997.

[48] M. Mandal and R. R. Breaker. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.*, 11(1):29–35, 2003.

[49] F. Di Palma, S. Bottaro, and G. Bussi. Kissing loop interaction in adenine riboswitch: insights from umbrella sampling simulations, 2014. submitted manuscript.

[50] R. T. Batey, S. D. Gilbert, and R. K. Montange. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, 432 (7015):411–415, 2004.

[51] A. Serganov, Y. R. Yuan, O. Pikovskaya, A. Polonskaia, L. Malinina, A. T. Phan, C. Hobartner, R. Micura, R. R. Breaker, and D. J. Patel. Structural basis for discriminative regulation of gene expression by adenine-and guanine-sensing mrnas. *Chem. Biol.*, 11(12):1729–1741, 2004.

[52] F. Di Palma, F. Colizzi, and G. Bussi. Ligand-induced stabilization of the aptamer terminal helix in the add adenine riboswitch. *RNA*, 19(11):1517–1524, 2013.

[53] F. Harada and S. Nishimura. Possible anticodon sequences of tRNAHis, tRNAAsn, and tRNAAsp from escherichia coli. universal presence of nucleoside Q in the first position of the anticodons of these transfer ribonucleic acid. *Biochemistry*, 11(2): 301–308, 1972.

[54] J. K. Wickiser, M. T. Cheah, R. R Breaker, and D. M. Crothers. The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry*, 44(40):13404–13414, 2005.

[55] J. K. Wickiser, W. C. Winkler, R. R. Breaker, and D. M. Crothers. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Molecular cell*, 18(1):49–60, 2005.

[56] S. D. Gilbert, C. D. Stoddard, S. J. Wise, and R. T. Batey. Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *J. Mol. Biol.*, 359(3):754, 2006.

[57] R. Rieder, K. Lang, D. Graber, and R. Micura. Ligand-induced folding of the adenosine deaminase a-riboswitch and implications on riboswitch translational control. *Chembiochem*, 8(8):896–902, 2007.

[58] J. Noeske, C. Richter, M. A. Grundl, H. R. Nasiri, H. Schwalbe, and J. Wöhnert. An intermolecular base triple as the basis of ligand specificity and affinity in the guanine-and adenine-sensing riboswitch RNAs. *P. Natl. Acad. Sci. U. S. A.*, 102 (5):1372–1377, 2005.

[59] D. R. Forsdyke. A stem-loop "kissing" model for the initiation of recombination and the origin of introns. *Molecular biology and evolution*, 12(5):949–958, 1995.

[60] J. Nowakowski and I. Tinoco Jr. RNA structure and stability. In *Seminars in virology*, volume 8, pages 153–165. Elsevier, 1997.

[61] R. K. Montange and R. T. Batey. Structure of the s-adenosylmethionine riboswitch regulatory mRNA element. *Nature*, 441(7097):1172–1175, 2006.

[62] D. Leipply and D. E. Draper. Effects of Mg^{2+} on the free energy landscape for folding a purine riboswitch RNA. *Biochemistry*, 50(14):2790–2799, 2011.

[63] M. K. Lee, M. Gal, L. Frydman, and G. Varani. Real-time multidimensional NMR follows RNA folding with second resolution. *P. Natl. Acad. Sci. U. S. A.*, 107(20): 9192–9197, 2010.

[64] K. Neupane, H. Yu, D. A. N. Foster, F. Wang, and M. T. Woodside. Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism. *Nucleic Acids Res.*, 39(17):7677–7687, 2011.

[65] J-C. Lin, C. Hyeon, and D. Thirumalai. RNA under tension: Folding landscapes, kinetic partitioning mechanism, and molecular tensegrity. *The journal of physical chemistry letters*, 3(23):3616–3625, 2012.

[66] M. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulations*. Oxford University Press, 2008.

[67] J. C. Lin and D. Thirumalai. Relative stability of helices determines the folding landscape of adenine riboswitch aptamers. *Journal of the American Chemical Society*, 130(43):14080–14081, 2008.

[68] M. Sharma, G. Bulusu, and A. Mitra. Md simulations of ligand-bound and ligand-free aptamer: Molecular level insights into the binding and switching mechanism of the add a-riboswitch. *RNA*, 15(9):1673–1692, 2009.

[69] U. Priyakumar and A. D. MacKerell. Role of the adenine ligand on the stabilization of the secondary and tertiary interactions in the adenine riboswitch. *J. Mol. Biol.*, 396(5):1422–1438, 2010.

[70] Z. Gong, Y. Zhao, C. Chen, and Y. Xiao. Role of ligand binding in structural organization of add a-riboswitch aptamer: a molecular dynamics simulation. *J. Biomol. Struct. Dyn.*, 29(2):403, 2011.

[71] O. Allnér, L. Nilsson, and A. Villa. Loop-loop interaction in adenine-sensing riboswitch: A molecular dynamics study. *RNA*, 19:916–926, 2013.

[72] J-C. Lin, C. Hyeon, and D. Thirumalai. Sequence-dependent folding landscapes of adenine riboswitch aptamers. *Phys. Chem. Chem. Phys.*, 16:6376–6382, 2014.

[73] J. L. Boots, M. D. Canny, E. Azimi, and A. Pardi. Metal ion specificities for folding and cleavage activity in the schistosoma hammerhead ribozyme. *RNA*, 14 (10):2212–2222, 2008.

[74] D. Thirumalai, N. Lee, S. A. Woodson, and D. K. Klimov. Early events in RNA folding. *Annual review of physical chemistry*, 52(1):751–762, 2001.

[75] S. A. Woodson. Metal ions and RNA folding: a highly charged topic with a dynamic future. *Current opinion in chemical biology*, 9(2):104–109, 2005.

[76] J. A. Beebe, J. C. Kurz, and C. A. Fierke. Magnesium ions are required by Bacillus subtilis ribonuclease P RNA for both binding and cleaving precursor tRNAAsp. *Biochemistry*, 35(32):10493–10505, 1996.

[77] A. L. Feig and O. C. Uhlenbeck. 12 The role of metal ions in RNA biochemistry. *Cold Spring Harbor Monograph Archive*, 37:287–319, 1999.

[78] D. E. Draper. A guide to ions and RNA structure. *RNA*, 10(3):335–343, 2004.

[79] D. E. Draper, D. Grilley, and A. M. Soto. Ions and RNA folding. *Annu. Rev. Biophys. Biomol. Struct.*, 34:221–243, 2005.

[80] D. Grilley, A. M. Soto, and D. E. Draper. Mg^{2+}−RNA interaction free energies and their relationship to the folding of RNA tertiary structures. *Proceedings of the National Academy of Sciences*, 103(38):14003–14008, 2006.

[81] O. Allnér, L. Nilsson, and A. Villa. Magnesium ion–water coordination and exchange in biomolecular simulations. *Journal of Chemical Theory and Computation*, 8(4):1493–1502, 2012.

[82] F. Di Palma, F. Colizzi, and G. Bussi. Using reweighted pulling simulations to characterize conformational changes in riboswitches, 2014. submitted manuscript.

[83] F. Di Palma and G. Bussi. Structural and thermodynamics investigation of the role of ligand binding and Mg^{2+} in the adenine riboswitch folding. manuscript in preparation.

[84] S. Bottaro, F. Di Palma, and G. Bussi. The role of nucleobase interactions in RNA structure and dynamics, 2014. submitted manuscript.

[85] R. O. Dror, C. Young, and D. E. Shaw. Anton, a special-purpose molecular simulation machine. In *Encyclopedia of Parallel Computing*, pages 60–71. Springer, 2011.

[86] C. Dellago and G. Hummer. Computing equilibrium free energies using non-equilibrium molecular dynamics. *Entropy*, 16(1):41–61, 2013.

[87] C. Abrams and G. Bussi. Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy*, 16(1):163–199, 2013.

[88] H. Grubmuller, B. Heymann, and P. Tavan. Ligand-binding-molecular mechanics calculation of the streptavidin biotin rupture force. *Science*, 271(5251):997–999, 1996.

[89] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23 (2):187–199, 1977.

[90] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.

[91] Y. Sugita and Y. Okamoto. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chemical Physics Letters*, 329(3):261–270, 2000.

[92] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78(14):2690, Apr 1997.

[93] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E*, 56:5018–5035, Nov 1997.

[94] F. Colizzi and G. Bussi. RNA unwinding from reweighted pulling simulations. *J. Am. Chem. Soc.*, 134(11):5173–5179, 2012.

[95] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *J. Comput. Chem.*, 13(8):1011–1021, 1992.

[96] B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *The Journal of Chemical Physics*, 27(5):1208–1209, 1957.

[97] B. J. Alder and T. E. Wainwright. Studies in molecular dynamics. I. general method. *The Journal of Chemical Physics*, 31(2):459–466, 1959.

[98] A. Rahman. Correlations in the motion of atoms in liquid argon. *Physical Review*, 136(2A):A405, 1964.

[99] F. H. Stillinger and A. Rahman. Improved simulation of liquid water by molecular dynamics. *The Journal of Chemical Physics*, 60(4):1545–1557, 1974.

[100] J. A. McCammon, B. R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 267:585–590, 1977.

[101] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Physical review letters*, 55(22):2471, 1985.

[102] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical review*, 136 (3B):B864, 1964.

[103] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133, 1965.

[104] B. Isralewitz, M. Gao, and K. Schulten. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struc. Biol.*, 11(2):224–230, 2001.

[105] David D. L. Minh and J. A. McCammon. Springs and speeds in free energy reconstruction from irreversible single-molecule pulling experiments. *The Journal of Physical Chemistry B*, 112(19):5892–5897, 2008.

[106] J. Gore, F. Ritort, and C. Bustamante. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proceedings of the National Academy of Sciences*, 100(22):12564–12569, 2003.

[107] G. Hummer and A. Szabo. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proceedings of the National Academy of Sciences*, 98 (7):3658–3661, 2001.

[108] D. D. L. Minh and A. B. Adib. Optimized free energies from bidirectional single-molecule force spectroscopy. *Physical review letters*, 100(18):180602, 2008.

[109] T. N. Do, P. Carloni, G. Varani, and G. Bussi. RNA/peptide binding driven by electrostatics insight from bidirectional pulling simulations. *Journal of Chemical Theory and Computation*, 9(3):1720–1730, 2013.

[110] DJ Klein, PB Moore, and TA Steitz. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.*, 340(1):141–177, 2004.

[111] P. Gendron, S. Lemieux, and F. Major. Quantitative analysis of nucleic acid three-dimensional structures. *J. Mol. Biol.*, 308(5):919–936, 2001.

[112] A Barducci, G. Bussi, and M. Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Physical review letters*, 100 (2):020603, 2008.

[113] M. Bonomi, A. Barducci, and M. Parrinello. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *Journal of computational chemistry*, 30(11):1615–1621, 2009.

[114] S. Piana and A. Laio. A bias-exchange approach to protein folding. *The Journal of Physical Chemistry B*, 111(17):4553–4559, 2007.

[115] M. R. So and A. F. Voter. Temperature-accelerated dynamics for simulation of infrequent events. *The Journal of Chemical Physics*, 112(21):9599–9606, 2000.

[116] L. Maragliano and E. Vanden-Eijnden. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chemical physics letters*, 426(1):168–175, 2006.

[117] L. Wang, R. A. Friesner, and B. J. Berne. Replica exchange with solute scaling: A more efficient version of replica exchange with solute tempering (rest2). *The Journal of Physical Chemistry B*, 115(30):9431–9438, 2011.

[118] U. H. E. Hansmann. Parallel tempering algorithm for conformational studies of biological molecules. *Chemical Physics Letters*, 281(1):140–150, 1997.

[119] G. Bussi, A. Laio, and M. Parrinello. Equilibrium free energies from nonequilibrium metadynamics. *Physical review letters*, 96(9):090601, 2006.

[120] G. Bussi. Hamiltonian replica exchange in GROMACS: a flexible implementation. *Molecular Physics*, 112(3-4):379–384, 2014.

[121] C. Dellago and P. G. Bolhuis. Transition path sampling and other advanced simulation techniques for rare events. *Adv. Polym. Sci.*, 221:167–233, 2009.

[122] E. Darve and A. Pohorille. Calculating free energies using average force. *The Journal of chemical physics*, 115(20):9169–9183, 2001.

[123] J. F. Lemay, G. Desnoyers, S. Blouin, B. Heppell, L. Bastet, P. St-Pierre, E. Massé, and D. A Lafontaine. Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. *PLoS Genet.*, 7(1):e1001278, 2011.

[124] J-C. Lemay and D. A. Lafontaine. Core requirements of the adenine riboswitch aptamer for ligand binding. *RNA*, 13(3):339–350, 2007.

[125] J. Wang, P. Cieplak, and P. A. Kollman. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, 21(12):1049–1074, 2000.

[126] A. Pérez, I. Marchán, D. Svozil, J. Šponer, T. E. Cheatham III, C. A. Laughton, and M. Orozco. Refinement of the amber force field for nucleic acids: Improving the description of α/γ conformers. *Biophys. J.*, 92(11):3817–3829, 2007.

[127] M. Zgarbová, M. Otyepka, J. Šponer, A. Mládek, P. Banáš, T. E. Cheatham III, and P. Jurečka. Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J Chem Theory Comput*, 7(9):2886–2902, 2011.

[128] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general amber force field. *J. Comput. Chem.*, 25(9):1157–1174, 2004.

[129] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model. *J. Phys. Chem.*, 97(40):10269–10280, 1993.

[130] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A.

Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision C.02, 2004. Gaussian, Inc., Wallingford, CT.

[131] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. Lincs: a linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18(12):1463–1472, 1997.

[132] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An n·log(n) method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089, 1993.

[133] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926, 1983.

[134] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1):014101–014107, 2007.

[135] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81: 3684, 1984.

[136] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, and M. Parrinello. Plumed: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.*, 180(10):1961–1972, 2009.

[137] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi. PLUMED 2: New feathers for an old bird. *Computer Physics Communications*, 185(2):604–613, 2014.

[138] N. B Leontis and E. Westhof. Geometric nomenclature and classification of rna base pairs. *RNA*, 7(4):499–512, 2001.

[139] A. C. Pan, T. M. Weinreich, Y. Shan, D. P. Scarpazza, and D. E Shaw. Assessing the accuracy of two enhanced sampling methods using egfr kinase transition pathways: the influence of collective variable choice. *Journal of Chemical Theory and Computation*, 10(8):2860–2865, 2014.

[140] N. B. Leontis, J. Stombaugh, and E. Westhof. The non-watson–crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, 30(16):3497–3531, 2002.

[141] D. H. Mathews, M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker, and D. H. Turner. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *P. Natl. Acad. Sci. U. S. A.*, 101(19):7287–7292, 2004.

[142] D. H. Turner and D. H. Mathews. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.*, 38 (suppl 1):D280–D282, 2010.

[143] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, 125(2):167–188, 1994.

[144] R. Lorenz, S. H. Bernhart, C. H. Zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, and IL. Hofacker. ViennaRNA Package 2.0. *Algorithm. Mol. Biol.*, 6(1): 26, 2011.

[145] G. Bussi, F. L. Gervasio, A. Laio, and M. Parrinello. Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *Journal of the American Chemical Society*, 128(41):13435–13441, 2006.

[146] N. Foloppe and A. D. MacKerell Jr. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal of Computational Chemistry*, 21(2):86–104, 2000.

[147] A. D. MacKerell and N. K. Banavali. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *Journal of Computational Chemistry*, 21(2):105–120, 2000.

[148] A. Grossfield. Wham: an implementation of the weighted histogram analysis method. version 2.0.9, 2013. http://membrane.urmc.rochester.edu/content/wham.

[149] J. Šponer, J. Leszczynski, and P. Hobza. Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Biopolymers*, 61(1):3–31, 2001.

[150] P. Banáš, A. Mládek, M. Otyepka, M. Zgarbová, P. Jurečka, D. Svozil, F. Lankaš, and J. Šponer. Can we accurately describe the structure of adenine tracts in B-DNA? reference quantum-chemical computations reveal overstabilization of stacking by molecular mechanics. *Journal of Chemical Theory and Computation*, 8(7): 2448–2460, 2012.

[151] N. Salim, R. Lamichhane, R. Zhao, T. Banerjee, J. Philip, D. Rueda, and A. L. Feig. Thermodynamic and kinetic analysis of an rna kissing interaction and its resolution into an extended duplex. *Biophysical journal*, 102(5):1097–1107, 2012.

[152] W. Stephenson, P. N. Asare-Okai, A. A. Chen, S. Keller, R. Santiago, S. A. Tenenbaum, A. E. Garcia, D. Fabris, and P. T. X. Li. The essential role of stacking adenines in a two-base-pair rna kissing complex. *Journal of the American Chemical Society*, 135(15):5602–5611, 2013.

[153] A. L. Edwards and R. T. Batey. A structural basis for the recognition of 2'-deoxyguanosine by the purine riboswitch. *Journal of molecular biology*, 385(3): 938–948, 2009.

[154] S. D. Gilbert and R. T. Batey. Monitoring RNA–ligand interactions using isothermal titration calorimetry. In *Riboswitches*, pages 97–114. Springer, 2009.