HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Conformational dynamics of the ligand-free state of a riboswitch investigated by molecular simulations

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Christian Alexander Hanke
aus Mutlangen

Düsseldorf, August 2016

aus dem Institut für Pharmazeutische und Medizinische Chemie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Holger Gohlke

Korreferent: Prof. Dr. Martin Lercher

Tag der mündlichen Prüfung:

30.09.2016

Eidesstattliche Erklärung

lch versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Diese Dissertation wurde in der vorgelegten oder einer ähnlichen Form noch bei keiner anderen Institution eingereicht, und es wurden bisher keine erfolglosen Promotionsversuche von mir unternommen.

Düsseldorf, 30. August 2016

Table of Contents

List of publications

This thesis is based on the following manuscripts (Contribution in parentheses):

I. Hanke, C.A. (70%), Gohlke, H. - Force field dependence of riboswitch dynamics; Methods in Enzymology 2015, Vol. 553, p. 163-191.

Impact factor reported for 2014: 2.088

II. Hanke, C.A. (70%), Gohlke, H. - Ligand-mediated and tertiary interactions cooperatively stabilize the P1 region in the guanine-sensing riboswitch; PLOS ONE, 2017. **12**(6): p. e0179271.

Impact factor reported for 2015/2016: 3.54

III. Hanke, C.A. (70%), Gohlke, H. - Tertiary interactions in the unbound guanine-sensing riboswitch focus functional conformational variability to the binding site

Submitted (2017).

Abbreviations

CD	circular dichroism
c-di-AMP	cyclic di-adenosine monophosphate
ci-di-GMP	cyclic di-guanosine monophosphate
CRISPR	clustered regulatory interspaced short palindromic repeats
DNA	deoxyribonucleic acid
FIRST	Floppy Inclusions and Rigid Substructure Topography
FMN	flavin mononucleotide
Gsw	guanine-sensing riboswitch aptamer
Gsw^{apt}	guanine-sensing riboswitch aptamer wild type variant
Gsw^{loop}	guanine-sensing riboswitch aptamer G37A/C61U mutant variant
MBAR	multistate Bennet acceptance ratio
MD	molecular dynamics
miRNA	microRNA
mRNA	messenger RNA
NMR	nuclear magnetic resonance
PDB	protein data bank
PMF	potential of mean force
preQ$_1$	pre-queuosine$_1$
REMD	replica exchange molecular dynamics
RMSD	root-mean-square deviation
RMSF	root-mean-square fluctuation
RNA	ribonucleic acid
rRNA	ribosomal RNA
SAH	S-adenosylhomocysteine
SAM	S-adenosylmethionine
siRNA	small interfering RNA
TPP	thiamine pyrophosphate
tRNA	transfer RNA
WHAM	weighted histogram analysis method

Abstract

Riboswitches are a class of regulatory RNA molecules mainly found in bacterial mRNA. They typically consist of an aptamer domain and an expression platform and regulate gene expression based on the availability of a ligand. Upon binding of the ligand to the aptamer domain, the expression platform undergoes a conformational change, which results in altered gene expression. Understanding of the unbound state of riboswitches is crucial as a first step towards understanding the riboswitch-based gene regulation process. However, detailed structural knowledge of the unbound state is still scarce despite experimental and computational efforts, likely due to the marginally stable character of the riboswitch that is required for fast switching after ligand binding. In this thesis, I investigated the unbound state of a guanine-sensing riboswitch aptamer domain using molecular simulations. I particularly focused on the stability of tertiary interactions formed between two loops and their influence on the aptamer domain. To this end, I used a wildtype guanine-sensing riboswitch aptamer and a mutant, which features destabilized tertiary interactions and exhibits a Mg^{2+}-dependent ligand binding behavior.

Initially, I had to decide for a setup for the molecular dynamics (MD) simulations with respect to the force field and the initial configuration of the simulation system. The chosen setup was required to reproduce the experimentally observed differences between wildtype and mutant. Based on this setup, I performed MD simulations of the two aptamer variants at different Mg^{2+} concentrations. From these simulations, I observed that the mutation has a long-range destabilizing effect from the loop region to the ~30 Å distant ligand binding site. This destabilizing effect is compensated by the presence of Mg^{2+} ions. Furthermore, I found that stable tertiary interactions together with the stabilization of the ligand binding site can cooperatively stabilize the terminal P1 region. The results from this study suggest that the information on the stability of the tertiary interactions and on the ligand binding are not transferred via large conformational changes, but rather via changes in dynamics. This would allow for a fast response of a preorganized aptamer upon ligand binding.

In order to gain insights into the conformational heterogeneity of the unbound state, I performed replica exchange MD simulations and calculated free energy landscapes for the aptamers. From these results I found that for the wildtype, a homogeneous conformational ensemble close to the native bound state is energetically favorable, which exhibits local flexibility to allow for ligand binding. In contrast, for the mutant I observed a more heterogeneous structural ensemble in the unbound state, which includes alternative base pairs and folds. Overall, the results suggest that the

destabilized tertiary interactions in the mutant shift the folding kinetics from a directed folding without notable traps for the wildtype towards a slower folding, which includes alternative conformations.

In my thesis, I gained insights into the unbound state, long-range effects in the aptamer and the role of the tertiary interactions in the stability and conformational heterogeneity of the guanine-sensing riboswitch aptamer. Such insights could be useful for the development of novel antibiotic drugs targeting riboswitches, where information about the ligand binding site is crucial, or for the design of synthetic riboswitches, where a particular behavior of the aptamer is desired.

Zusammenfassung

Riboswitches sind eine Gruppe von regulatorischen RNA-Molekülen, die überwiegend in bakterieller mRNA vorkommt. Sie bestehen üblicherweise aus einer Aptamerdomäne und einer Expressionsplattform und regulieren die Expression von Genen abhängig von der Verfügbarkeit eines Liganden. Das Binden des Liganden an die Aptamerdomäne induziert eine Konformationsänderung in der Expressionsplattform, wodurch die Genexpression beeinflusst wird. Das Verständnis des ungebundenen Zustandes eines Riboswitches ist ein wichtiger erster Schritt hin zum Verständnis der Riboswitch-basierten Genregulation. Trotz experimentellen und computergestützten Anstrengungen ist das detaillierte strukturelle Verständnis des ungebundenen Zustandes noch begrenzt. Ein möglicher Grund hierfür ist der marginal stabile Charakter des Riboswitches, der für eine schnelle Reaktion nach der Ligandbindung notwendig ist.

In dieser Arbeit habe ich den ungebundenen Zustand einer Guanin-bindenden Riboswitch-Aptamerdomäne mittels molekularer Simulationen untersucht. Hierbei konzentrierte ich mich insbesondere auf die Stabilität von tertiären Wechselwirkungen, die zwischen zwei Loop-Regionen gebildet werden, und deren Einfluss auf die Aptamerdomäne. Hierfür verwendete ich eine Wildtyp-Aptamerdomäne eines Guanin-bindenden Riboswitches und eine Mutante, welche destabilisierte tertiäre Wechselwirkungen aufweist und hierdurch ein Mg^{2+}-abhängiges Ligandbindungsverhalten zeigt.

Zunächst musste ich ein geeignetes Setup bezüglich des Kraftfeldes und der Startkonfiguration für die Moleküldynamik (MD)-Simulationen wählen. Das gewählte Setup musste in der Lage sein, experimentell beobachtete Unterschiede zwischen dem Wildtyp und der Mutante reproduzieren zu können. Aufbauend auf diesem Setup führte ich MD-Simulationen der beiden Aptamervarianten bei unterschiedlichen Mg^{2+} Konzentrationen durch. In diesen Simulationen beobachtete ich, ausgehend von der destabilisierten Loop-Region, eine Destabilisierung der ~30 Å entfernten Ligandbindestelle. Dieser destabilisierende Effekt konnte durch die Anwesenheit von Mg^{2+}-Ionen in der Simulation ausgeglichen werden. Darüber hinaus fand ich heraus, dass stabile tertiäre Wechselwirkungen in der Loop-Region zusammen mit einer stabilisierten Ligandbindestelle einen kooperativen, stabilisierenden Einfluss auf die P1-Region haben, an welche die Expressionsplattform anschließt. Die Ergebnisse dieser Studie legen nahe, dass Informationen über die Stabilität der tertiären Wechselwirkungen und über die Ligandbindung nicht durch große strukturelle Änderungen, sondern vielmehr durch Änderungen in der Dynamik übertragen werden.

Solch eine Übertragung würde eine schnelle Reaktion auf die Bindung eines Liganden erlauben. Um einen Einblick in die konformationelle Heterogenität des ungebundenen Zustandes zu erhalten, führte ich replica exchange MD-Simulationen durch und berechnete freie Energie-Landschaften. Die Ergebnisse dieser Berechnungen zeigen, dass der ungebundene Zustand des Wildtyps aus einem homogenen Strukturensemble besteht, welches dem gebundenen Zustand ähnlich ist. Gleichzeitig besitzen die Strukturen jedoch lokale Flexibilität um die Bindung eines Liganden zu erlauben. Im Gegensatz hierzu beobachtete ich für den ungebundenen Zustand der Mutante ein heterogeneres Strukturensemble, welches alternative Basenpaarungen und Faltungen einschließt. Insgesamt deuten diese Ergebnisse darauf hin, dass die destabilisierten tertiären Wechselwirkungen in der Mutante die Faltungskinetik von einem gerichteten Faltungsprozess ohne nennenswerte Intermediate hin zu einem langsameren Faltungsprozess verschiebt, welcher alternative Konformationen beinhaltet.

Durch die Ergebnisse meiner Arbeit erhielt ich tiefgehende Einblicke in den ungebundenen Zustand, langreichweitige Effekte und den Einfluss der tertiären Wechselwirkungen auf die Stabilität und die konformationelle Heterogenität der Guanin-bindenden Riboswitch Aptamerdomäne. Solche Einblicke können sich für die Entwicklung neuer Antibiotika, die auf Riboswitches abzielen, als hilfreich erweisen. Darüber hinaus können diese Einblicke für die Entwicklung von synthetischen Riboswitches nützlich sein, bei denen ein bestimmtes Verhalten der Aptamerdomäne gewünscht ist.

1 Introduction

Ribonucleic acid (RNA) molecules are involved in a variety of tasks in biological systems. This includes information transfer, as well as catalytic and regulatory functions [1, 2]. In the case of regulatory functions, riboswitches raised increasing interest since their discovery about a decade ago [3-7]. Riboswitches are RNA-based regulatory elements typically found in bacterial mRNA. They undergo a conformational change in response to binding of a specific ligand, and subsequently influence the expression of the genes under their control. Since the riboswitch-based regulation usually happens in early stages of bacterial protein biosynthesis, riboswitches are proposed to be interesting drug targets [5, 7-9]. Accordingly, a detailed understanding of the regulatory decision induced by riboswitches would be valuable for the development of novel drugs targeting riboswitches.

For riboswitches regulating gene expression on the transcriptional level, the unbound state is of particular interest. This is because the riboswitch has to fold, bind a ligand, and induce the conformational change, before the RNA polymerase is out of reach for the regulation [10-13]. The guanine-sensing riboswitch from the *xpt-pbuX* operon of *Bacillus subtilis* is a transcriptionally acting riboswitch, which was subject to a variety of experimental and computational studies [11]. For this riboswitch, tertiary interactions play an important role for the structural stability and influence the ligand binding ability of the aptamer. However, atomic-level structural knowledge of the unbound state is still scarce. In particular, the amount of preorganization required for a quick response, and whether the unbound state is formed by a heterogeneous set of conformations is still under discussion [14, 15].

In my thesis, I set out to gain a deeper understanding of the unbound state of the guanine-sensing riboswitch, in particular with respect to the influence of the tertiary interaction stability on the ligand binding site and on the connection to the regulatory decision. For this, I applied molecular dynamics (MD) simulations and enhanced sampling methods, as these allow for insights into the dynamics of the RNA on the atomic level. As the setup and input for a MD simulation strongly influence the results, I initially searched for the optimal setup to reproduce experimental findings on the system (Section 4). Based on these findings, I performed MD simulations to investigate the differences between the wildtype aptamer and a G37A/C61U mutant, which possesses destabilized tertiary interactions and shows a Mg^{2+} dependence of its ligand binding ability (Section 5). Finally, I performed replica exchange molecular dynamics simulations to overcome the limited sampling capability of standard MD simulations. In this way, it is possible to gain insight into the conformational space and the free energy landscape accessible to the wildtype and the mutant (Section 6) and to determine the conformational heterogeneity of the unbound aptamers.

2 Background

2.1 Nucleic acids in biological systems

Besides proteins, carbohydrates, and lipids, nucleic acids represent one of the important groups of macromolecules found in biological systems. Both deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are formed by highly similar building blocks, however they form different three-dimensional structures. While DNA forms regular, double stranded helices [16] (Figure 1A), RNA forms a variety of complex structures [2] (Figure 1B-E). This might be surprising at first sight, since both nucleic acids are formed by nucleotides, which consist of a nucleobase (adenine (A), cytosine (C), guanine (G), and thymine (T) or uracil (U) for DNA and RNA, respectively) attached to a sugar-phosphate backbone. The differences in the three-dimensional structures formed by DNA and RNA are due to a difference in the sugar-phosphate backbone. In the case of DNA, the sugar is deoxyribose, while for RNA it is ribose. The presence of the hydroxyl group in the 2' position in the case of RNA results in a less stable and more flexible structure compared to DNA [1]. Due to this flexibility, RNA can form different secondary structural motifs, such as loops, bulges, junctions, but also single-stranded and paired regions [17-21]. For both DNA and RNA three-dimensional structures are stabilized by base pairs. In the case of DNA, these are typically Watson-Crick base pairs (G-C and A-T), while for RNA, Watson-Crick base pairs (G-C and A-U) as well as non-Watson-Crick base pairs are found [22, 23].

The difference in three-dimensional structure results in a diversity of tasks carried out by DNA and RNA. While DNA serves as a stable information storage, RNA carries out a variety of tasks in cells, such as information transport, regulation, and catalytic processes [1, 2].

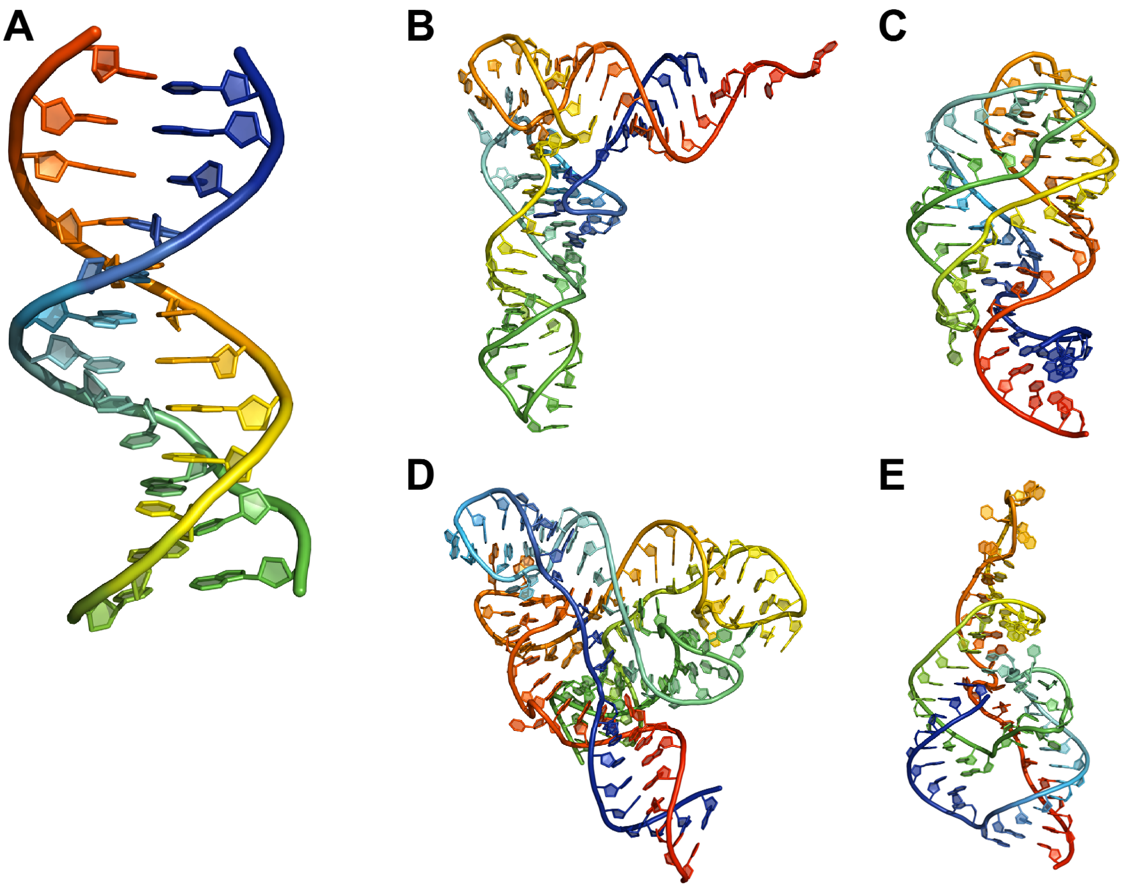


Figure 1: Structures of DNA and RNA. A: DNA dodecamer (PDB ID: 1BNA [24]); B: tRNA (PDB ID: 4TNA [25]); C: guanine-sensing riboswitch (PDB ID: 4FE5 [26]); D: flavin-mononucleotide-sensing riboswitch (PDB ID: 5C45 [27]); E: HDV ribozyme (PDB ID: 4PR6 [28]). The structures are colored using a rainbow scheme from the 5' (blue) to the 3' (red) end.

2.2 Importance of RNA in cellular processes

For many years, the role of ribonucleic acid in biological systems was underestimated and reduced to its role in protein biosynthesis [29]. In the central dogma of protein synthesis, which describes the information flow from DNA to RNA to protein [30, 31], RNA plays three main roles: 1) The messenger RNA (mRNA) transports the information encoded in the DNA towards the ribosome, where the protein is synthesized. 2) The ribosomal RNA (rRNA) makes up large parts (up to 60% [32]) of the ribosome, in particular the active site [33]. 3) The transfer RNA (tRNA) transports amino acids to the ribosome for the incorporation into the new polypeptide chain. rRNA and tRNA both belong to the large group of non-coding RNA [29], i.e., they do not encode for proteins. Nowadays, a large number of non-coding RNAs is known [31], such as ribozymes, small interfering RNA (siRNA), microRNA (miRNA), riboswitches, and clustered regulatory interspaced short palindromic repeats (CRISPR). These RNAs perform a variety of tasks, including regulatory functions. Ribozymes catalyze biochemical reactions similar to protein enzymes [34]. siRNA and miRNA bind to target mRNA in order to inhibit translation of the mRNA and to accelerate its degradation [4, 35]. Riboswitches, which are the main topic of this thesis and which will be discussed later in more detail, bind metabolites and allosterically regulate gene expression [4, 10, 36-39]. CRISPRs are a form of bacterial immune system, where fragments of viral DNA are incorporated in certain regions of the RNA [40]. An overview of these and additional non-coding RNAs is given in Ref. [31]. The finding that RNA can perform regulatory functions, which were initially assigned to proteins exclusively, supports the RNA world hypothesis [41-43]. This hypothesis describes the possibility that early organisms might have relied on RNA exclusively, without the need for DNA and proteins for information storage and regulatory functions. In the RNA world, riboswitches could have played an important regulatory role [39].

The basis to perform the aforementioned variety of tasks is the RNA's ability to form complex three-dimensional structures [2]. The formation of three-dimensional structures is especially important for riboswitches, since they have to form marginally stable aptamers that are able to bind ligands.

2.3 Riboswitches

Riboswitches are short gene regulatory elements of noncoding RNA, which have the ability to sense and bind ligand molecules. The binding of a ligand molecule induces a conformational change of the RNA molecule, which results in the regulation of gene expression [4, 10, 36-39, 44]. Riboswitches are usually found in the 5'-untranslated region of the mRNA, and are thus located upstream of the genes they regulate [45]. Typical ligands for riboswitches are coenzymes, nucleotide derivatives, and amino acids. In this way, many riboswitches regulate genes involved in metabolic pathways [3, 4].

A typical riboswitch consists of two functionally important domains (Figure 2): an aptamer domain, which senses the cellular environment for its ligand and binds it with high specificity, and a downstream expression platform, which undergoes a conformational change when the ligand is bound to the aptamer domain. The aptamer domain and the expression platform often overlap to some extent [4], which suggests the existence of two mutually exclusive riboswitch states [46, 47] regulating the gene expression: ligand-bound and ligand-free state. Upon ligand binding, riboswitches can either activate or repress gene expression, depending on their setup, i.e., on the composition of aptamer domain and expression platform. Therefore, the ligand-bound state can either be the on- or the off-state for the genes. Often riboswitches regulate genes that are involved in the synthesis or transport of the molecules they sense [4, 48]. In this way, riboswitches can represent a feedback mechanism regulating the concentration of metabolites in the cells. A negative feedback, i.e., switching off gene expression upon ligand binding, is more common than a positive feedback, where ligand binding would induce gene expression [49].

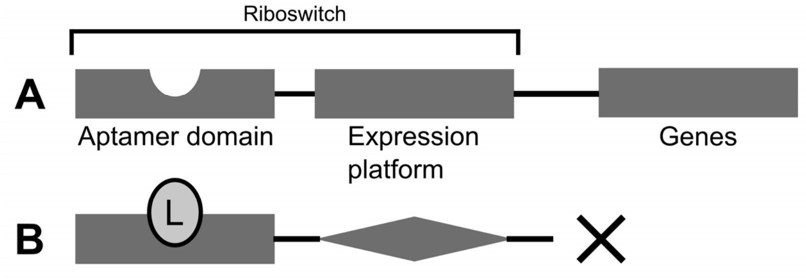


Figure 2: Schematic representation of the composition of an off-switch. The ligand binding-capable aptamer domain and the regulatory expression platform are both located upstream of the genes that are to be regulated. A: In the ligand-unbound state, the genes are expressed. B: If the ligand (L) is bound to the aptamer domain, the expression platform undergoes a conformational change, and thus the genes are not expressed.

The riboswitch aptamer domain is typically evolutionary highly conserved, due to its need for highly specific binding of the ligand molecule [50]. In contrast, the expression platform may vary based on the mechanism employed for gene regulation. For example, the adenine-sensing *pbuE* (also called *ydhL*) riboswitch in *Bacillus subtilis* [51] and the *add* riboswitch in *Vibrio vulnificus* [52] adopt a highly similar aptamer domain and bind adenine with high specificity. At the same time, these riboswitches regulate gene expression via activation upon ligand binding on the transcriptional and on the translational level, respectively [14].

However, aptamer domains of riboswitches sensing the same ligands do not always share structural similarity. There exist several classes of S-adenosylmethionine (SAM) binding riboswitches, which show substantially different folds [50, 53] (Figure 3).

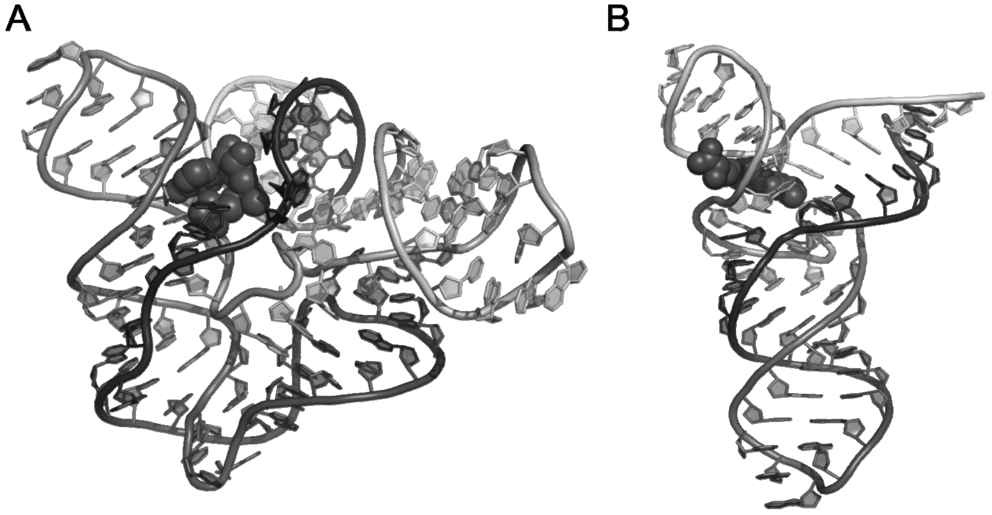


Figure 3: 3D-structures of S-adenosylmethionine (SAM) binding riboswitches. A: SAM-I riboswitch (PDB ID: 2GIS [54]) bound to S-adenosylmethionine (spheres); B: SAM-III riboswitch (PDB ID: 3E5E [55]) bound to S-adenosylhomocysteine (spheres).

Overall, riboswitches sense a variety of molecules, and to date more than 20 classes of riboswitches are reported [39, 50, 56]. The most common type of ligand are small molecules, such as coenzymes (adenosylcobalamin [57, 58], thiamine pyrophosphate (TPP) [59, 60], flavin mononucleotide (FMN) [59, 61], S-adenosylmethionine (SAM) [62-70], S-adenosylhomocysteine (SAH) [71], tetrahydrofolate [72], and molybdenum and tungsten cofactors [73]), nucleotide derivatives [14, 74] (guanine [3, 26], adenine [51], pre-queuosine$_1$ (preQ$_1$) [75, 76], deoxyguanosine [77], cyclic di-guanosine monophosphate (c-di-GMP) [78, 79], and cyclic di-adenosine monophosphate (c-di-AMP) [80]), amino acids (lysine [81-83], glycine [84], and glutamine [85]), and a phosphorylated sugar (glucosamine-6-phosphate [86, 87]). In addition, there are riboswitches that sense the cellular concentration of ions (Mg^{2+} [86, 88, 89], and the halide anion F^- [90]) [49, 56, 91, 92]. Finally, there exist 'orphan' riboswitches, whose ligands are not identified yet [50, 93].

While almost all riboswitches were discovered in bacteria only, riboswitches sensing TPP were also identified in eukaryotes [94], in particular in fungi [95, 96], algae [97], and higher plants [98, 99]. However, no riboswitches have been discovered in humans to date.

2.4 *Mechanisms of gene regulation by riboswitches*

In bacteria, riboswitches are usually found in the 5'-untranslated region of mRNA, where they control gene regulation by various mechanisms [4, 39, 93]. The two most common modes by which riboswitches regulate gene expression in bacteria are the transcriptional and the translational gene regulation. Here, riboswitches can act as *cis*-acting on- or off-switch for the downstream genes. In the following, I will describe the mechanisms for transcriptional gene regulation, translational gene regulation, and I will briefly mention additional mechanisms for riboswitch regulation.

2.4.1 Intrinsic transcriptional gene regulation

In the case of the (rho-independent) transcriptional gene regulation, an intrinsic transcription terminator [100, 101] element is part of the expression platform. The terminator element is composed of a terminator stem followed by five to nine uridines [9, 100, 101]. When the RNA polymerase encounters the transcription terminator loop during transcription, it stops and eventually dissociates from the DNA strand, thereby halting mRNA production [39]. Depending on whether the ligand is bound to the aptamer domain, one of two competing structural elements is formed: the transcription terminator or the so-called antiterminator, which shares part of its sequence with the terminator. If the antiterminator is formed, the formation of the terminator loop is prevented, and thus the RNA polymerase produces the full-length mRNA (Figure 4).

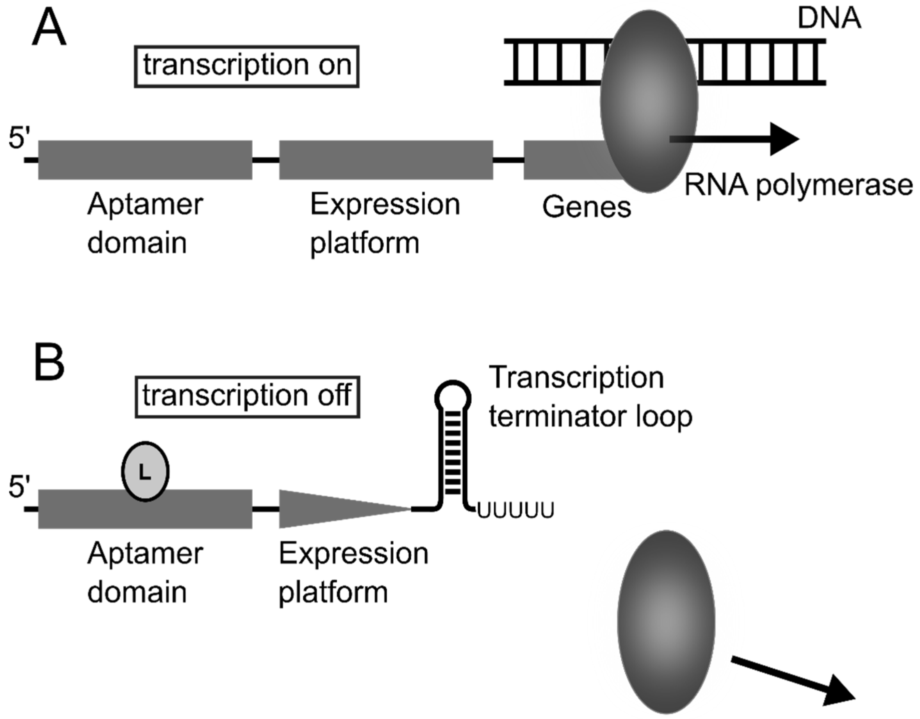


Figure 4: Transcriptional riboswitch gene regulation, exemplarily depicted with an off-switch acting via intrinsic transcription termination. A: Transcription is switched on. The aptamer and the expression platform are transcribed by the RNA polymerase. Since no ligand is bound to the aptamer domain, an anti-terminator loop is formed, and the RNA polymerase transcribes the genes. B: Transcription is switched off due to a ligand (L) bound to the aptamer domain. The transcription terminator loop is formed using a part of the expression platform; this halts RNA polymerase action.

For the transcriptional gene regulation, the regulatory decision whether the downstream genes should be transcribed or whether the transcription should be stopped, has to be made in a small time window of typically only a few seconds [10-12] (Figure 5). Over the course of transcription, first the riboswitch aptamer domain gets transcribed by the RNA polymerase. When (partially) transcribed, the aptamer domain folds, typically supported by Mg^{2+} ions. As soon as the aptamer domain is folded correctly, the riboswitch enters the so-called "sensing phase" [13]. In this phase, the aptamer domain scans the cellular environment for a potential ligand. The sensing phase ends when the RNA polymerase has passed the point in sequence where the transcription terminator loop would be formed, or when the RNA polymerase encountered the transcription terminator. If no terminator is formed, the RNA polymerase will resume transcribing the downstream genes. Thus, the events of encountering and binding of a ligand molecule, transmission of the binding information to the expression platform, and the subsequent structural rearrangement involving the expression platform, that are crucial for gene regulation, occur during the sensing phase. The length of the sensing phase depends on the length of the riboswitch sequence, on specific features of the DNA sequence such as RNA polymerase pause sites [102, 103], and on the rate of transcription of the RNA polymerase. Typical RNA polymerase transcription rates for *Escherichia coli* can be up to ~50 nucleotides per second, and depend on several environmental factors [104-106].

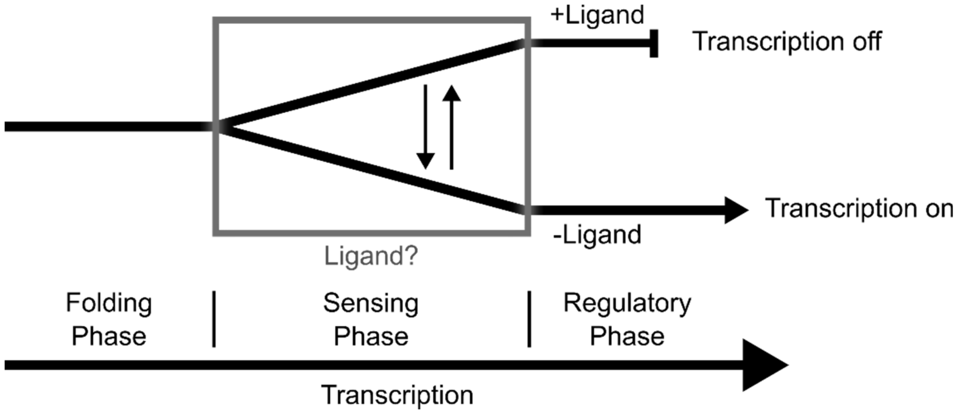


Figure 5: Phases during transcription for a transcriptionally acting off switch. During the folding phase, the riboswitch aptamer folds. When (partially) folded, the aptamer enters the sensing phase. Here, the aptamer searches the cellular environment for a ligand. The sensing phase ends when the RNA polymerase reaches the point in sequence where the transcription terminator loop would be formed. If the aptamer domain encounters a ligand and binds it (Top; +Ligand), the RNA polymerase stops the transcription. Otherwise, the RNA polymerase continues the transcription of the downstream genes (Bottom; -Ligand).

2.4.2 Translational gene regulation

For the translational gene regulation, the expression platform contains the Shine-Dalgarno sequence, which is the ribosomal binding site on the mRNA. Depending on the ligand binding to the aptamer domain, the Shine-Dalgarno sequence is exposed or occluded via alternative base pairing (Figure 6), such that the small ribosomal subunit can or cannot bind to the mRNA and initiate translation [39, 93, 94]. Here, the available time window for the regulatory decision is slightly longer than for the transcriptional regulation, as the ribosome can only bind when the ribosomal binding site in the expression platform is transcribed.

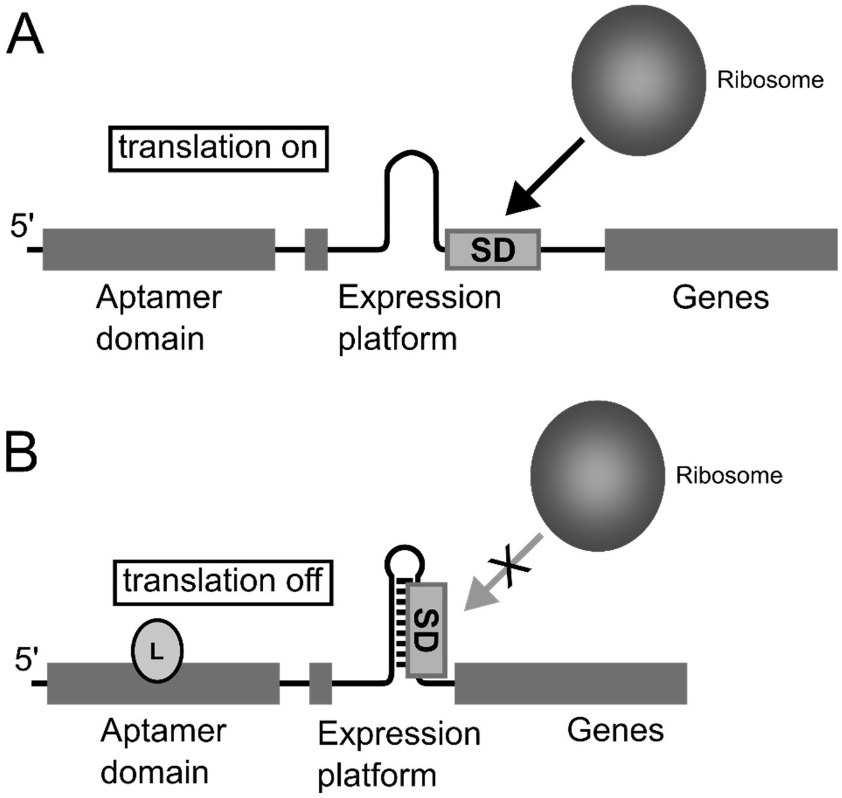


Figure 6: Gene regulation by a translational riboswitch exemplarily depicted with a translationally acting off-switch. A: Translation is switched on. Since no ligand is bound, the ribosomal binding site (SD) is accessible for the ribosome. B: Translation is switched off. If a ligand (L) binds to the aptamer domain, the ribosomal binding site (SD) is occluded via base pairing within the expression platform. In this case, the ribosome cannot bind to the mRNA and thus translation is hindered.

2.4.3 Other mechanisms of gene regulation

Besides these two most frequently found modes of riboswitch gene regulation in bacteria, several other regulation mechanisms by riboswitches were found [91, 93].

In addition to the intrinsic transcription termination described before (Section 2.4.1), which is rho-independent, i.e., does not require the rho-factor protein for transcription termination, rho-dependent transcription termination was observed for a Mg^{2+} sensing riboswitch [107].

The glmS riboswitch-ribozyme mRNA, which senses glucosamine-6-phosphate, was found to undergo self-cleavage upon binding of its ligand [86], representing a negative feedback for glucosamine-6-phosphate production.

In contrast to riboswitches that mainly sense and regulate molecules involved in metabolic pathways [108], riboswitches sensing the second-messenger c-di-GMP were found to be capable of broader regulatory function [92, 93]. Riboswitches sensing c-di-GMP were found in proximity of genes involved in bacterial motility, virulence, and other processes [79]. Some of these riboswitches are located adjacent to group-I self-splicing introns [78], where they induce a complex response to binding of c-di-GMP that involves the self-splicing intron itself. In particular, binding of the c-di-GMP to the aptamer domain of these riboswitches induces a conformational change allowing for self-excision of the intron. This step yields a translatable mRNA by joining the previously split ribosomal binding site [109].

While most riboswitches act in a *cis*-regulatory fashion by influencing the mRNA downstream of the aptamer, the S-adenosylmethionine sensing riboswitches SreA and SreB from *Listeria monocytogenes* were found to be *trans* acting. As such, these riboswitches bind to a distally located mRNA after transcription termination upon ligand binding, thereby downregulating the expression of the virulence regulator PrfA [110].

In eukaryotes, such as fungi [95, 96], algae [97], and higher plants [98, 99], thiamine pyrophosphate sensing riboswitches have been found to control gene expression via alternative splicing.

Additionally, there are riboswitches that do not control gene expression as single units, but employ a tandem riboswitch setup, where two coupled aptamer domains or entire riboswitches are used for gene regulation [111]. An example of a riboswitch, where two aptamer domains control a single expression platform is the glycine riboswitch, for which cooperative binding of glycine to the aptamer sites was suggested [84]. The tandem arrangement found in this riboswitch might allow for increased sensitivity of gene regulation in response to small changes in glycine concentration [112]. However, later studies questioned the cooperativity in this riboswitch when looking at full-length glycine riboswitches [113, 114]. A recent model of the glycine riboswitch mechanism suggests that dimerization of the aptamer domains results in a response in the expression platform [111, 115].

Furthermore, other types of tandem riboswitches have been found to bind different ligands but yet regulate the same genes. The metE tandem riboswitch found in *Bacillus clausii* contains a S-adenosylmethionine and an adenosylcobalamin sensing part, and binding of either of the two ligands results in the repression of gene expression [116].

2.4.4 Riboswitch classification based on binding kinetics

Besides the so far presented classification of riboswitches based on their mechanism of gene regulation, another possibility to classify riboswitches is based on the rate of transcription elongation with respect to the speed of ligand binding. Here, riboswitches can be classified in two groups: 1) If the aptamer domain has enough time to reach thermodynamic equilibrium with its ligand before a regulatory decision is required, a riboswitch is said to be under thermodynamic control. 2) Otherwise, the riboswitch is under kinetic control [6, 12, 103, 117]. In this case, ligand binding has to be fast compared to the speed of RNA transcription. Therefore, a much higher ligand concentration than the in vitro determined K_d value is required to induce a regulatory response [117].

Due to the short time window for the regulation decision for transcriptionally acting riboswitches (Section 2.4.1), these are typically under kinetic control, whereas several translationally acting riboswitches are assumed to work under thermodynamic control [14, 117].

2.5 Ligand binding and the unbound state as key to the regulatory decision

Riboswitch aptamer domains typically bind their ligand molecules by a network of hydrogen bonds or by base stacking interactions, sometimes supported by metal ions [49, 118, 119]. High-resolution structural information about the ligand-bound state of riboswitch aptamer domains arises from a number of crystal structures available for almost all riboswitch classes described before [120]. In addition to these ligand-bound crystal structures, also crystal structures of ligand-free aptamer domains have been reported, e.g., for preQ$_1$-, SAM-I-, lysine-, and glycine-responsive riboswitches [120]. Unbound and bound conformations of the same aptamers are surprisingly similar to each other [49, 120]. One possible reason might be the crystal lattice, which selects minor conformations from solution [14]. Many aptamers, such as the purine- [26, 52, 104], preQ$_1$- [121], lysine- [122, 123], or glycine- [112] sensing riboswitches, envelope their ligands into a binding pocket, suggesting that for ligand binding a conformational change is required. This suggests that the crystal structures of the unbound state might not resemble the true unbound state of the aptamer. Furthermore, it is assumed that the unbound state of the riboswitch is rather comprised of an ensemble of conformations [13, 49, 120, 124]. Some of these conformations might be ligand binding-competent, and ligand binding stabilizes the bound state out of the ensemble. This way of binding is also called "conformational capture" or "conformational selection" [13, 49].

Accordingly, the picture of riboswitches as two-state gene regulation switches had to be adapted recently. Reining *et al.* found a third state for the full-length *add* adenine-sensing riboswitch of *Vibrio vulnificus* [125, 126]. Here, a ligand-bound conformation, which corresponds to the on-state for the gene expression, and two distinct ligand-free conformations, where gene expression is switched off, were observed. One of the two ligand-free conformations was shown to be ligand binding-incompetent, while the other one can bind the ligand and induce translational activation. The relative population of the two unbound states was found to be temperature-dependent, suggesting that the three-state regulation might be important to maintain control of the gene regulation over a range of temperatures [125, 126].

The study by Reining *et al.* also highlights the importance of considering the expression platform to improve understanding of the unbound state of a riboswitch and the regulatory decision [11]. However, so far the expression platform did not receive as much attention as the aptamer domain. In particular, high-resolution structural information such as from X-ray crystallography for the expression platform of riboswitches is still scarce.

In summary, knowledge about the unbound state of most riboswitch aptamers is limited, even though some unbound crystal structures are known. At the same time, such knowledge would be highly important for understanding the riboswitch-controlled gene regulation, particularly because of the importance of the unbound state during the sensing phase (Figure 5) and the ligand binding process.

2.6 Prospects of riboswitches as antibiotic drug targets

The increasing number of drug-resistant bacterial strains highlights the urgency to discover and develop novel antibiotic drugs, and to find novel modes of action for the antibiotic drugs currently available [127]. Due to their role in early stages of bacterial gene regulation, i.e., on the transcriptional and the translational level, and since they often regulate genes important for cell metabolism, riboswitches are considered promising antibacterial drug targets [8]. Accordingly, several studies showed that riboswitches are able to act as antimicrobial targets [9, 128].

In particular, some of these studies showed that known antibacterial compounds, with previously unknown mode of action, actually target riboswitches [5, 129]. Other studies designed compounds specifically targeted at riboswitches [27, 130-132]. Here, I will briefly describe examples of compounds that show antimicrobial activity by binding to bacterial riboswitches.

In 2005, Sudarsan *et al.* showed that pyrithiamine, which is toxic for fungi and bacteria, acts by regulating gene expression via binding to the TPP riboswitch [129]. To this end, the authors used a reporter gene under the control of the TPP riboswitch, whose expression was inhibited by the presence of pyrithiamine. Similarly, Roseoflavin, a natural compound that is synthesized by *Streptomyces davawenensis* and shows antibacterial activity, was found to bind to a FMN riboswitch *in vitro*, and to be able to regulate a reporter gene under the control of a FMN riboswitch [5, 133, 134].

In these studies, the riboswitch-based mode of action was found retrospectively, only after the discovery of the antimicrobial activity of the respective compounds. However, there are also studies describing the rational design of compounds targeting a riboswitch. Kim *et al.* presented an approach in which they rationally designed guanine analogues binding to the guanine-sensing riboswitch, using an available aptamer crystal structure [130]. Several of their analogues were shown to bind to the riboswitch aptamer *in vitro* with affinities comparable to guanine and were able to inhibit bacterial growth. In addition, they found that one of their compounds (6-*N*-hydroxyl-aminopurine) was able to repress the expression of a reporter gene under the control of a guanine riboswitch in *Bacillus subtilis*, suggesting that this compound indeed inhibits bacterial growth via binding to the guanine riboswitch. Instead of guanine analogues, Mulhbacher *et al.* designed a pyrimidine that mimics guanine interactions in the aptamer domain, with the aim of targeting a specific guanine riboswitch [131]. The pyrimidine compound PC1 (2,5,6-triaminopyrimidin-4-one) was found to inhibit bacterial growth in species that used a guanine riboswitch to control the expression of *guaA*, a GMP synthetase. Such bacterial species include *Staphylococcus aureus* and *Clostridium difficile*, which are two clinically highly relevant pathogens [131]. Mulhbacher *et al.* were also able to show that their compound was active in a mouse model. This pyrimidine compound represents a successful case of a rationally designed antibiotic compound targeting a riboswitch in clinically relevant bacterial species. Furthermore, the compound PC1 was tested in milk cows, where it showed promising activity in reducing *Staphylococcus aureus* concentrations, even though stability, dosage, and formulation still need improvements [135]. Recently, Blount *et al.* presented a study in which they characterize a riboflavin analog (5FDQD) that binds to a FMN riboswitch and shows bactericidal activity against *Clostridium difficile in vitro* and in mouse experiments, while at the same time it only modestly affects the cecal flora in mice. Current work focuses on the determination of the exact mode of action of this analog *in vivo* and the importance of FMN riboswitch binding [132]. Howe *et al.* reported the discovery of ribocil, a chemical modulator of bacterial FMN riboswitches, which is structurally distinct from the natural ligand FMN [27].

Ribocil was discovered by performing a phenotypic screening of ~57,000 synthetic small molecules with antibacterial activity. Howe *et al.* showed that the antibacterial activity of ribocil is also given in a mouse model, and furthermore, they were able to co-crystallize an *E. coli* FMN aptamer with ribocil.

In addition to rational design, other approaches aimed at finding potential compounds targeting riboswitches could be used, such as high-throughput screening of compound libraries or fragment-based screening [56]. The previously described cases show that compounds targeting riboswitches may indeed be used as antibacterial drugs. However, further investigations and improvements of the compounds are still required.

2.7 The guanine-sensing riboswitch

In this chapter, I will introduce the guanine-sensing riboswitch and the structural features of its aptamer domain, which were important for my later investigations.

2.7.1 The guanine-sensing riboswitch and other purine-sensing riboswitches

The class of purine-sensing riboswitches contains riboswitches binding guanine, adenine, and 2'-deoxyguanosine [11, 14]. The aptamer domains of all three members of this class share a similar global fold [74]. In the following, I will describe the features of members of this class based on the guanine-sensing riboswitch from the *xpt-pbuX* operon of *Bacillus subtilis*, as I used this system for my investigations. This operon encodes genes that are involved in guanine-metabolism [52]. In this way, an off-switch that senses guanine can represent a negative feedback for the guanine metabolism. Nucleotide sequence numbers or identities might differ between the classes and even between representatives of the guanine-sensing riboswitches [104]. However, nucleotides in structurally important elements, such as the ligand binding site and the interactions between the loops, are highly conserved [14, 104].

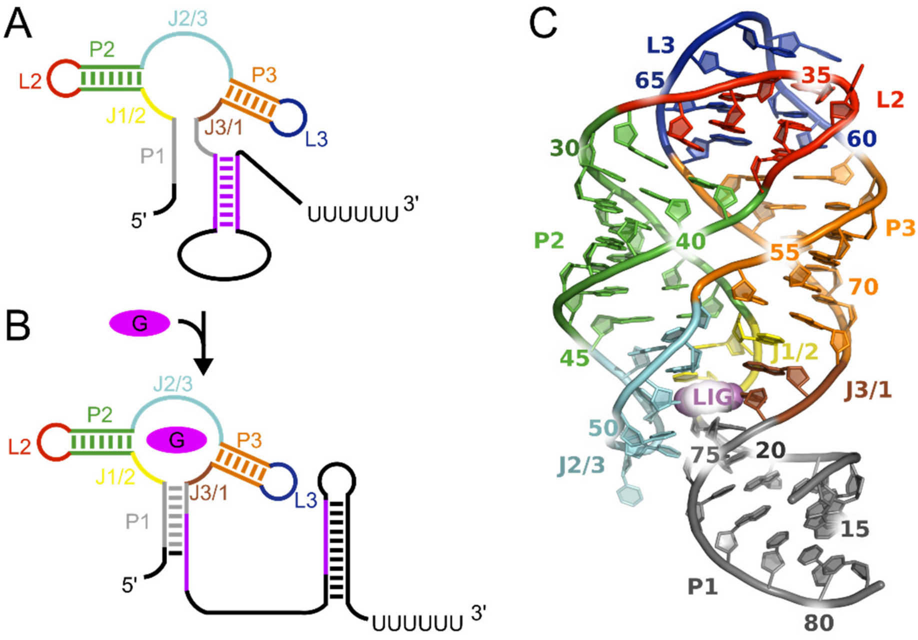


Figure 7: Schematic overview and 3D structure of the guanine-sensing riboswitch aptamer domain. A: Unbound state of the aptamer. The secondary structure elements (color-coded) are assigned according to Ref. [26]. The switching sequence, which is involved in the formation of the terminator loop or the P1 region, is colored in purple. B: Schematic representation of the guanine-bound state of the aptamer. Colors as in A. Guanine is shown as magenta oval with G. C: 3D structure of the aptamer from PDB ID 1U8D [26]. The figure is adapted from Ref. [136].

The guanine-sensing riboswitch can bind guanine, hypoxanthine, and closely related molecules with high specificity [26, 137]. The aptamer domains consist of a three-way junctional fold, which is formed by three paired helices P1, P2, and P3, two loops L2 and L3, which form tertiary interactions, and three joining regions J1/2, J2/3, and J3/1, connecting the paired helices and forming the junction (Figure 7 and Table 1, nomenclature according to [26]). Nucleotides within the loops and the joining regions show the highest conservation [104]. Reviews detailing further structural features of the aptamer domain can be found in Ref. [11, 14, 15, 74, 138].

Table 1: Secondary structure elements and corresponding nucleotides in the aptamer domain of the *xpt-pbuX* guanine-sensing riboswitch.

Secondary structure element[a]	Nucleotides[b]
P1	15-21, 75-81
P2	25-31, 39-45
P3	54-59, 67-72
L2	32-38
L3	60-66
J1/2	22-24
J2/3	46-53
J3/1	73-74

[a] Following assignment in Ref. [26]

[b] Nucleotide numbering according to *xpt-pbuX* guanine-sensing riboswitch aptamer

In the following, I will describe two structurally and functionally highly important structural elements of the guanine-sensing riboswitch aptamer domain: the ligand binding site and the tertiary interactions formed by the loops L2 and L3.

2.7.2 Ligand binding by the guanine-sensing riboswitch aptamer domain

The guanine-sensing riboswitch aptamer domain (Gsw) binds guanine with high affinity (K_d in the order of 5 nM for the *xpt-pbuX* riboswitch from *B. subtilis* [3]), and hypoxanthine with a lower affinity (K_d ≈50 nM for *xpt-pbuX* [3]). The affinity of the guanine-sensing riboswitch aptamer for adenine is ~10,000 fold higher than for guanine [11]. In contrast, the pbuE adenine riboswitch aptamer binds adenine and 2,6-diaminopurine with affinities of 300 nM and 10 nM, respectively, while its affinity for guanine is > 10,000 nM [52]. Interestingly, studies on purine-sensing riboswitches found that the ligand binding pocket can adapt to a certain degree to allow non-cognate ligands, but only interacts weakly with them [137, 139]. Furthermore, the ligand is not completely fixed in the binding pocket, but can sample several conformations in the ligand binding pocket [140].

From a structural point of view, ligand binding within the three-way junction of the guanine-sensing riboswitch is accomplished by the nucleotides U47, U51, C74, and the 2′-hydroxyl group of U22 [14, 26, 141] (Figure 8). Here, the guanine is bound through a Watson-Crick interaction to nucleotide C74, which is always cytosine in guanine-sensing aptamers while it is always uracil in adenine-sensing aptamers. Thus, nucleotide 74 was suspected, and later confirmed, to be responsible for the specificity of the aptamer [52, 142]. Indeed, exchanging C74 by a uracil inverts the specificity of the aptamer towards binding of adenine instead of guanine [142], whereas in the wildtype adenine binding is disfavored by a factor of 10,000 [3, 11] . In this way, structural and ligand binding studies on the (mutated) guanine-sensing aptamer domain were possible using the fluorescent molecule 2-aminopurine [104, 137, 140]. Nucleotide 74 is furthermore assumed to act as initial binding partner for the ligand [143]. Nucleotides 47 and 51, which are also involved in ligand binding, belong to the joining region J2/3. This region is suggested to be the entry point for the ligand, as this is the most flexible region in the absence of the ligand [143].

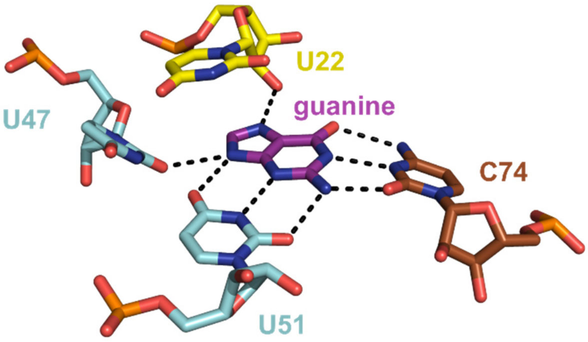


Figure 8: Nucleotides directly binding the guanine ligand (purple) in the guanine-sensing riboswitch aptamer domain. The main binding interactions are formed by nucleotides U47 and U51 in the joining region J2/3 (carbon atoms in cyan), and C74 in J3/1 (carbon atoms in brown). Furthermore a hydrogen bond with the O2' atom of U22 (carbon atoms in yellow) in the J1/2 region can be formed. The colors of the joining regions are the same as in Figure 7.

Current models for the ligand binding suggest an initial low-affinity encounter complex and subsequent interaction of the ligand with nucleotide 74. After this initial binding, the J2/3 region is assumed to act as a lid, closing the entry of the binding pocket, interacting with the ligand via nucleotides 47 and 51, and encapsulating the ligand in the binding pocket [14, 142-145].

While most models agree with this overall process of ligand binding, the extent of preorganization of the ligand binding pocket in the unbound state is still under discussion. Based on a lack of hydrogen bond signals in NMR experiments, Noeske *et al.* described the binding pocket in the unbound aptamer as unstructured [146]. In contrast, Ottink *et al.* observed a more organized ligand binding core [144]. However, they suggested that the observed differences might be due to different buffer conditions, especially with respect to the presence of Mg^{2+} ions in their experiments.

Some degree of preorganization of the ligand binding pocket is likely necessary to allow for fast and efficient binding of the ligand, in particular with respect to the short time window available for productive ligand binding and information transmission.

2.7.3 Tertiary interactions in the guanine-sensing riboswitch aptamer loop regions

The tertiary interactions formed between the loops L2 and L3 are a second important structural feature of purine-sensing riboswitches. In the aptamer domain of the guanine-sensing riboswitch, these tertiary interactions between the L2 and L3 loops (Figure 9A) are formed via two base quadruples (Figure 9B,C), which are located on top of each other and form a complex hydrogen bond network (Figure 9) [26, 52]. Each base quadruple is formed by a Watson-Crick base pair (G37-C61 and G38-C60 for the upper and lower base quadruple, respectively) and a non-canonical base pair (U34-A65 and A33-A66 for the upper and lower base quadruple, respectively) that interacts with the Watson-Crick base pair. Mutational studies found the Watson-Crick base pairs to be the most crucial components of the tertiary loop-loop interactions [147].

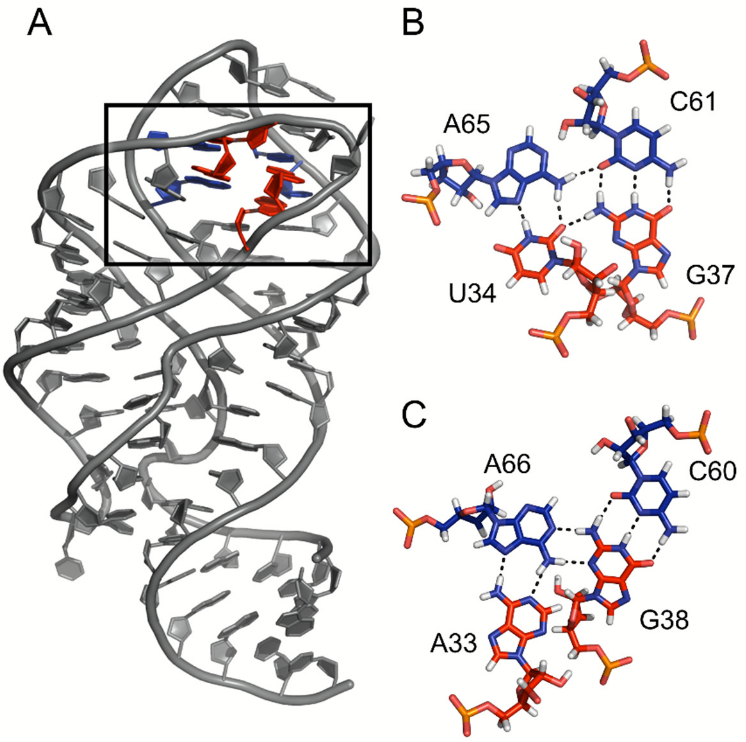


Figure 9: Hydrogen bond network formed by the tertiary interactions between L2 and L3. A: The location of the loop-loop interactions is indicated by a black rectangle; B: Upper base quadruple; C: Lower base quadruple. Hydrogen bonds are indicated by black dashed lines. The nucleotides are colored according to the loop they belong to (L2: red, L3: blue).

The tertiary loop-loop interactions in the guanine-sensing riboswitch were found to be stable even in the ligand-free state in the absence of Mg^{2+} ions [146]. In contrast, the adenine-sensing riboswitch comprises less stable tertiary interactions between the loops. Here, Mg^{2+} is required to shift the conformational equilibrium to structures with formed loop-loop interactions [148, 149]. In the guanine-sensing riboswitch, the interactions between L2 and L3 were found to aid in stabilizing the global tertiary fold of the aptamer domain by keeping P2 and P3 in a parallel arrangement [14, 150]. In this way, the structural basis for the formation of the ligand binding pocket is set by bringing the joining regions in proximity. Interestingly, the tertiary interactions were found to facilitate ligand binding at the distant binding site [143]. This effect is suggested to likely happen via restriction of the conformational freedom in the ligand binding site [143, 151].

Modification of nucleotides involved in the loop-loop interactions mostly results in reduced ligand binding ability of the aptamer [147]. The replacement of the L2 and L3 loops by stable UUCG tetraloops, for example, results in a lack of the tertiary interactions and abolished hypoxanthine binding [26]. Introduction of a G37A mutation in the L2 loop strongly reduces the binding ability for 2,6-diaminopurine in a C74U mutant [147], and for guanine and hypoxanthine in a wildtype variant [143]. Stoddard *et al.* observed that the G37A mutant is able to bind guanine but not hypoxanthine at a temperature of 20°C. In this case, guanine binding resulted in a stabilization of the suboptimal tertiary element. At a higher temperature of 45°C, binding of both guanine and hypoxanthine by the G37A mutant was abolished [143].

Another interesting mutant is the double mutant G37A/C61U created and characterized by Schwalbe *et al.* [146, 151, 152]. In this mutant, the exchange of the stable G37-C61 base pair by the A37-U61 base pair results in a destabilization of the hydrogen bond network forming the tertiary interactions (Figure 10). Interestingly, this mutant shows a Mg^{2+}-dependence of loop-loop interaction formation and of ligand binding [151, 152]. In the absence of Mg^{2+}, the G37A/C61U mutant does not bind the ligand, but the ligand binding ability can be restored by addition of Mg^{2+}. A more detailed description of this mutant and a comparison to the wildtype follows in Section 2.7.5.

Interestingly, some mutations in the P2 regions were found to have a stabilizing influence on the loop-loop interactions, which results in preformation of the interactions in the unbound state [138, 148]. Other studies investigated further mutations in the purine sensing aptamer domains [104, 147, 153]. As I did not investigate these in my work, I will not describe them in detail here.

2.7.4 The aptamer structure in the unbound state

The overall structure of the unbound guanine-sensing riboswitch aptamer domain and its degree of preorganization is still not clear. Different experiments [3, 51, 142, 144-146, 148, 149, 154] suggest that in the absence of ligand and Mg^{2+} ions, the aptamer domains of adenine- and guanine-sensing riboswitches consist of a heterogeneous mixture of conformers in solution [153]. Based on fluorescence experiments, Brenner *et al.* suggested that P2 switches dynamically between two conformations with respect to P1 and P3, and the addition of Mg^{2+} results in an ordered arrangement of P2 next to P3 [155]. Furthermore, it is assumed that the guanine-sensing aptamer is more ordered, and the adenine-sensing aptamer is more conformationally flexible [146, 148]. In particular, NMR and inline-probing experiments suggested that the overall fold of the unbound guanine-sensing riboswitch aptamer is globally preorganized [138, 146], whereas the ligand binding site is largely disordered [142, 143]. The global preorganization in the unbound state is attributed to the stable tertiary interactions in the guanine-sensing riboswitch [146, 148]. In contrast, the adenine-sensing riboswitch requires Mg^{2+} ions for the stable formation of the tertiary interactions [148, 149].

2.7.5 The G37A/C61U mutant and structural and functional differences with respect to the wildtype

In this section, I will point out the major differences between the wildtype and the G37A/C61U mutant investigated by Schwalbe *et al.*, since these aptamers were important for my later investigations.

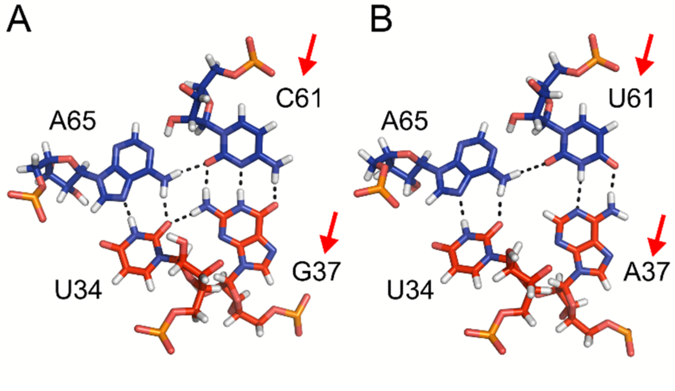


Figure 10: Upper base quadruple for the wildtype (A) the G37A/C61U mutant (B). The mutated nucleotides are marked by a red arrow. The carbon atoms of the nucleotides are colored according to the loops they belong to (L2: red, L3: blue). Black dashed lines indicate hydrogen bonds.

It is worth mentioning, that for both wildtype and G37A/C61U aptamers used in their studies, Schwalbe *et al.* introduced additional, structurally silent mutations (Figure 11). Some of these positions were mutated to stabilize the aptamer in their NMR experiments (U25C, C26U, and A45G in P2 [141, 144]), while others were added to P1 to allow for more efficient transcription

[141, 152]. An indirect influence by the mutations on the aptamer cannot be excluded. However, for comparisons between wildtype and mutant these effects should cancel out, since both variants carry the mutations.

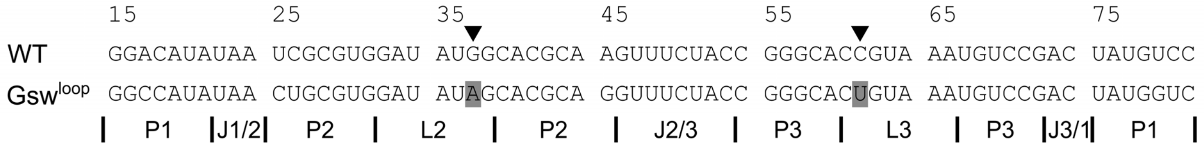


Figure 11: Nucleotide sequences for guanine-sensing aptamer domains. WT corresponds to the wildtype *xpt-pbuX* guanine-sensing riboswitch from [26]. Gsw^{loop} corresponds to the sequence of the G37A/C61U mutant of the guanine-sensing riboswitch from PDB ID 3RKF [152]. Nucleotides 37 and 61 are highlighted in Gsw^{loop} with a grey background. Below the sequences, the assignment to secondary structure elements are shown according to Ref. [26].

The G37A/C61U mutant (also referred to as Gsw^{loop}) was characterized by NMR studies, a crystal structure, and further experiments in the working group of Prof. Schwalbe [138, 146, 151, 152]. The double mutation, located in the L2 and L3 loops, was designed to destabilize the hydrogen bond network connecting the two loops [146, 151, 152]. However, the mutation does not change the overall fold of the structure in the bound crystal structure [152], and thus the two crystal structures of wildtype and mutant are very similar with a root mean square deviation (RMSD) of ~0.85 Å [152]. Interestingly, the mutation induces a Mg^{2+} dependence of the formation of the loop-loop interactions and furthermore, of the aptamer ligand binding ability. In detail, three Mg^{2+} concentration regimes could be detected [151, 152], which yield different structural and functional properties of the aptamer:

- In the **absence of Mg^{2+} ions**, the tertiary loop-loop interactions are not formed. Furthermore, the Gsw^{loop} mutant does not show ligand binding competence for hypoxanthine. Finally, the initial encounter complex [151] observed in NMR experiments for the wildtype [138, 154], is not observed for Gsw^{loop} in the absence of Mg^{2+} ions.
- At **medium Mg^{2+} concentrations** (corresponding to a $[Mg^{2+}]$:[RNA] ratio up to 18:1 [151]), the Gsw^{loop} aptamer is ligand binding-competent, but the transiently formed loop-loop interactions only form stably when the ligand binds.
- At **high Mg^{2+} concentrations** ($[Mg^{2+}]$:[RNA] ratio above 18:1 [151]), the Gsw^{loop} aptamer behaves similar to the wildtype aptamer. In this case, the loop-loop interactions in Gsw^{loop} are preformed, and the aptamer is ligand binding-competent. In addition, the low-affinity initial encounter complex is observed as for the wildtype.

In summary, the G37A/C61U mutation disturbs the formation of the loop-loop interactions, which results in a loss of ligand binding ability of the aptamer. The ligand binding ability can, however, be restored by the addition of Mg^{2+} ions. Thus, the destabilization caused by the mutation in the loop-loop interactions, which influences the distant ligand binding site, could be compensated by the presence of Mg^{2+} ions [138, 146, 151, 152]. In contrast to the Gsw^{loop} aptamer, the wildtype aptamer (Gsw^{apt}) does not show large changes with respect to different Mg^{2+} concentrations. Here, the aptamer possesses preformed loop-loop interactions and is ligand binding-competent even in the absence of Mg^{2+} [151].

In order to investigate the functional stability of the aptamer domains, Buck *et al.* applied circular dichroism (CD) melting experiments to obtain temperature-dependent unfolding transitions [151] for Gsw^{apt} and Gsw^{loop}. They observed two unfolding transitions for the RNA structures, where the tertiary loop-loop interactions are preformed, such as for Gsw^{apt} with (first/second transition at ~21°C/~68°C) and without Mg^{2+} (first/second transition at ~66°C/~79°C), and Gsw^{loop} in the presence of high Mg^{2+} concentrations (first/second transition at ~37°C/~74°C). For structures where the loop-loop interactions are not stably formed, i.e., Gsw^{loop} in the absence of Mg^{2+} or at low Mg^{2+} concentrations, only a single unfolding transition was identified at ~68°C. Combined with their NMR experiments, Buck *et al.* could assign the unfolding transition at lower temperature to the unfolding of the tertiary interactions, whereas the one (in the case of Gsw^{loop} without Mg^{2+} or with low Mg^{2+} concentration the only) at higher temperature could be assigned to the unfolding of the secondary structure elements. Furthermore, the presence of hypoxanthine does not have a significant influence on the unfolding transitions for either of the aptamers. In contrast, the presence of Mg^{2+} has a significant influence on the unfolding transitions, especially with respect to the unfolding of the tertiary interactions. Interestingly, the Mg^{2+}-induced tertiary structure stabilization seems to be more pronounced for Gsw^{apt} than for Gsw^{loop}, which was ascribed to the reduced number of tertiary contacts due to the mutation in Gsw^{loop} [151].

The experiments show that the Gsw^{loop} aptamer represents an interesting test case that can be used to probe the influence of the loop-loop interaction stability on the behavior of the aptamer domain.

2.8 Computational studies of riboswitches

In addition to experimental studies, riboswitches are also subject to several computational studies. Examples of these studies reach from the identification and discovery of riboswitches [156, 157] to the investigation of riboswitch dynamics. The dynamics of riboswitches are often investigated using molecular simulations [158]. By the use of such computational methods, it is possible to gain atomic-level insight into riboswitch states that are experimentally difficult to study. Here, the ligand in complex with the aptamer [159-164] but also the conformational flexibility of the unbound state [13, 161, 165] can be investigated. Furthermore, it is possible to investigate the contribution of specific nucleotides to ligand binding [166, 167], or the influence of a bound ligand on the terminal P1 region [168, 169] or distant tertiary interactions [170]. Additionally, specific processes, such as the breaking of the tertiary interactions in a purine-sensing riboswitch [171], or cotranscriptional folding in coarse-grained simulations [172] can be studied computationally.

Computational studies on the guanine-sensing riboswitch investigated differences in the aptamer hydrogen bond network in the bound and unbound state using MD simulations [159], and studied how different ligands could bind to the aptamer using docking and MD simulations [160], quantum chemical calculations [166], and free energy calculations [167]. Non-equilibrium simulations of the guanine-sensing riboswitch aptamer were used to study the first steps of ligand unbinding [159], and found correlations between the ligand binding site and the tertiary interactions when studying the temperature-induced energy flow in the RNA [170].

Computational studies would be well-suited to investigate the influence of experimentally studied mutations, such as in the P2 tune box [104], on the dynamics and the energetics of the Gsw. These results could yield atomic level insights into how such mutations influence ligand binding. Furthermore, computational investigations of the expression platform and its influence on the aptamer domain might yield important findings on how the aptamer domain and the expression platform interact. Currently, the scarcity of structural information on the expression platform largely excludes MD simulations from such studies without prior modeling the structure of the expression platform.

2.9 Computational methods

2.9.1 Molecular dynamics simulations

Molecular dynamics (MD) simulations are the computational method of choice, if information on biomolecular dynamics on the atomic level is desired. Here, the movement of the atoms in the simulated system is calculated based on Newton's laws of motion [173]. To this end, a force field is used to calculate the potential energy of the system and subsequently to derive the forces acting on each atom of the system.

2.9.1.1 Force fields for MD simulations

The results of a MD simulation is strongly influenced by the choice of the underlying force field [174-177]. In this section, I will briefly describe the components of a common MD force field, using the Amber force field as an example.

A typical MD force field consists of three components: the atom types, the functional form, and the force field parameters [173, 178]. The atom types of a force field describe different chemical environments for the atoms. For the MD simulation, each atom is assigned an atom type. In this way, a carbon atom in an aromatic ring will get a different atom type than a carbon atom in a methyl group [178]. The functional form of the force field describes the computation of the potential energy of a system based on the coordinates of its atoms. The standard functional form for the Amber force field [178] contains terms for bonded and nonbonded interactions:

$$\begin{aligned} E_{total} = & \underbrace{\sum_{bonds} K_r \left(r - r_{eq}\right)^2}_{A} + \underbrace{\sum_{angles} K_\theta \left(\theta - \theta_{eq}\right)^2}_{B} + \\ & \underbrace{\sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\omega - \gamma)]}_{C} + \underbrace{\sum_{i<j} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}}_{D} + \underbrace{\sum_{i<j} \frac{q_i q_j}{\varepsilon R_{ij}}}_{E} \end{aligned} \tag{1}$$

Here, the first three terms (A-C) describe the bonded interactions for bond lengths between two atoms (A), for angles between three covalently bound atoms (B), and for dihedral angles between four atoms (C) (Figure 12). Term D describes the interaction of uncharged particles with respect to their distance, while term E describes Coulombic interactions between charged particles.

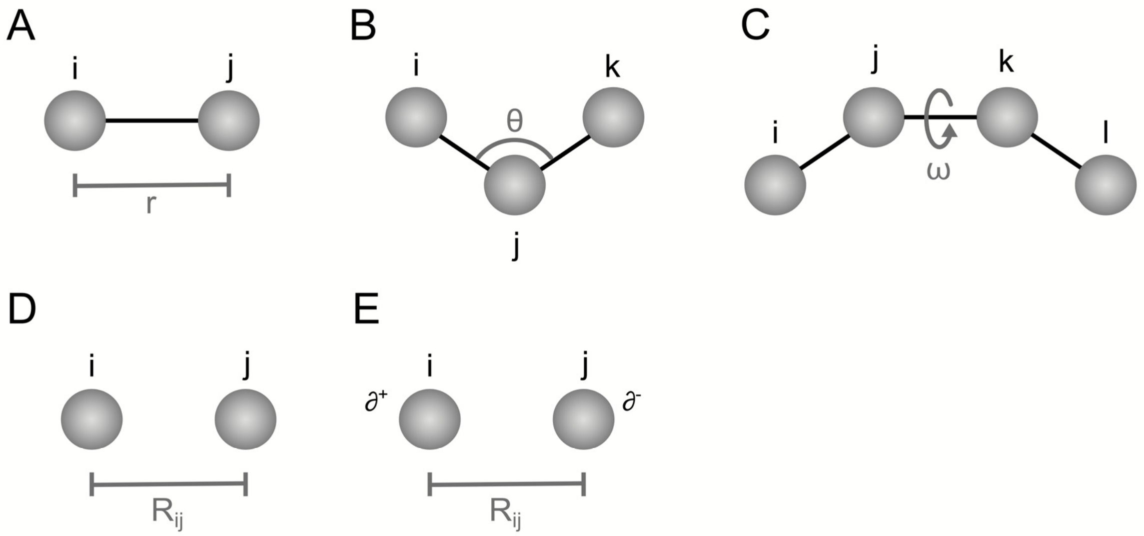


Figure 12: Bonded (A, B, and C) and nonbonded (D and E) interactions described by the force field terms in equation (1). A: Bond length $\boldsymbol{r}$ between two atoms $\boldsymbol{i}$ and $\boldsymbol{j}$; B: Bond angle $\boldsymbol{\theta}$ between three atoms $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$; C: Dihedral angle (four atoms $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$, and $\boldsymbol{l}$) describing the rotation around the bond between atoms $\boldsymbol{j}$ and $\boldsymbol{k}$; D: Van der Waals interactions between two uncharged atoms $\boldsymbol{i}$ and $\boldsymbol{j}$; E: Electrostatic interactions between two atoms $\boldsymbol{i}$ and $\boldsymbol{j}$, here with partial charges $\boldsymbol{\partial^+}$ and $\boldsymbol{\partial^-}$ assigned exemplarily.

The first term (A) in equation (1) describes the energy of a bond as a function of its length between two atoms (Figure 12A); the bond is modeled by a harmonic potential. The term contains two atom type-specific parameters K_r and r_{eq}. K_r represents a force constant for the harmonic potential and describes how much the bond length between the atoms can vary, while r_{eq} is the equilibrium bond length describing the typical length of the bond between the atoms. In addition, r represents the current length of the bond. The second term (B) in equation (1) describes the energy of a bond as a function of its angle between three atoms (Figure 12B). Similarly to the first term, K_θ is a force constant of a harmonic potential, θ_{eq} is the equilibrium angle, and θ is the current angle. The third term (C) in equation (1) describes the energy of a bond as a function of a dihedral angle, i.e., the rotation around one bond considering four atoms (Figure 12C). Here, V_n is again an atom type-specific force constant. Torsion angles are modelled by a cosine series expansion, where n represents the multiplicity and describes the number of energy minima, γ is the phase describing the location of the minima, and ω is the current value of the dihedral angle.

The nonbonded interactions, i.e., the last two terms (D, E) in equation (1), are distance-dependent. Here, R_{ij} represents the distance between the two atoms i and j. Term D describes van der Waals interactions between uncharged atoms (Figure 12D). The given terms with the exponents 12 and 6 are also known as Lennard-Jones potential [173, 179] and represent a repulsive (A_{ij}) and an attractive term (B_{ij}), respectively. Term E of the equation describes Coulombic interactions between charged atoms i and j possessing charges q_i and q_j (Figure

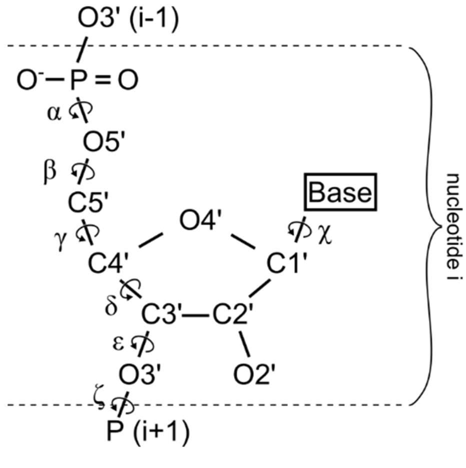
12E). Since the nonbonded interactions are calculated pair-wise, their calculation takes most of the computational time needed to calculate the potential energy according to equation (1).

The force field parameters are the set of atom type-specific parameters in each term in equation (1). Due to their use in the calculation of the potential energies, which yields the forces acting on the atoms, the atom type-specific parameters influence the behavior of the system in the MD simulation. The determination of these parameters, i.e., the parameterization of the force field, is a highly complex task, since terms influence each other, and a single change in one parameter might have a huge impact on the behavior of the biomolecule in the simulation [178, 180].

2.9.1.2 Molecular dynamics force fields for RNA simulations

In this section, I will describe the most recent Amber force field modifications for RNA simulations. The following description is adapted from Publication 1 (Section 4) [136].

For MD simulations of nucleic acids, in particular RNA, the ff99 force field [181] was the force field of choice for several years in the Amber program package [182]. The ff99 force field [181], as well as other current nucleic acid force fields, performs well in describing canonical DNA and RNA helices, while they might show different success for more complex RNA molecules [180, 183, 184].



Dihedral	Atoms
α (i)	O3'(i-1) - P(i) - O5'(i) - C5'(i)
β (i)	P(i) - O5'(i) - C5'(i) - C4'(i)
γ (i)	O5'(i) - C5'(i) - C4'(i) - C3'(i)
δ (i)	C5'(i) - C4'(i) - C3'(i) - O3'(i)
ε (i)	C4'(i) - C3'(i) - O3'(i) - P(i+1)
ζ (i)	C3'(i) - O3'(i) - P(i+1) - O5'(i+1)
χ (i)	for purine base: O4'(i) - C1'(i) - N9(i) - C4(i) for pyrimidine base: O4'(i) - C1'(i) - N1(i) - C2(i)

Figure 13: Dihedral angles in a nucleotide (left). Right table: Atoms involved in dihedral angles in the nucleotide. The atom names and dihedral definitions are according to IUPAC [185].

In recent years, modifications of the force field parameters of the ff99 force field were published in order to tackle structural issues occurring in nucleic acid MD simulations during long simulation times. One such refinement is the parmbsc0 modification (also referred to as bsc0) for the ff99 force field [186]. In bsc0, a reparameterization of the α/γ dihedrals in the nucleic acid backbone (Figure 13) was introduced in order to prevent the accumulation of irreversible non-native γ-*trans* dihedral states, which resulted in the collapse of B-DNA. Initially, it was not clear whether bsc0 offers advantages for RNA simulations [187], as such degradation has never substantially affected canonical A-RNA simulations [188, 189]. With the increase in computational power, and consequently extending simulation times, a sudden irreversible transition of RNA helices to ladder-like structures (Figure 14) was observed in some simulations [190]. In order to avoid the formation of such non-physiological structures, a reparameterization of the χ-dihedral (Figure 13) was developed. This reparameterization was designed in order to suppress *anti* to *high-anti* shifts in RNA [183] and was published as parmχ_{OL3} [191]. Even though there exist alternative modifications of χ-dihedral parameters [192, 193] when using the ff99 force field, the currently recommended combination is to use the ff99 force field together with the bsc0 and the parmχ_{OL3} refinements [194]. In the context of RNA simulations in Amber, this combination is also referred to as ff10, ff12 (in Amber12), and ff14 (in Amber14).

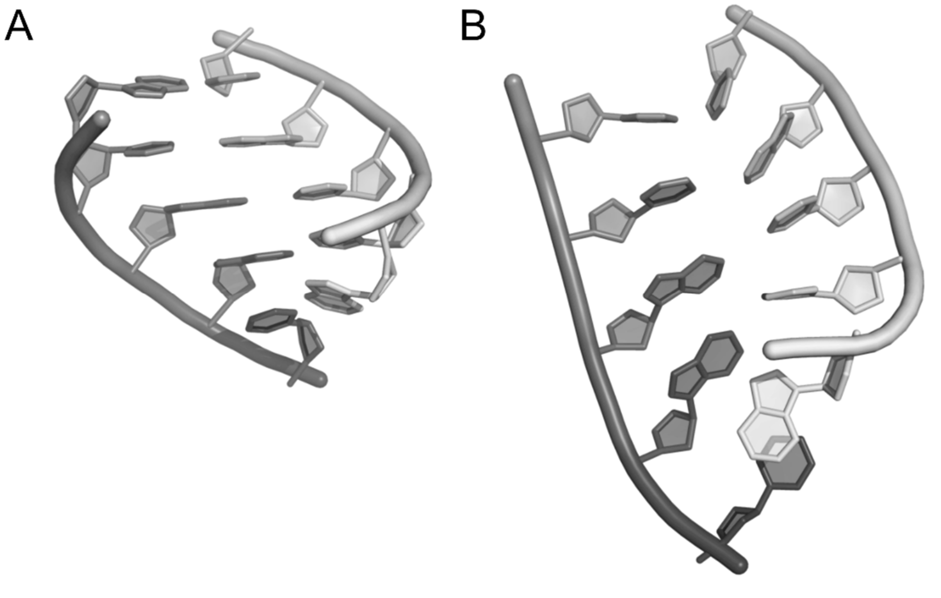


Figure 14: Structural comparison between a RNA helix (A) and a "ladder-like" structure, which lost most of its helical structure and is caused by the force field (B).

2.9.1.3 Mg^{2+} in MD simulations of RNA

Another point of interest in nucleic acid MD simulations are positively charged ions. These ions serve the purpose of charge compensation, which is necessary due to the strong negative charge of the nucleic acid backbone. In biological systems and many experiments involving nucleic acids, Mg^{2+} ions are one of the most important divalent ions used [2, 195, 196]. This is due to their importance in stabilizing nucleic acid structures [197, 198].

Here, I introduce the challenges of modeling Mg^{2+} ions in MD simulations. The following section is adapted from Publication 1 (Section 4) [136]. The treatment of Mg^{2+} ions in MD simulations is challenging due to their slow exchange kinetics of their first shell ligands on the order of µs [199], their low diffusion coefficient [200], and difficulties in modeling polarization and charge-transfer effects in classical fixed-charge additive force fields [183, 201]. Due to these difficulties, the initial positioning of Mg^{2+} ions in the MD simulation is important. If experimental information on Mg^{2+} binding sites is available, the placement of these ions is comparably simple. However, if there is no information about inner-sphere coordinated Mg^{2+} ions, they should be assumed freely moving. If a Mg^{2+} ion is placed too close to the negatively charged RNA backbone, this ion will most likely form an inner-sphere contact and will presumably not move away from there over the course of a MD simulation.

Despite the existence of more sophisticated models [202], the nonbonded model description for Mg^{2+}, i.e., representing their interactions using Coulombic and Lennard-Jones terms, is still widely used. This is also due to the simplicity of the nonbonded model, since only Lennard-Jones parameters have to be determined. More than 20 years ago Åqvist developed Lennard-Jones parameters for Mg^{2+} ions [203], which were then adopted for the use in Amber force fields. These parameters result in an ion-water exchange rate of the first solvation shell, which is two orders of magnitude slower than experimental values [204]. Therefore, Allner *et al.* developed other Mg^{2+} parameters to reproduce this ion-water exchange rate [204]. More recently, Li *et al.* designed Lennard-Jones parameters for Mg^{2+} ions, which are now recommended to be used in Amber [202, 205, 206]. These parameters were designed to reproduce the hydration free energy and the ion-oxygen distance in the first hydration shell. However, these hydration free energies for Mg^{2+} deviate more from the experimental value than those computed with the Amber force field-adopted Åqvist parameters.

2.9.2 Replica exchange molecular dynamics

2.9.2.1 Idea of replica exchange molecular dynamics

MD simulations are a powerful tool to gain insights into the dynamics of a biomacromolecule on the atomic level. However, the sampling capability of standard MD simulations is limited due to its problems to overcome local maxima at typical simulation temperatures around 300 K. Higher temperature simulations are able to access new areas of the conformational space not available to room-temperature MD simulations, but the high temperature makes sampling of local energy minima difficult (Figure 15). In order to overcome these sampling issues of standard MD simulations at one temperature, replica exchange molecular dynamics (REMD) simulations can be used [207-209].

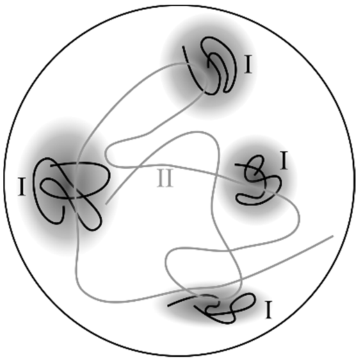


Figure 15: Sampling of the free energy landscape by different simulations and principle of a REMD simulation. A: Local sampling (I, black) of simulations that are trapped in local minima (grey shaded areas). A higher temperature simulation (II, grey) has the ability to overcome local maxima but has difficulties to sample local minima. Figure based on Ref. [209].

REMD, also called parallel tempering, is an enhanced sampling method, in which multiple copies of the simulation system are simulated in parallel at different temperatures [209]. At regular intervals, an exchange between two neighboring replicas is attempted. An exchange between two replicas i and $j = i + 1$ is accepted with probability

$$P = \min\{1,\ \exp[(\beta_i - \beta_j)(U(r_i^N) - U(r_j^N))]\}$$

where $\beta_i = 1/(k_B\,T_i)$ is the reciprocal temperature with k_B the Boltzmann constant and T_i the temperature of replica i. $U(r_i^N)$ is the potential energy based on the coordinates r_i for N atoms. Thus, the probability for an exchange between two replicas depends on the temperatures of the two replica and the potential energies of the corresponding systems.

The exchange between the temperature replicas allows a system to go to a higher temperature in order to overcome local barriers and afterwards go to lower temperatures again for improved local sampling (Figure 16).

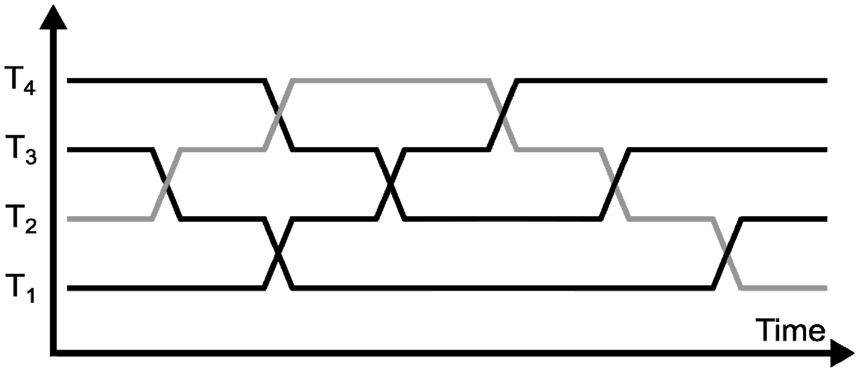


Figure 16: Principle of a REMD simulation. Four replica at different temperatures (T_1, T_2, T_3, T_4). At regular intervals, exchanges between two neighboring replica are attempted. One replica is shown in grey for visibility.

If an exchange attempt is successful, the momenta for the replica are scaled according to

$$p^{(i)\prime} = \sqrt{\frac{T_{new}}{T_{old}}} p^{(i)}$$

where $p^{(i)\prime}$ are the new momenta after the exchange, $p^{(i)}$ are the momenta before the exchange, T_{old} is the temperature of the system before the exchange, and T_{new} is the new temperature after the exchange [209].

In order to allow for exchanges between neighboring replica, the temperatures of the replica have to be chosen carefully [210-213]. The temperatures should be chosen, such that the potential energy distributions of the replica at the different temperatures overlap. Typically an exchange rate between 20% and 30% is desired [209, 211]. The frequency of exchange attempts is another often discussed parameter [214, 215].

REMD simulations were successfully applied to biomolecules, which typically possess a rugged energy landscape [207, 209]. Applications include the investigation of folding pathways [208, 216-218], increased conformational sampling [13, 219, 220], and the determination of free energy landscapes [221] for proteins and nucleic acid structures.

2.9.2.2 Analysis of REMD simulations

REMD simulations produce large amounts of data due to the parallel simulation of multiple copies of the system. The results of a REMD simulation can be analyzed by following one replica through different temperatures, or by extracting snapshots from the systems at a certain temperature. By analyzing the snapshots at a target temperatures (e.g. 300 K), two subsequent snapshots are not necessarily connected in time. This is due to the exchanges between the temperatures, where a system can go to a higher temperature and then come back to lower temperatures. However, one obtains a thermodynamic ensemble of structures at the target temperature. Yet, using only the structures from one temperature means to omit all the data collected for the other temperatures. In order to make use of these data, it is possible to reweight the results from all temperatures to obtain an unbiased result [221, 222]. Such a reweighting can be done using the weighted histogram analysis method (WHAM) [223, 224] or the multistate bennet acceptance ratio (MBAR) method [225, 226]. These methods can also be applied in a similar way to reweight the results of Umbrella sampling simulations and estimate an unbiased result [226].

Using these methods, it is possible to generate a potential of mean force (PMF) for various observables, such as the radius of gyration or the number of native contacts. Furthermore, a two-dimensional PMF can be generated, which allows to identify energetically favorable regions with respect to two chosen observables [227-230].

2.9.3 Rigidity theory

In order to investigate the flexibility and rigidity of biomolecules, graph-theory based approaches proved efficient and successful [231-240]. Here, the biomolecule is converted to a network, where the vertices represent the molecule's atoms and the edges represent interatomic interactions, i.e., covalent bonds, hydrogen bonds, or hydrophobic interactions. From this network representation, flexible and rigid regions of the molecule can then be determined. It is important to note that rigidity and flexibility describe the statics of a molecule, i.e., the possibility to move without actually moving parts of the molecule [237].

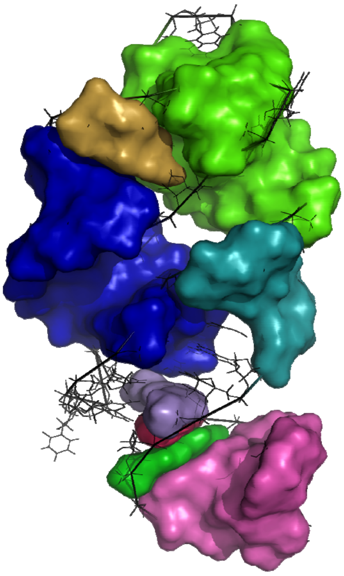


Figure 17: Rigid cluster decomposition of the Gsw aptamer domain. The differently colored objects represent different rigid clusters. The largest rigid cluster is colored blue. Nucleotides shown in black belong to flexible regions connecting the rigid clusters.

The *Floppy Inclusions and Rigid Substructure Topography* (FIRST) software [232] performs such a decomposition of a given molecule. After the conversion of the input structure to a network representation, the 3D-pebble game algorithm [241, 242] is used to perform the decomposition of the network into rigid parts and flexible connections. FIRST has been successfully applied to the investigation and prediction of flexibility and rigidity characteristics of proteins and RNA molecules [231-239]. For RNA molecules, a RNA-specific parameterization was developed, to account for the different structural properties of RNA and proteins [237, 243]. As it was found that rigidity analyses are sensitive to the input structure, it is recommended to perform rigidity analyses on ensembles of structures [231].

The results of a FIRST analysis can, for example, be evaluated by looking at the rigid cluster decomposition of the biomolecule (Figure 17). Here it is possible to trace the membership of an atom in the largest rigid cluster. When analyzing an ensemble of structures with FIRST, a probability for an atom i to belong to the largest rigid cluster can be calculated as

$$p_{lrc}(i) = \frac{n_1(i)}{N} \tag{2}$$

where $n_1(i)$ is the number of occurrences of atom i as part of the largest rigid cluster, calculated over N structures [231].

3 Scope of the thesis

In order to gain insights into the regulatory decision by riboswitches, a detailed understanding of the first step of this process, the unbound state, is crucial. However, due to the dynamic nature of riboswitches, the unbound state has remained largely elusive. In my thesis, I investigated the conformational heterogeneity in the unbound state of the guanine-sensing riboswitch aptamer domain using molecular simulations. For this, I used a wildtype guanine-sensing riboswitch aptamer ($\mathrm{Gsw}^{\mathrm{apt}}$) and a mutant ($\mathrm{Gsw}^{\mathrm{loop}}$), which exhibits a Mg^{2+} dependent destabilization in experiments.

The first step for my work was to decide for a system setup that can reproduce experimental findings on the riboswitch, in particular on differences between wildtype and mutant (Publication I, Section 4). Therefore, I compared three force field variants available in the Amber software with respect to their agreement with experimental data. Furthermore, I developed an improved placement procedure for Mg^{2+} ions in nucleic acid simulations to prevent an initial bias due to their placement, in particular, if no information about ion binding sites is available.

As a next step, I used this system setup to compare the behavior of $\mathrm{Gsw}^{\mathrm{apt}}$ and $\mathrm{Gsw}^{\mathrm{loop}}$ in the unbound state (Publication II, Section 5). In this way, I investigated long-range effects from the tertiary interactions to the ~30 Å distant ligand binding site. Furthermore, I applied a rigidity theory-based approach to study the connection between the stability of the tertiary interaction, ligand binding, and the terminal P1 region.

As a next step towards the understanding of the unbound state, I performed REMD simulations, which allow for enhanced sampling of the conformational space compared to standard MD simulations (Publication III, Section 6). From this, I calculated free energy landscapes, which allowed me to gain insights into the influence of the tertiary interactions on the conformational heterogeneity of the unbound aptamer.

4 Publication I - Force field dependence of riboswitch dynamics

Hanke C.A., Gohlke, H.

in Computational methods for understanding riboswitches, S.-J. Chen and D.H. Burke-Aguero, Editors. 2015, Academic Press: Burlington. p. 163-191.

This section is based on the above-mentioned publication. Manuscript see pages 70-99

4.1 *Background*

Molecular dynamics (MD) simulations are a very useful method to gain insights about the dynamics of a biomolecular system. However, the choice of the force field and other simulation parameters strongly influences the outcome of a MD simulation [174-177]. Similarly, positively charged ions, especially Mg^{2+} ions, play an important stabilizing role for RNA structures due to their ability to compensate the negative charges of the nucleic acid backbone [2, 195-197]. However, the treatment of Mg^{2+} ions in MD simulations poses several difficulties as explained in Section 2.9.1.3. For this study, I used the two guanine-sensing riboswitch aptamer variants Gsw^{apt} and Gsw^{loop}, representing the wildtype and the G37A/C61U mutant (Section 2.7.5). The G37A/C61U mutation in Gsw^{loop} was developed to destabilize the tertiary loop-loop interactions and in this way the entire aptamer [146, 151, 152].

The aim of the study was to find a set of force field and Mg^{2+} ion parameters that provide results in very good agreement with experimental observations. Therefore, I thoroughly investigated the influence of initial simulation conditions and the choice of parameters on the dynamics of the unbound guanine-sensing riboswitch aptamer domain.

To this end, I tested three recent RNA force field variants available in the Amber package [182], which show only minor differences in their parameterization: ff99, bsc0, and ff10 (Section 2.9.1.2). Furthermore, I developed an improved placement procedure for Mg^{2+} ions for RNA simulations, since the default placement procedure implemented in the *leap* program [244] (available in the Amber software package [182]) can result in an initial bias due to close contacts between Mg^{2+} ions and the RNA. In summary, for this study, I performed 58 MD simulations to test the influence of force field parameters, Mg^{2+} ion parameters, and Mg^{2+} ion placement on the behavior of Gsw^{apt} and Gsw^{loop}. The total aggregate simulation time for this study amounts to more than 11 µs, which allows for a thorough comparison of different parameters.

4.2 *Results*

An improved placement procedure for Mg^{2+} results in a better description of the ion mobility

Despite their importance for charge compensation in biological systems, Mg^{2+} ions are not widely used in MD simulations [201, 245]. This is due to the difficulty in modeling these ions

(Section 2.9.1.3), which makes their initial placement highly important for nucleic acid MD simulations. Using the default placement procedure in the *leap* program [244], which is typically used to set up simulation systems within the Amber package [182], Mg^{2+} ions are often placed very closely to the negatively charged nucleic acid backbone. The reason for this is that the program calculates an electrostatic potential and subsequently places the ions at the electrostatically most favorable positions. However, I observed that this can strongly impair the dynamics of the RNA in the simulations due to quick formation of inner-sphere complexes. In particular, when placed between two RNA backbones, the Mg^{2+} ion might bridge the gap between the phosphorus atoms. Such a bridging interaction severely disturbs the local dynamics of the RNA molecule. In order to avoid such an initial bias, I developed an alternative placement procedure for Mg^{2+} ions in my simulations. The placement procedure is not restricted to riboswitch simulations but can be also applied to the setup of other nucleic acid simulations. Here, I initially placed dummy atoms with the same charge as a Mg^{2+} ion, but with a larger radius of 4 Å. This larger radius is similar to the radius of a hexahydrated Mg^{2+} ion [246]. The dummy atom was then replaced by a Mg^{2+} ion with a first hydration shell of six water molecules. The initial hydration shell around the Mg^{2+} ion was maintained during the thermalization phase by the use of positional restraints.

As shown in Figure 18A, the movement of the Mg^{2+} ions when using the default placement procedure is strongly restricted due to stable direct interactions with the RNA. Larger movements observed in this case are due to the movement of the RNA. In contrast, in Figure 18B, the Mg^{2+} ions can initially move freely. Over the course of the MD simulation, several ions lost part of their hydration shell and directly interact with the RNA. After 200 ns, at least 60% of the Mg^{2+} ions were still hexahydrated. In this way, the ions can initially equilibrate and find preferred locations independent from their starting positions. This reduces the likelihood to introduce errors due to suboptimal initial ion placement.

In comparison to the obvious differences caused by the different placement procedures of the Mg^{2+} ions, I did not observe a large influence of different ion parameters on the RNA structural dynamics. I compared two sets of Mg^{2+} parameters available with Amber force fields: the long-used parameter set developed by Åqvist [203] and a more recent set of parameters by Allner *et al.* [204]. The two parameter sets result in different amounts of inner-sphere contacts of the ions with RNA, where the parameter set by Allner *et al.* does not show any inner-sphere contacts. Even though the number of inner-sphere contacts for the Åqvist parameter set might be too high, the lack of any inner-sphere contacts in the case of the parameters by Allner *et al.* was surprising, since such contacts do occur in crystal structures of related purine-sensing riboswitches [52, 247]. From the data, we could not finally determine which parameter set is superior. Therefore, we decided to use the Åqvist parameters for further simulations due to the longer experience with these parameters.

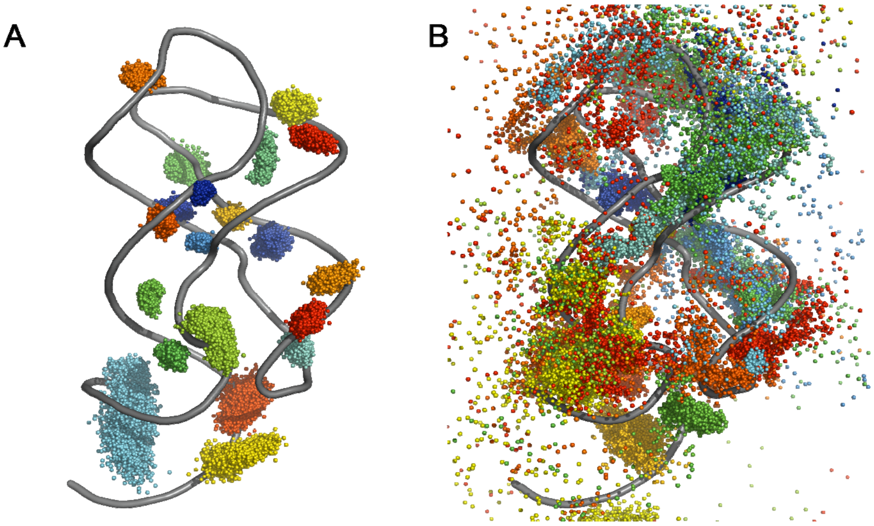


Figure 18: Positions of 20 Mg^{2+} ions during 100 ns of MD. Each color corresponds to one Mg^{2+} ion. The RNA was overlaid. A: The Mg^{2+} ions were placed using the default placement procedure. B: The Mg^{2+} ions were placed using the improved placement procedure developed in Publication I. This Figure was taken from [136].

The choice of the force field parameters affects the behavior of the RNA

Due to the variety of force field variants available, the question arises which of them to use. Here, I tested three recent force field variants available in the Amber program package [182] with respect to how simulation results obtained with them agree with experimental data: ff99, bsc0, and ff10 (for details, see Section 2.9.1.2).

The ff10 force field modification was developed in order to avoid the formation of ladder-like conformations in RNA structures. I observed the formation of ladder-like structures only in one out of 18 simulations (after ~200 ns) for the ff99 force field, and in two out of 18 simulations (after ~100 ns) for the bsc0 force field. Here, the ladder-like structures were formed at the end of the terminal P1 region. Such structures can however be identified easily, and the corresponding trajectories were excluded from subsequent analyses. As expected, I did not observe the formation of ladder-like structures when using ff10.

For the influence of the force field parameters on the RNA, I calculated atomic fluctuations in terms of root-mean-square fluctuation (RMSF) values for the RNA. Figure 19 shows the mean RMSF for different force field variants and Mg^{2+} concentrations (0, 12, and 20 Mg^{2+} ions per RNA molecule). Interestingly, I found statistically significant larger mean RMSF values for Gsw^{loop} than for Gsw^{apt} in the absence of Mg^{2+} ions for both ff99 and bsc0 force fields. In the case of ff99, I also observed a significantly larger mean RMSF value for Gsw^{loop} than for Gsw^{apt} in the presence of Mg^{2+} ions. A less stable Gsw^{loop} aptamer in comparison to Gsw^{apt} is in agreement with experimental findings [151]. Furthermore, I observed significant differences between the absence and the presence of Mg^{2+} ions for Gsw^{apt} and Gsw^{loop} in the case of ff99 and bsc0. Since Mg^{2+} ions are known to stabilize nucleic acid structures [196, 198, 248], such a decrease in atomic fluctuations is not surprising. For ff10, however, I did not observe significant differences in atomic fluctuations for either the comparison between Gsw^{apt} and Gsw^{loop} or the comparison between the absence and the presence of Mg^{2+} ions. For all systems using ff10, the RMSF values are comparably low. Thus, ff10 results in a largely rigid RNA, irrespective of the presence of the destabilizing mutation or the absence of Mg^{2+} ions.

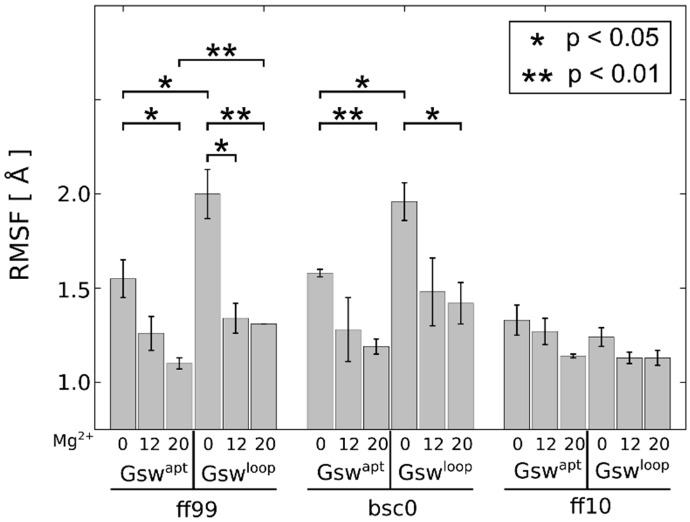


Figure 19: Mean RMSF±SEM of the 80% least fluctuating nucleotides of the RNA. The mean RMSF values are shown for different numbers of Mg^{2+} ions and Gsw variants, grouped by the used force field. The stars indicate statistically significant differences between the mean values (*: $p < 0.05$; **: $p < 0.01$). Figure taken from [136].

Since the G37A/C61U mutation was designed to disturb the native hydrogen bond network between the L2 and L3 loops, I compared the occupancy of hydrogen bonds within the two base quadruples that form the loop-loop interactions (Figure 9). All three force fields show a stable hydrogen bond network in the case of Gsw^{apt}, with ff10 showing the highest occupancy values. For ff99 and bsc0, I observed the largest decreases (up to 60% for ff99 and up to 80% for bsc0) for the occupancies when comparing Gsw^{loop} to Gsw^{apt}, while ff10 still shows values above 55% occupancy for all hydrogen bonds for Gsw^{loop}. Only in the case of ff99 and bsc0, the destabilization of the hydrogen bond network introduced by the G37A/C61U mutation in the upper base quadruple extends to the lower base quadruple, while for ff10 a similar stability is seen in Gsw^{loop} as in Gsw^{apt}.

The results for ff99 and bsc0 show a reduced stability of the tertiary interaction in Gsw^{loop}. This is in line with the disturbance of the native hydrogen bond network in the loop region of Gsw^{loop} caused by the G37A/C61U mutation. However, our results are at variance with findings from

NMR experiments that the interactions between the L2 and L3 loops are not formed in the absence of Mg^{2+} ions [151, 152]. It is important to note here, that we simulated the unbound aptamer domain based on the ligand-bound crystal structure after removing the ligand. In this structure, the stable hydrogen bond network is already formed. In addition, our simulation times of 200 ns are still short with respect to movements of an RNA on a rugged energy landscape [183] and thus may be insufficient to completely break the initial hydrogen bonds. Furthermore, using ff99 I was able to observe an increased stabilization of the hydrogen bond network for Gsw^{apt} in the presence of Mg^{2+}, manifested in higher hydrogen bond occupancies. Such an observation is in agreement with NMR experiments of Gsw^{apt} [151, 152].

From the similar behavior of ff99 and bsc0 and from comparison of α, γ, and χ dihedral distributions, I concluded that the rigidification observed in ff10 is due to the change of the χ torsion angle parameter. While this modification is successful in suppressing the formation of ladder-like structures, it has a detrimental effect on the mobility of the marginally stable Gsw.

4.3 Conclusions and significance

In this study, I showed that the initial setup and the choice of force field parameters significantly affect the dynamics of RNA in the simulations. Furthermore, I developed an improved procedure for the placement of Mg^{2+} in nucleic acid simulations, which prevents a potential initial bias due to the placement of the ions around the nucleic acids. The main conclusions from this work are:

- The initial positioning of ions, especially of Mg^{2+} ions, is important during the system setup, in order to avoid impairment of the RNA dynamics.
- The MD simulations (using ff99 and bsc0) are sensitive enough to reproduce experimental differences between Gsw^{apt} and Gsw^{loop} and between different Mg^{2+} concentrations. This enables subsequent, more detailed investigations of the behavior of the aptamers.
- The ff99 force field reproduces experimental differences between Gsw^{apt} and Gsw^{loop} and different Mg^{2+} concentrations best, followed closely by bsc0.
- The ff10 force field results in a more rigid RNA, which does not reflect the marginally stable nature of the guanine-sensing riboswitch aptamer domain.

Based on the findings from this study, I used the ff99 force field together with the improved placement procedure for Mg^{2+} ions for my further investigations of Gsw.

5 Publication II - Ligand-mediated and tertiary interactions cooperatively stabilize the P1 region in the guanine-sensing riboswitch

Hanke C.A., Gohlke, H.

PLOS ONE, 2017. **12**(6): p. e0179271.

This section is based on the above-mentioned publication. Manuscript see pages 100-142.

Parts of this work have been published in the Proceedings of the NIC Symposium 2014 [249].

5.1 Background

In the previous study (Publication I, Section 4) [136], I investigated the influence of parameter choices for the MD force field and Mg^{2+} ions on the behavior of the guanine-sensing riboswitch aptamer. From this, I chose a setup that can reproduce experimental findings on Gsw^{apt} and Gsw^{loop}.

Here, I will present findings on long-range effects in the Gsw, in particular the effects of destabilizing the tertiary interactions in the loop region of Gsw and possible stabilizing effects by the presence of a ligand. To this end, I performed MD simulations of the unbound state of Gsw^{apt} and Gsw^{loop} at three different Mg^{2+} concentrations (0, 12, and 20 Mg^{2+} ions per RNA molecule). The simulation setup for this study was similar to the previously presented study, but the simulations were extended with respect to simulation time and analyses. For this study, I performed three independent simulations for each of the six systems with a simulation time of 550 ns each, summing up to a total simulation time of almost 10 µs.

5.2 Results

The Mg^{2+} ions preferably occupy experimentally determined Mg^{2+} binding site

For the setup of the MD simulations, I used the improved placement procedure introduced in Publication I (Section 4) [136]. This allowed for initial mobility of the ions as previously described (Figure 18) [136], and enabled occasional exchanges between hexahydrated mobile Mg^{2+} ions (Figure 20A). I observed that the sites preferably occupied by the ions are in agreement with experimentally determined ion binding sites from crystal structures [104, 152] (red and orange spheres in Figure 20B). Furthermore, these sites are close to nucleotides showing chemical shift changes upon Mg^{2+} titration in NMR experiments [152] (blue nucleotides in Figure 20B).

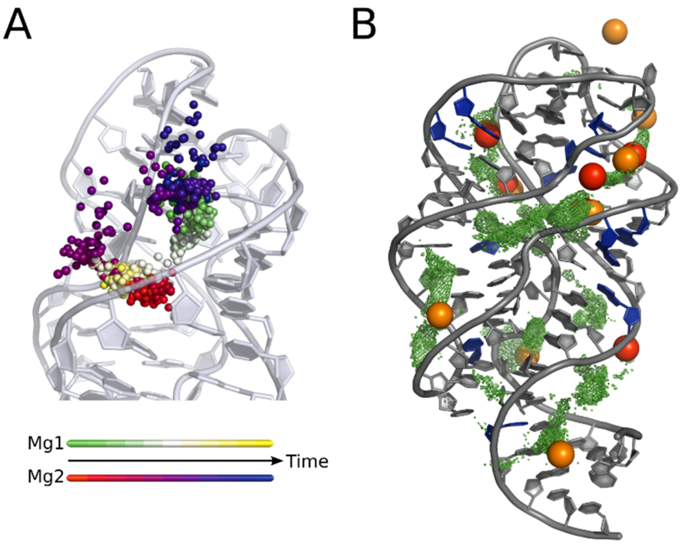


Figure 20: Exchange of Mg^{2+} ions and preferred sites of occupancy. A: Exchange of positions of two Mg^{2+} ions over a simulation time of 8 ns. The position of the ions are color-coded by time using two different color scales. The RNA is shown in light grey. B: Preferred sites of occupancy of Mg^{2+} around the Gsw (grey) during 300 ns of MD simulations (green) together with experimentally determined ion binding sites (red/orange: binding of cobalt hexammine ions in X-ray structures with PDB ID 4FE5 [104]/3RKF [152]) and nucleotides showing chemical shift changes upon Mg^{2+} titration in NMR experiments [152]. Figure taken from [250].

Influence of the G37A/C61U mutant on the aptamer

Similarly to the observations from Publication I (Section 4) [136], I observed a destabilization of the tertiary interactions between the loops of the G37A/C61U mutant Gsw^{loop} in the simulations of this study. This destabilization can be seen in terms of lower hydrogen bond occupancies. Interestingly, the destabilization by the mutation located in the upper base quadruple even extends to the lower base quadruple. Furthermore, I observed a stabilization of the tertiary interaction for Gsw^{loop} in the presence of Mg^{2+} ions. These findings are in agreement with experimental findings that the tertiary interactions in Gsw^{loop} are not stably formed in the absence of Mg^{2+} ions, but can form above a [Mg]:[RNA] ratio of 18:1 [151, 152]. In contrast to experiments, I did not observe a complete disruption of the tertiary interactions in the absence of Mg^{2+} ions. This is likely because the simulation starts from a structure with formed tertiary interactions, and the simulation length does not allow the complete breaking of these.

In my simulations, I did not observe large conformational changes or structural differences between Gsw^{apt} and Gsw^{loop} on a global level, measured in terms of RMSD and radius of gyration. Thus, it is likely that the influence of the tertiary interactions on ligand binding and the effect of Mg^{2+} ions can be found on the level of dynamics or small local rearrangements.

The G37A/C61U mutation increases, whereas Mg^{2+} ions decrease the structural dynamics of the aptamer

In order to investigate the influence of the tertiary interaction stability on the dynamics of the aptamer, I calculated RMSF values. In general, Gsw^{loop} shows higher mobility than Gsw^{apt}, especially in the absence of Mg^{2+} ions. For both aptamer variants, I observed a decrease of the structural dynamics of the RNA in the presence of Mg^{2+} ions, which is in line with the stabilizing role of Mg^{2+} ions in RNA structures [197]. These data suggest that the G37A/C61U mutant in Gsw^{loop} has a long-range effect in terms of increasing the structural dynamics of the entire aptamer, which can be compensated by the presence of Mg^{2+} ions. When comparing Gsw^{apt} and Gsw^{loop} in the absence of Mg^{2+} ions, I observed higher mobility in the junction region of Gsw^{loop}, especially around J3/1 (nucleotides 73, 74, and 75) (Figure 21A). Interestingly, this region, together with the loop regions, shows decreased structural dynamics in the presence of Mg^{2+} ions in Gsw^{loop} compared to the absence of Mg^{2+} ions (Figure 21B).

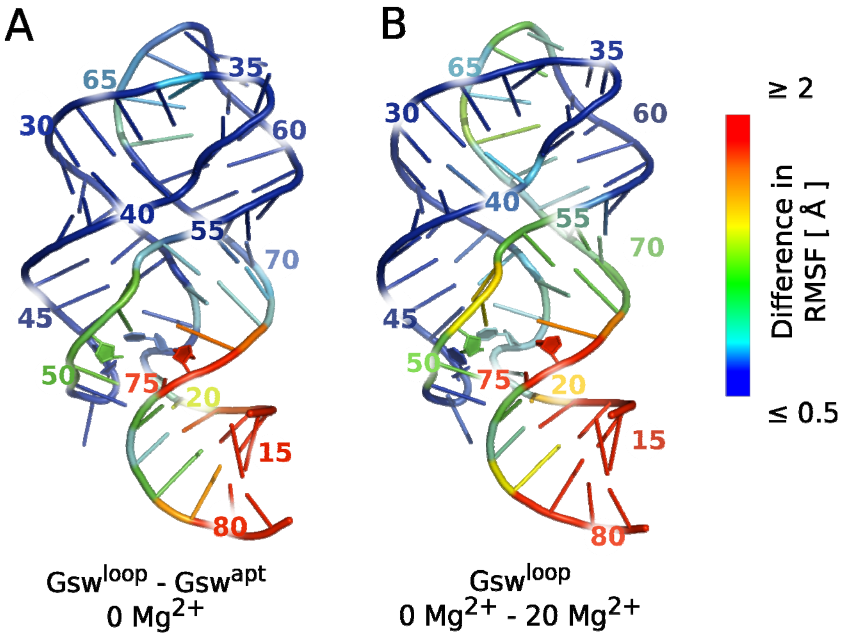


Figure 21: Per-nucleotide differences in atomic fluctuations mapped on the aptamer. A: Difference in RMSF between Gsw^{apt} and Gsw^{loop} in the absence of Mg^{2+} ions. B: Difference in RMSF between the presence and the absence of Mg^{2+} ions for Gsw^{loop}. Figure adapted from [250].

The J3/1 region contains nucleotide 74, which was suggested to be involved in the initial binding of the ligand [142, 143]. Thus, these findings could explain why Gsw^{loop} is not able to bind hypoxanthine in the absence of Mg^{2+} ions, while the ligand binding ability is restored in their presence [151, 152]. In the absence of Mg^{2+} ions, nucleotide 74 might exhibit too high structural dynamics for productive binding of a ligand. This would be in line with experimental findings that the formation of the initial encounter complex requires a restriction of the conformational dynamics and a partially organized ligand binding core [151]. Such a conformational restriction together with the partial organization of the ligand binding core might be induced through the stabilization of remote regions [151], which is the case for the loop-loop interactions in Gsw^{loop} in the presence of 20 Mg^{2+} ions. In summary, the destabilization of the tertiary interactions has a long-range effect, increasing the structural dynamics of the ~30 Å distant ligand binding site. This increase of the structural dynamics can be compensated by the presence of Mg^{2+} ions.

The presence of a ligand rigidifies the binding pocket and its surrounding

In order to investigate the influence of a ligand in the binding pocket without the need for additional simulations, I applied a graph theory-based approach using the FIRST software [232]. The FIRST software decomposes the constraint network representation of an input structure into rigid and flexible regions (Section 2.9.3). For the rigidity analyses, I introduced additional constraints to the network in order to mimic the presence of a ligand in the binding pocket. The constraints were chosen to represent the interactions formed by the nucleotides when binding the ligand. I calculated the probability to be in the largest rigid cluster $p_{lrc}(i)$ (equation (2), Section 2.9.3) for both, the absence ($p_{lrc}(i)_{apo}$) and the presence of the ligand ($p_{lrc}(i)_{lig}$) in the structural ensemble from the MD simulation. From the difference of the p_{lrc} values ($\Delta p_{lrc}(i) = p_{lrc}(i)_{lig} - p_{lrc}(i)_{apo}$), it is possible to identify the rigidification caused by the presence of the ligand constraints (Figure 22). In the simulations, I observed that the ligand has a stabilizing influence on the ligand binding site, which extends to the neighboring nucleotides. A stabilizing influence on nucleotides of the P1 region is in line with experimental data, suggesting that the ligand stabilizes the ligand binding pocket and thereby stabilizes the P1 region via formation of base triples and stacking interactions [14]. The rigidifying effect of the ligand on the P1 region is stronger in case of Gsw^{apt}, where the loop-loop interactions are more stable than in Gsw^{loop}. Furthermore, the stabilizing effect becomes stronger in Gsw^{loop} when Mg^{2+} is present, which stabilizes the tertiary interactions as described before. This suggests that the influence of the bound ligand on the P1 region might be dependent on the stability of the tertiary loop-loop interactions.

Stable tertiary interactions and a stable ligand binding site cooperatively stabilize P1

As a next step, I investigated whether the stable tertiary interactions of Gsw^{apt} together with the stabilization of the ligand binding site have a cooperative effect on the stability of the terminal P1 region. For this, I calculated $Coop(i)$, which describes the cooperativity between the stabilization of the tertiary interactions and the stabilization of the ligand binding pocket:

$$Coop(i) = \underbrace{\ln\left(\frac{p_{lrc}(i)^{Gsw^{apt}}_{lig}}{p_{lrc}(i)^{Gsw^{loop}}_{apo}}\right)}_{A} - \left[\underbrace{ln\left(\frac{p_{lrc}(i)^{Gsw^{apt}}_{apo}}{p_{lrc}(i)^{Gsw^{loop}}_{apo}}\right)}_{B} + \underbrace{ln\left(\frac{p_{lrc}(i)^{Gsw^{loop}}_{lig}}{p_{lrc}(i)^{Gsw^{loop}}_{apo}}\right)}_{C}\right] \tag{3}$$

Here, Gsw^{loop} in the absence of the ligand constraints was used as a ground state, since in this case neither the tertiary interactions nor the ligand binding pocket is stabilized. Gsw^{apt} corresponds to the case where the tertiary interactions are stabilized, and the presence of a ligand results in a stabilization of the ligand binding pocket. Accordingly, term A describes the influence of the tertiary interactions together with the ligand, term B describes the influence of the tertiary interactions alone, and term C describes the influence of the ligand alone. A positive $Coop(i)$ value means that the simultaneous stabilization of the tertiary interactions and the ligand binding site (through the ligand constraints) do not simply add up, but result in a larger stabilizing effect.

In the presence of Mg^{2+} ions, only $Coop(i)$ values for nucleotides in the terminal P1 region were significantly different from zero. Here, the $Coop(i)$ values are positive, which shows that the effects of the stabilization of the tertiary interactions and the stabilization of the ligand binding pocket have a synergistic stabilizing effect. Thus, the increase in rigidity of nucleotides in the terminal P1 region is larger if both sites are stabilized at the same time.

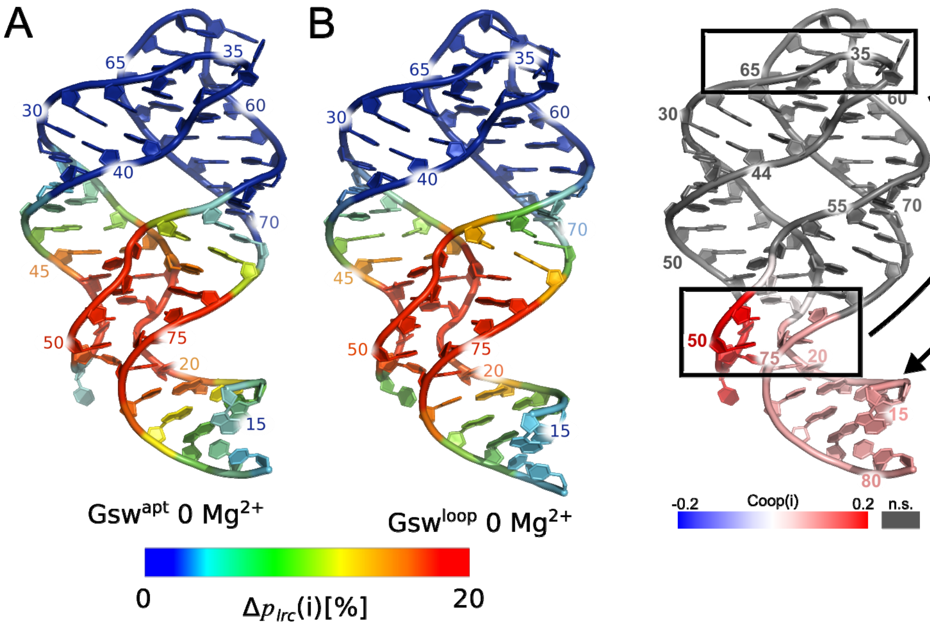


Figure 22: Results from rigidity analyses and cooperativity in the guanine-sensing riboswitch aptamer. A,B: Difference in rigidity between the presence and the absence of the ligand constraints (calculated by equation (2)) for Gsw^{apt} (A) and Gsw^{loop} (B) in the absence of Mg^{2+} ions. The rigidifying effect towards the terminal P1 region is larger for Gsw^{apt} than for Gsw^{loop}. C: $\boldsymbol{Coop(i)}$ values calculated according to equation (3) for the systems in the presence of Mg^{2+} ions. The tertiary loop-loop interactions (indicated by the upper rectangle) together with the ligand binding site (lower rectangle) cooperatively influence the stability of the P1 region. Grey nucleotides show $\boldsymbol{Coop(i)}$ values that are not significantly different from zero ($\boldsymbol{p} > 0.05$). Figure adapted from [250].

5.3 Conclusions and significance

In this study, I investigated the long-range effects of the tertiary interactions and a stabilized ligand binding site on the aptamer domain. The results of this study revealed for the first time, that there is a structurally stabilizing interplay not only between pairs, but between all three of the investigated sites: the tertiary interactions in the loop region, the ligand binding site, and the terminal P1 region. The main results from this work are:

- The G37A/C61U mutation in $\mathrm{Gsw}^{\mathrm{loop}}$ results in an increase of the structural dynamics of the aptamer and in particular of the ~30 Å distant ligand binding site. The increase in structural dynamics can be compensated by the presence of Mg^{2+} ions.
- The increase structural dynamics of the ligand binding site together with the Mg^{2+}-induced decrease of the structural dynamics might explain the lack of ligand binding ability in $\mathrm{Gsw}^{\mathrm{loop}}$ due to increased conformational flexibility.
- The presence of a ligand (via constraints in the rigidity analysis) results in a more rigid ligand binding site and its surrounding.
- The tertiary interactions and ligand interactions have a cooperative, stabilizing effect on the terminal nucleotides in the P1 region. Thus, a more stable P1 region could influence the switching sequence and the following expression platform.

These results suggest that large conformational changes are not required to transmit information on the stability of the tertiary interaction to the ligand binding site, and from there together with the information on a bound ligand to the terminal P1 region. Instead, changes in dynamics of the RNA might be sufficient for the transmission of allosteric information from the ligand binding core to the P1 region. Such an allosteric mechanism without large structural changes [251] is in line with models suggesting a high degree of preorganization for the unbound state of the aptamer domain [138, 144, 146]. When relying on changes in dynamics without the need for large structural changes, the transmission of ligand binding information may be fast enough to induce the formation of the transcription terminator loop in the expression platform in time for the regulation of the RNA polymerase. This is particularly important with respect to the ligand binding time that is on the order of the transcription time for the terminator by the RNA polymerase and the kinetic regulation mechanism proposed for this riboswitch [117].

The findings on the stabilization of the RNA by the tertiary interactions may also be interesting when studying other RNA structures also carrying a RNA three-way junction together with remote tertiary interactions, such as the TPP-sensing riboswitch or the hammerhead ribozyme [150].

6 Publication III - Tertiary interactions in the unbound guanine-sensing riboswitch focus functional conformational variability to the binding site

Hanke C.A., Gohlke, H.

Submitted (2017).

This section is based on the above-mentioned manuscript. Manuscript see pages 143-179

6.1 Background

In the previous study (Publication II, Section 5), I showed that changes in dynamics of the guanine-sensing riboswitch aptamer may be sufficient for a fast response upon ligand binding without the need for large conformational rearrangements.

The unbound state of the wildtype aptamer and the G37A/C61U mutant may however differ structurally due to the destabilization introduced by the mutation. Such a difference may not have been accessible to the MD simulations in Publication II, since the unbound state was simulated starting from the bound crystal structures. Thus, the aptamers may be too stable to show large conformational changes on typical MD time scales of several hundreds of ns. In order to increase the sampling of the conformational space, I applied REMD simulations and subsequently calculated free energy landscapes for each of the systems. For this, I performed REMD simulations with an aggregate simulation time of ~16 µs.

6.2 Results

The number of replica as well as the choice of the temperature steps influence the exchange frequency between neighboring replica. In general, an exchange frequency of ~20% is recommended [209, 211]. In my simulations, I achieved exchange frequencies of ~20-30% using 64 replica and a temperature range from 273 to 363 K, where the highest temperature was chosen based on the experimental unfolding temperatures [151].

Unfolding on a global level

As a first step, I investigated the unfolding of the aptamers on a global level. To this end, I calculated Q values for each structure, which is a measure for the number of native contacts [252, 253]. In order to investigate the global unfolding, I calculated Q_{all} considering all nucleotides. From the Q_{all} values for each structure, I determined the fraction of folded structures at each temperature replica and determined melting curves. A structure was assumed to be folded if its Q_{all} value was larger than 0.75. In the absence of Mg^{2+} ions, the number of folded structures decreases with increasing temperature. This is expected, since at higher temperature structures are more likely to be unfolded and less native-like. Furthermore, I observed that at each temperature a larger number of structures are unfolded for Gsw^{loop} than for Gsw^{apt}. This agrees with the expected destabilization of the aptamer by the G37A/C61U mutation in Gsw^{loop} [151, 152]. These results show that the difference in structural stability between Gsw^{apt} and Gsw^{loop}, as well as the Mg^{2+}-induced structural stabilization, is maintained in the REMD simulations. Thus, the REMD results can be used to investigate the different behavior of the RNA with respect to the difference in stability of the tertiary interactions and at different Mg^{2+} concentrations.

Influence of tertiary interaction stability on the free energy landscape

As a next step, I calculated a two-dimensional potential of mean force (2D-PMF) for each system (Figure 23) using a reweighting scheme [225] in order to obtain an unbiased result from the REMD results. As reaction coordinates for the 2D-PMF, I used the number of native contacts within the loop-loop interactions ($Q_{loopinteractions}$) and the number of native contacts at the ligand binding site ($Q_{bindingsite}$). This allowed me to track these two structurally important features of the aptamer. The free energy landscapes of Gsw^{apt} (Figure 23A, B) contain one predominant free energy basin for both cases, when Mg^{2+} ions are present or absent. This basin is found close to the native bound structure, with a narrow range of $Q_{loopinteractions}$ values (≈ 0.98), which indicates that the tertiary interactions are stable in most structures. Additionally, the basin shows a spread with respect to $Q_{bindingsite}$ without larger energy barriers in between. This spread shows that the native contacts for the ligand binding site found in the bound crystal structure are less conserved. This is not surprising, since the unbound aptamer requires some flexibility to allow for ligand binding, which is assumed to involve the J2/3 region [14, 142-144]. Additionally, the free energy landscape is rather steep, which suggests a downhill folding pathway without notable alternatively folded conformations. In contrast, the free energy landscapes of Gsw^{loop} are flatter and contain several energetically equally favorable areas (Figure 23C, D). The positions of these basins suggest a structural variability in the tertiary interactions and in the ligand binding site within the structural ensemble of the basins. Thus, for Gsw^{loop}, structures that do not possess the native contacts within the tertiary interactions and/or the ligand binding site are also energetically favorable. These results, together with the small energy barriers between the basins (< 1 kcal/mol), suggest that the conformational ensemble of Gsw^{loop} is more heterogeneous than for Gsw^{apt}, and energetically favorable conformations can interconvert.

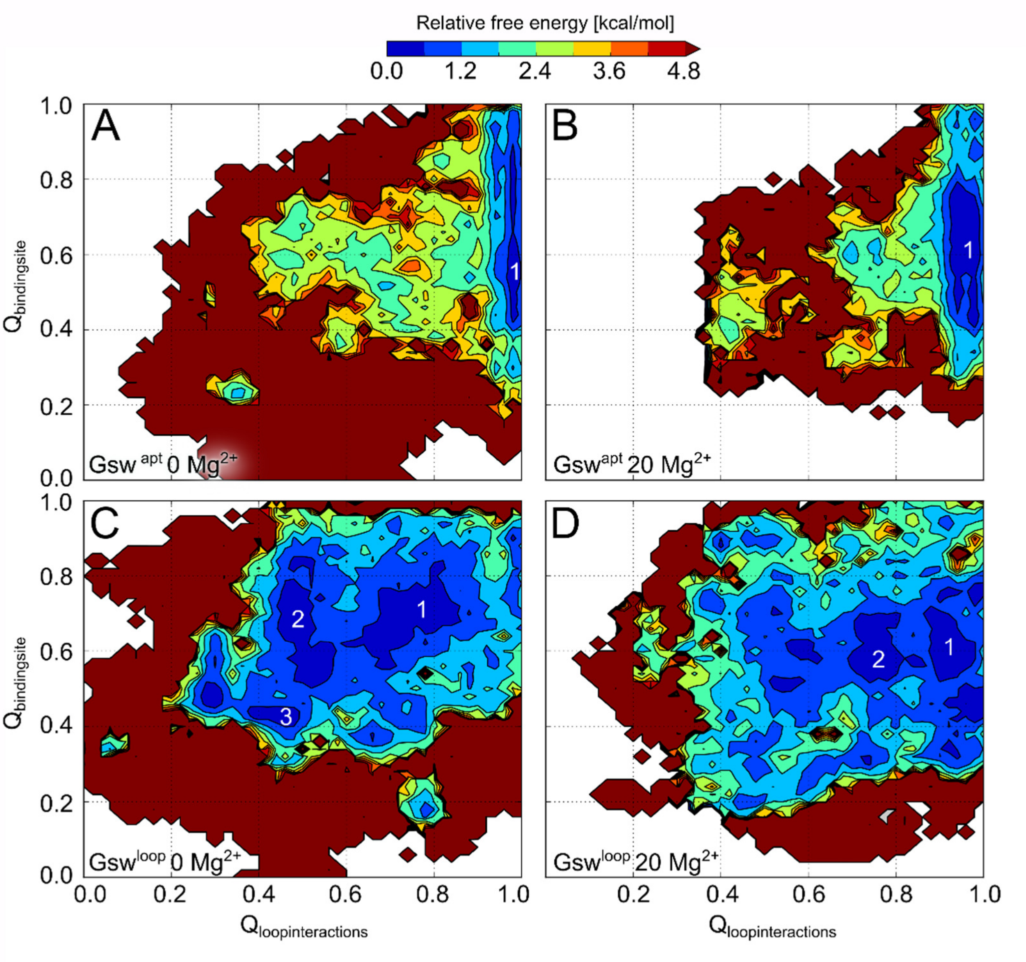
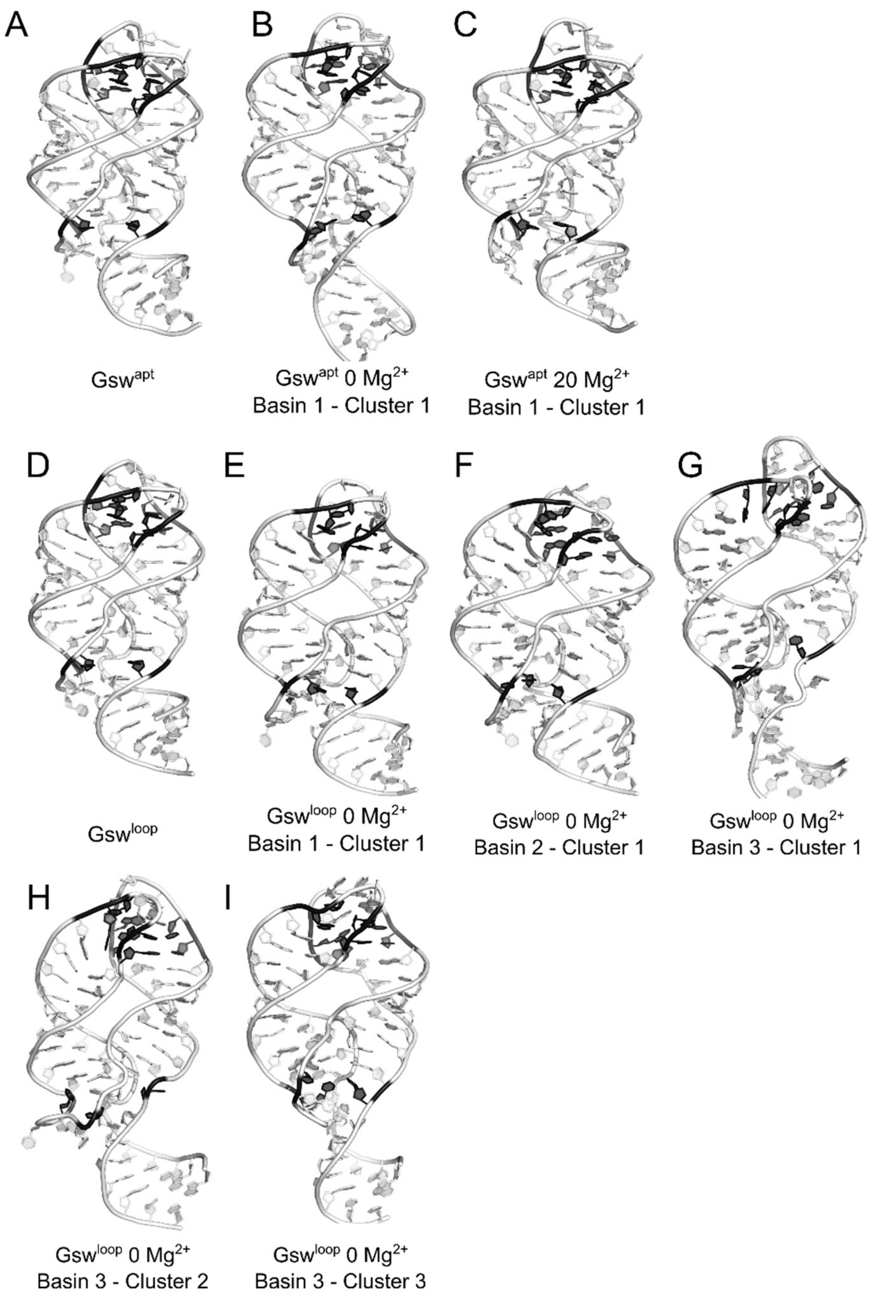


Figure 23: 2D-PMF calculated for the observables $Q_{bindingsite}$ and $Q_{loopinteractions}$ for Gsw^{apt} in the absence and the presence of 20 Mg^{2+} ions per RNA molecule (A and B, respectively) and for Gsw^{loop} in the absence and the presence of 20 Mg^{2+} ions per RNA molecule (C and D, respectively). The white numbers identify local basins. The top color scale represents the relative free energies with respect to the lowest basin. White areas were not sampled in the simulations. Figure taken from the Manuscript (page 144).

Cluster analysis of the structural ensemble within the free energy basins based on RMSD of the phosphorus atoms showed that for Gsw^{apt} one predominant conformation is found in the basins (Figure 24B, C). Together with the spread in the $Q_{bindingsite}$ values, this shows that the aptamer in Gsw^{apt} has a largely stable global fold with a structurally slightly variable ligand binding site. Furthermore, the cluster representative is very similar to the bound crystal structure (Figure 24A). These findings are in agreement with my findings from the previous study (Publication II, Section 5) that only changes in the local flexibility might be sufficient to transmit the ligand binding information towards the expression platform, without the need for large conformational changes.

In contrast, for Gsw^{loop} in the absence of Mg^{2+} ions (Figure 24E-I), only two of the free energy basins show one larger cluster, whereas one free energy basin shows three different clusters. While the cluster representatives of basin 1 and 2 (Figure 24E, F) are still similar to the initial structure (Figure 24D) (RMSD ~2.5 Å), the cluster representatives for basin 3 (Figure 24G-I) differ from the bound crystal structure by 3.4 - 5.0 Å. Thus, the structures in the basins do not only differ from the initial structure in their ligand binding site and their loop-loop interactions as seen from the location on the free energy landscape, but also in their overall fold. Additionally, the native hydrogen bond network in the loop region is destabilized for the structural ensembles in the basins, and alternative base pairs within the loops are formed. In summary, the structural ensemble for Gsw^{loop} in the absence of Mg^{2+} ions contains energetically favorable conformations that differ from the native fold with respect to the loop-loop interactions, the binding site conformation, and the overall fold.



A: Gsw^{apt}

B: Gsw^{apt} 0 Mg^{2+} Basin 1 - Cluster 1

C: Gsw^{apt} 20 Mg^{2+} Basin 1 - Cluster 1

D: Gsw^{loop}

E: Gsw^{loop} 0 Mg^{2+} Basin 1 - Cluster 1

F: Gsw^{loop} 0 Mg^{2+} Basin 2 - Cluster 1

G: Gsw^{loop} 0 Mg^{2+} Basin 3 - Cluster 1

H: Gsw^{loop} 0 Mg^{2+} Basin 3 - Cluster 2

I: Gsw^{loop} 0 Mg^{2+} Basin 3 - Cluster 3

Figure 24: Structures of the cluster representatives from free energy basins. The nucleotides involved in the formation of the loop-loop interactions and the nucleotides binding the ligand are colored in black (top black nucleotides and bottom black nucleotides, respectively). A: Initial structure of Gsw^{apt}, which is highly similar to the crystal structure with PDB ID 4FE5 [26, 104]; B: Cluster representative of the largest cluster from basin 1 for Gsw^{apt} in the absence of Mg^{2+} ions; C: Cluster representative of the largest cluster from basin 1 for Gsw^{apt} with 20 Mg^{2+} ions per RNA molecule; D: Initial structure of Gsw^{loop}, corresponding to the crystal structure with PDB ID 3RKF [152]; E-I: Cluster representatives for Gsw^{loop} in the absence of Mg^{2+} ions. Cluster representatives for the largest clusters for basin 1 (E) and basin 2 (F), and for basin 3 for the largest (G), the second-largest (H) and the third-largest cluster (I) are shown. Figure adapted from the Manuscript (page 144).

6.3 Conclusions and significance

In this study, I performed REMD simulations to study the influence of tertiary interaction stability on the accessible conformational space and on the free energy landscape of Gsw^{apt} and Gsw^{loop} in the absence and in the presence of Mg^{2+} ions. This study represents the most comprehensive computational analysis on the differences caused by a mutation in an unbound riboswitch aptamer domain using free energy landscapes calculated from REMD simulations. The main results from this work are:

- The destabilized tertiary interactions in Gsw^{loop} result in an overall less stable aptamer, as seen from the melting curves.
- The free energy landscapes for Gsw^{apt} show one predominant free energy basin close to the native state for both, the absence and the presence of Mg^{2+} ions. The steep energy landscape suggests a downhill folding pathway for the aptamer towards the native bound-like state without notable intermediate or misfolded structures.
- In the case of Gsw^{loop}, the free energy landscape is flatter and exhibits several basins. This suggests that for Gsw^{loop} in the absence of Mg^{2+} ions, folding to the native state might be slower due to the presence of intermediate or misfolded states.
- The presence of Mg^{2+} ions in Gsw^{loop} rescues the stability of the aptamer to some extent and shifts the energetically favorable areas towards the native state.

Overall, these data support the picture that the unbound Gsw^{apt} is more stable [153], and exhibits only local flexibility to allow for ligand binding. In contrast, the Gsw^{loop} aptamer with destabilized tertiary interactions forms a more heterogeneous structural ensemble comparable to what is expected for the adenine-sensing riboswitch where the tertiary interactions are not stable in the absence of Mg^{2+} ions [14, 120, 138, 148]. Thus, the lack of stable tertiary interactions could shift the kinetics of folding from a quick folding without traps or intermediate states for Gsw^{apt} to a slower folding including alternative folds and intermediates for Gsw^{loop}.

7 Summary and perspectives

In my thesis, I investigated the influence of long-range interactions on the behavior of the unbound guanine-sensing riboswitch aptamer using molecular simulations. Here, I particularly focused on the tertiary interactions formed between the loops L2 and L3, which are known to stabilize the aptamer and play a supporting role in ligand binding [143]. For this, I used a wildtype guanine-sensing riboswitch aptamer domain (Gsw^{apt}) and a G37A/C61U mutant (Gsw^{loop}), which possesses destabilized tertiary interactions. Since Gsw^{apt} and Gsw^{loop} are initially identical except for the G37A/C61U mutation, the observed differences between the two variants can be attributed to this mutation and thus the destabilization of the tertiary interactions.

As a first step, I needed to decide for a system setup for the simulations, which can reproduce experimentally found differences between Gsw^{apt} and Gsw^{loop} and between different Mg^{2+} concentrations (Section 4). Here, I found that the ff99 force field variant reproduces these differences best, while this was not possible using the ff10 variant. In addition, I developed an improved placement procedure for Mg^{2+} ions for nucleic acid simulations, which avoids impairment of RNA dynamics by the ions due to their initial placement.

Based on these findings, I moved on to investigate the influence of the tertiary interaction stability on the dynamics of the aptamer (Section 5). In this study, I found that the destabilization of the tertiary interactions in Gsw^{loop} causes an increase of the structural dynamics at the ~30 Å distant ligand binding site, which can be compensated by the presence of Mg^{2+} ions. This long-range effect is in agreement with experimental findings on the Mg^{2+}-dependent ligand binding ability of Gsw^{loop} and the suspected stabilization of the ligand binding site by the tertiary interactions [151]. Furthermore, I found that the tertiary interactions together with the stabilization of the ligand binding site cooperatively stabilize the terminal P1 region, which is the connection to the expression platform. A stabilization of the P1 region is suspected to be involved in the information transmission from the ligand binding site to the expression platform [14]. Overall, these results suggest that large conformational changes might not be necessary to transfer the information on the tertiary interaction stability and on the bound ligand to the expression platform, but changes in the aptamer's dynamics might be sufficient. If no large conformational changes are required to transfer the ligand binding information, this might accelerate the riboswitch response. This is in particular important for the riboswitch investigated here, since the time window required for ligand binding is on the order of the transcription time of the RNA polymerase for the transcription terminator. Thus, a fast response to ligand binding allows for reliable transcriptional gene regulation.

Since structural information at the atomistic level of the unbound state of the guanine-sensing riboswitch aptamer domain has remain elusive, I performed replica exchange molecular dynamics simulations (Section 6). This enhanced sampling method allowed me to sample a larger conformational space than in the previous MD simulations, and subsequently calculate free energy landscapes. The results from these calculations suggest one predominant conformation with local flexibility at the ligand binding site for Gsw^{apt}. In contrast, for Gsw^{loop}, the results suggest a heterogeneous conformational ensemble in the unbound state, which contains structures with alternative base pairs and partial alternative folds. Overall, the lack of stable tertiary interactions in Gsw^{loop} might shift the folding kinetics from a fast downhill folding without notable traps for Gsw^{apt} towards a slower folding with alternative conformations. Thus, the destabilized tertiary interactions might induce a more heterogeneous conformational ensemble, which is in line with the suspected more heterogeneous structural ensemble for the adenine-sensing riboswitch, which requires Mg^{2+} ions to form stable tertiary interactions [14, 120, 138, 148, 153].

In summary, I gained insights into the unbound state of a guanine-sensing riboswitch aptamer domain with respect to conformational heterogeneity and long-range effects of destabilized tertiary interactions.

Such knowledge about structural heterogeneity in the unbound state of riboswitches might help in the development of novel antibiotic drugs targeting riboswitches, since a structurally heterogeneous ensemble might possess ligand binding pockets with different characteristics. On the one hand, such a variety in the ligand binding pocket makes binding of a ligand, which targets one specific binding pocket conformation, more difficult. On the other hand, this might allow other molecules to bind the RNA, which could not bind the native ligand binding pocket. Furthermore, the knowledge about how tertiary interactions influence the behavior of an aptamer could be useful for the quickly emerging field of synthetic riboswitches [254, 255], where tuning the regulatory response to ligand binding could be achieved by varying the stability of the aptamer.

8 Acknowledgment

First of all, I would like to thank my supervisor Prof. Dr. Holger Gohlke for his support, guidance, and everything I learned during the last few years.
Furthermore, I would like to thank Dr. Christopher Pfleger, Dr. Emanuele Ciglia, and Giulia Pagani for reading this thesis, for their helpful suggestions, and for their constructive criticism. I would also like to thank my cooperation partners on the work that did not become part of this thesis: Prof. Dr. Claus A.M. Seidel, Dr. Hayk Vardanyan, and Dr. Simon Sindbert.

The last years were sometimes tough, but there are many people who made them an overall enjoyable and worthwhile time, which I would not want to miss. Here, I would like to particularly mention Giulia, Ana, Emanuele, Christopher, Ido, Jags, Michele, Alexandra, Kristyna, Caro, and Daniele. Thank you for your support, your advices, and for being a great company.

I am deeply grateful to my parents and my family for always supporting me and believing in me.
Finally, I would like to thank my wife Verena, for the constant support, patience (also when proofreading this thesis), and simply for being there.

9 Publications

9.1 Reprint permissions for publications

Publication I (pages 70-99)

Reprinted from "Force field dependence of riboswitch dynamics", Christian A. Hanke, Holger Gohlke, Methods in Enyzmology Vol. 553, p. 163-191, Copyright (2015), with permission from Elsevier.

Publication II (pages 100-142)

Reprinted from "Ligand-mediated and tertiary interactions cooperatively stabilize the P1 region in the guanine-sensing riboswitch", Christian A. Hanke, Holger Gohlke, PLOS ONE, 2017. **12**(6): p. e0179271.

Copyright (2017) Hanke, Gohlke. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

9.2 Publication I - Force field dependence of riboswitch dynamics

Hanke C.A., Gohlke, H.

in Computational methods for understanding riboswitches, S.-J. Chen and D.H. Burke-Aguero, Editors. 2015, Academic Press: Burlington. p. 163-191

CHAPTER SEVEN

Force Field Dependence of Riboswitch Dynamics

Christian A. Hanke, Holger Gohlke[1]
Mathematisch-Naturwissenschaftliche Fakultät, Institut für Pharmazeutische und Medizinische Chemie,
Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany
[1]Corresponding author: e-mail address: gohlke@uni-duesseldorf.de

Contents

Abstract

Riboswitches are noncoding regulatory elements that control gene expression in response to the presence of metabolites, which bind to the aptamer domain. Metabolite binding appears to occur through a combination of conformational selection and induced fit mechanism. This demands to characterize the structural dynamics of the *apo* state of aptamer domains. In principle, molecular dynamics (MD) simulations can give insights at the atomistic level into the dynamics of the aptamer domain. However, it is unclear to what extent contemporary force fields can bias such insights. Here, we show that the Amber force field ff99 yields the best agreement with detailed experimental observations on differences in the structural dynamics of wild type and mutant aptamer domains of the guanine-sensing riboswitch (Gsw), including a pronounced influence of Mg^{2+}. In contrast, applying ff99 with parmbsc0 and parmχ_{OL} modifications (denoted ff10) results in strongly damped motions and overly stable tertiary loop–loop interactions. These results are based on 58 MD simulations with an aggregate simulation time >11 μs, careful modeling of Mg^{2+} ions, and thorough statistical testing. Our results suggest that the moderate stabilization of the χ-*anti* region in ff10 can have an unwanted damping effect on functionally relevant structural dynamics of marginally stable RNA systems. This suggestion is supported by crystal structure analyses of Gsw aptamer domains that reveal χ torsions with high-*anti* values in the most mobile regions.

Methods in Enzymology, Volume 553
ISSN 0076-6879
http://dx.doi.org/10.1016/bs.mie.2014.10.056

We expect that future RNA force field development will benefit from considering marginally stable RNA systems and optimization toward good representations of dynamics in addition to structural characteristics.

1. INTRODUCTION

Riboswitches are noncoding genetic regulatory elements mostly found in the 5′ untranslated region of bacterial mRNA. Binding of a small molecule metabolite to the aptamer domain results in a large conformational change in the downstream expression platform, which modulates gene expression (Tucker & Breaker, 2005). Structural and related studies have revealed that riboswitches appear to bind their ligands through a combination of conformational selection and induced fit mechanisms (Serganov & Nudler, 2013). This demands to characterize the structural dynamics and energetics of the *apo* state of aptamer domains in order to understand the full process of gene regulation.

The role of the *apo* state is of particular importance for transcriptional gene regulation by riboswitches as there the decision whether the genes should be transcribed or not has to be made already during transcription (Stoddard et al., 2010). The guanine-sensing riboswitch (Gsw) found in the xpt-pbuX operon of *Bacillus subtilis* exerts transcriptional gene regulation and belongs to the class of purine sensing riboswitches (Batey, 2012; Batey, Gilbert, & Montange, 2004). Besides guanine, Gsw binds related ligands such as hypoxanthine (Batey et al., 2004; Gilbert, Stoddard, Wise, & Batey, 2006). Its aptamer domain consists of three paired regions P1, P2, and P3, two loops L2 and L3 capping the P2 and P3 regions, respectively, and three joining regions J1/2, J2/3, and J3/1 connecting the paired regions (Fig. 1A and B). The ligand is tightly bound in the binding pocket formed by the three joining regions (Batey et al., 2004). The aptamer terminal P1 helix connects to the downstream expression platform and is essential for the gene regulation (Batey et al., 2004). Tertiary interactions between the L2 and L3 loops are formed by two base quadruples stacked onto each other (Fig. 1C–F). These interactions were found to be present in the unbound state of the aptamer domain (Noeske, Schwalbe, & Wöhnert, 2007), and have been suggested to restrict the conformational freedom of the domain, thereby promoting ligand binding under physiological conditions (Stoddard, Gilbert, & Batey, 2008). In a mutant without these tertiary interactions, the aptamer's ability to bind

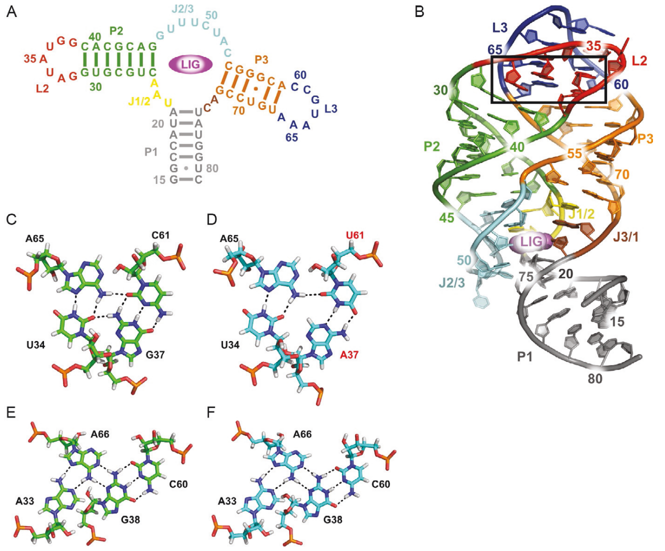


Figure 1 (A) Secondary structure and sequence used in this study of the aptamer domain of Gsw^{apt}; the secondary structure elements are assigned according to Batey et al. (2004). In addition, the ligand-binding site is indicated by the purple oval and denoted by LIG. (B) Tertiary structure of the aptamer domain of the Gsw bound to hypoxanthine (LIG; PDB ID 4FE5, Stoddard et al., 2013). The secondary structure elements are assigned as in (A). The black rectangle indicates the location of the base quadruples. (C and D) Upper base quadruple formed by the L2 and L3 loops for Gsw^{apt} (C) and Gsw^{loop} (D). Here, the G37A/C61U mutation is located (highlighted by the red labels). (E and F) Lower base quadruple formed by the L2 and L3 loops for Gsw^{apt} (E) and Gsw^{loop} (F).

hypoxanthine was abolished (Batey et al., 2004). In contrast, the G37A/C61U double mutation, reducing the number of hydrogen bonds within the upper base quadruple from seven in the wild type to five, led to an aptamer domain where formation of loop–loop interactions in the *apo* state and binding of hypoxanthine show a pronounced Mg^{2+} dependence (Buck, Noeske, Wöhnert, & Schwalbe, 2010; Noeske, Buck, et al., 2007). To what extent the loop–loop interactions contribute to the stabilization of the P1 helix, in addition to the ligand (Batey et al., 2004), has not been described so far.

With the long-term goal to gain insights at the atomistic level by molecular dynamics (MD) simulations into the function of the Gsw, we set out here to identify simulation conditions that allow appropriate modeling of the experimentally demonstrated marginal stability of the unbound aptamer domain, in particular when the G37A/C61U double mutation is present, and the pronounced influence of Mg^{2+} on the structural dynamics. With this we address one of the challenges that MD simulations of RNA face, the potential bias due to the applied force field (Sponer et al., 2014).

While contemporary force fields for RNA well describe canonical helices, they may show different successes on more complex RNAs (see references Cheatham & Case, 2013; Sponer et al., 2014; Sponer, Cang & Cheatham, 2012 for recent reviews). Here, we test the Amber (Case et al., 2005) force field ff99 (Wang, Cieplak, & Kollman, 2000) itself and together with the refinements parmbsc0 (Perez et al., 2007) and parmbsc0 + parmχ_{OL3} (Zgarbova et al., 2011). Note that the latter combination, ff99 + parmbsc0 + parmχ_{OL3}, has been termed ff10 in the context of RNA simulations in Amber (as of Amber12, this force field is also referred to as ff12). In parmbsc0, the α/γ dihedrals have been reparameterized in order to prevent the collapse of B-DNA due to the accumulation of irreversible nonnative γ-*trans* dihedral states (Perez et al., 2007; Sponer et al., 2014). As γ-*trans* degradation has never substantially affected canonical A-RNA simulations (Reblova et al., 2006), it was initially not clear if parmbsc0 brings any substantial changes for RNA simulations (Besseova, Otyepka, Reblova, & Sponer, 2009). However, long-timescale MD simulations of RNA revealed sudden irreversible transitions to ladder-like structures (Mlynsky et al., 2010), which required a reparameterization of the χ dihedral to suppress *anti* to high-*anti* χ shifts in RNA (Sponer et al., 2014), leading to the parmχ_{OL3} refinement (Zgarbova et al., 2011). At present, it is recommended to use parmχ_{OL3} together with parmbsc0 (Banas et al., 2010).

Another level of complexity arises in terms of describing the pronounced influence of Mg^{2+} ions on the structural dynamics of the aptamer domain of the Gsw (Buck et al., 2010). Mg^{2+} ions play an important role for the stability of nucleic acid structures (Draper, Grilley, & Soto, 2005; Pyle, 2002; Woodson, 2005). However, Mg^{2+} ions are challenging to treat in MD simulations due to the slow exchange kinetics of their first shell ligands (Ohtaki & Radnai, 1993), the low-diffusion coefficient (Mills & Lobo, 1989), and difficulties in modeling polarization and charge-transfer effects with fixed-charge additive force fields (Auffinger, 2012; Sponer et al., 2014). Here, we test two parameter sets for Mg^{2+} ions available for the use with Amber force fields: the long-used parameters developed by Aqvist (1992) and a set of parameters developed recently with the aim to improve agreement with experimentally determined kinetic properties of Mg^{2+} ions, especially with respect to the exchange of the first solvation shell (Allner, Nilsson, & Villa, 2012). Furthermore, we probe the influence of the method of Mg^{2+} ion placement on the simulation results. In total, we report on MD simulations with an aggregate sampling >11 μs (Table 1).

2. METHODS

2.1. Setup of the simulations

The starting structures for MD simulations of the aptamer domain of the Gsw in the *apo* state were obtained from the respective ligand bound crystal structures after removal of the ligand, ions, and crystal water molecules. The sequence of the wild type structure (PDB ID 4FE5, Stoddard et al., 2013) was adapted to match the sequence of the G37A/C61U mutant (PDB ID 3RKF, Buck et al., 2011). As a result, the wild type and mutant structures only differ in the nucleotides 37 and 61. The modified wild type will be referred to as Gsw^{apt}, and the G37A/C61U mutant will be referred to as Gsw^{loop}, according to Buck et al. (2011).

For both structures, three simulation systems with different Mg^{2+} concentrations (0, 12, and 20 Mg^{2+} ions per RNA molecule) were set up according to investigations on the Mg^{2+} dependence of Gsw^{loop} properties by Buck et al. (2010). For the setup, the *leap* program from the AmberTools suite of programs was used (Case et al., 2012).

For investigating the influence of the initial Mg^{2+} ion placement, one set of simulations was set up using the default placement procedure for ions of *leap*. The other simulations were set up using a modified protocol to place the Mg^{2+} ions with a first hydration shell of six water molecules: initially, a dummy atom was placed by *leap* instead of a Mg^{2+} ion, which had a 2+ charge, but a larger radius of 4 Å. This radius is similar to the one of a

Table 1 List of MD simulations presented in this work

Gsw variant	Number of Mg^{2+} ions[a]	Force field	Simulation time[b]	Mg^{2+} parameters
$\mathrm{Gsw}^{\mathrm{apt}}$	0	ff99[c]	3×200	Aqvist[d]
	12	ff99[c]	3×200	Aqvist[d]
	20	ff99[c]	3×200	Aqvist[d]
$\mathrm{Gsw}^{\mathrm{loop}}$	0	ff99[c]	3×200	Aqvist[d]
	12	ff99[c]	3×200	Aqvist[d]
	20	ff99[c]	3×200	Aqvist[d]
$\mathrm{Gsw}^{\mathrm{apt}}$	0	parmbsc0[e]	3×200	Aqvist[d]
	12	parmbsc0[e]	3×200	Aqvist[d]
	20	parmbsc0[e]	3×200	Aqvist[d]
$\mathrm{Gsw}^{\mathrm{loop}}$	0	parmbsc0[e]	3×200	Aqvist[d]
	12	parmbsc0[e]	3×200	Aqvist[d]
	20	parmbsc0[e]	3×200	Aqvist[d]
$\mathrm{Gsw}^{\mathrm{apt}}$	0	ff10[f]	3×200	Aqvist[d]
	12	ff10[f]	3×200	Aqvist[d]
	20	ff10[f]	3×200	Aqvist[d]
$\mathrm{Gsw}^{\mathrm{loop}}$	0	ff10[f]	3×200	Aqvist[d]
	12	ff10[f]	3×200	Aqvist[d]
	20	ff10[f]	3×200	Aqvist[d]
$\mathrm{Gsw}^{\mathrm{apt}}$	20	ff99[c]	3×100	Allner et al.[g]
$\mathrm{Gsw}^{\mathrm{apt}}$	20	ff99[c]	1×100	Aqvist[d,h]

[a] Per RNA molecule.
[b] In ns.
[c] Wang et al. (2000).
[d] Aqvist (1992).
[e] Perez et al. (2007).
[f] Zgarbova et al. (2011).
[g] Allner et al. (2012).
[h] The default placement procedure of *leap* was used for the Mg^{2+} ions in this simulation.

hexahydrated Mg^{2+} ion (Robinson, Gao, Sanishvili, Joachimiak, & Wang, 2000). Afterwards, the dummy atom was replaced by a Mg^{2+} ion with the correct radius and six water molecules surrounding it. Using a larger dummy atom initially in this two-step procedure results in a position of the Mg^{2+} ion that also allows accommodating the first hydration shell. Using a Mg^{2+} ion with the correct radius in the initial step instead would result in the ion be placed too closely to the RNA to accommodate the waters.

Using either placement method, 12 or 20 Mg^{2+} ions with parameters of Aqvist (1992) were added to Gsw^{apt} and Gsw^{loop}. To neutralize the simulation systems, also the ones without Mg^{2+} ions, Na^+ ions (Na^+ parameters for ff99 and ff99 + parmbsc0: Aqvist, 1990 and for ff10: Joung & Cheatham, 2008) were added by *leap*. The systems were then placed in an octahedral box of TIP3P water (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983) such that the distance between the edge of the water box and the closest RNA atom was at least 11 Å.

For each of the six simulation systems (Gsw^{apt} and Gsw^{loop} with 0, 12, or 20 Mg^{2+} ions per RNA molecule), three independent MD simulations were performed. These simulations were performed for each of the three force fields described earlier: ff99 (Wang et al., 2000), ff99 + parmbsc0 (Perez et al., 2007), and ff10 (Zgarbova et al., 2011) (Table 1).

Before the start of the MD simulation, each system was minimized by 200 steps of the steepest descent minimization, followed by 50 steps of conjugate gradient minimization. The particle mesh Ewald method (Darden, York, & Pedersen, 1993) was used to treat long-range electrostatic interactions, and bond lengths involving bonds to hydrogen atoms were constrained using the SHAKE algorithm (Ryckaert, Ciccotti, & Berendsen, 1977). The time step for all MD simulations was 2 fs, with a direct-space nonbonded cutoff of 9 Å. Applying harmonic force restraints with force constants of 5 kcal mol^{-1} $Å^{-2}$ to all solute atoms (and if present, also to the Mg^{2+} ions and their first hydration shell waters), we carried out canonical ensemble (NVT)-MD simulations for 50 ps, during which the system was heated from 100 to 300 K. Subsequent isothermal isobaric ensemble (NPT)-MD simulations were used for 50 ps to adjust the solvent density. Finally, the force constants of the harmonic restraints on positions of RNA, Mg^{2+} ions, and their first hydration shell waters were gradually reduced to 1 kcal mol^{-1} $Å^{-2}$ during 250 ps of NVT-MD simulations. This was followed by 50 ps of NVT-MD simulations without applying positional restraints. From the following 200 ns of NVT-MD simulations at 300 K performed with the GPU version of *pmemd* (Salomon-Ferrer, Gotz, Poole, Le Grand, & Walker, 2013) from the Amber 11 suite of programs (Case et al., 2005, 2010) conformations were extracted every 20 ps.

In order to investigate the influence of the Mg^{2+} parameters on the structural dynamics of the aptamer domain, we also performed MD simulations of Gsw^{apt} in the presence of 20 Mg^{2+} ions now using the recently developed Mg^{2+} parameters (Allner et al., 2012). The simulated system was otherwise identical in terms of the initial coordinates and the force field parameters for the RNA (ff99) and the water model (TIP3P). For the system, three independent MD simulations of 100 ns length each were performed.

2.2. Trajectory analysis

The MD trajectories were analyzed using the *cpptraj* tool (Roe & Cheatham, 2013) from the AmberTools suite of programs (Case et al., 2012). The initial 50 ns of each trajectory was not considered for analysis.

The radius of gyration (R_g), a measure for the compactness of a structure, was calculated for all atoms except the P1 region. Prior to the calculation of root-mean-square fluctuations (RMSF) over all atoms, global translational, and rotational differences between the structures along the trajectory were removed by least-squares fitting onto the average structure. In order to reduce the influence of very mobile regions on the picture of internal motions (Gohlke, Kuhn, & Case, 2004), conformations were root-mean-square fitted on those 80% of the nucleotides with the lowest fluctuations. These 54 nucleotides, which henceforth will be referred to as “core nucleotides”, were chosen from an initial calculation of the mean RMSF over the three simulations of Gsw^{apt} in the absence of Mg^{2+} ions. The root-mean-square deviations (RMSD) of atomic positions were calculated as a measure for the structural similarity with respect to the initial structure. The RMSD was calculated after aligning the conformations onto the starting structure using the core nucleotides; the RMSD was then calculated for all atoms of these core nucleotides (Table 2).

The occupancy of hydrogen bonds between bases in the base quadruples of the L2/L3 region was determined using default geometrical parameters (distance, 3.5 Å and angle, 120°). The statistical uncertainty within one trajectory was determined from block averaging over blocks of 10 ns length. Finally, the values reported in Table 3 were averaged over all hydrogen bonds between a pair of bases.

In order to characterize the behavior of Mg^{2+} ions, we calculated diffusion coefficients from the mean-square displacements of all Mg^{2+} ions as moving averages with varying time windows. Mg^{2+} ions with a reduced number of water molecules in the first hydration shell were identified as those that have at least one RNA atom within 3.5 Å.

Table 2 RMSD with respect to the initial structure of the core nucleotides of Gsw^{apt} and Gsw^{loop} using different force fields[a]

Simulated system		ff99	bsc0	ff10
Gsw^{apt}	0 Mg^{2+}[b]	2.9 ± 0.4	3.5 ± 0.5	2.7 ± 0.3
	12 Mg^{2+}[b]	2.7 ± 0.4	2.8 ± 0.3	2.3 ± 0.3
	20 Mg^{2+}[b]	2.1 ± 0.3	3.0 ± 0.4	2.2 ± 0.3
Gsw^{loop}	0 Mg^{2+}[b]	3.7 ± 0.6	4.1 ± 0.7	2.7 ± 0.3
	12 Mg^{2+}[b]	2.3 ± 0.3	2.9 ± 0.3	2.5 ± 0.3
	20 Mg^{2+}[b]	2.5 ± 0.3	3.6 ± 0.3	2.6 ± 0.2

[a] In Å; mean ± SEM over three trajectories of 200 ns length. Nucleotides 5–21, 23–33, 36–47, and 51–64 are taken as core nucleotides. All atoms of the aptamer domain were considered.

[b] Number of Mg^{2+} ions per RNA molecule.

Table 3 Hydrogen bond occupancy in the L2/L3 loop region of Gsw^{apt} and Gsw^{loop} using different force fields[a]

RNA	Pair of nucleotides	ff99	bsc0	ff10
Gsw^{apt}	U34–A65	80.9 ± 10.6	96.9 ± 0.7	95.7 ± 1.0
	U34–G37	86.1 ± 10.5	99.6 ± 0.2	99.2 ± 0.3
	C61–A65	74.9 ± 12.0	98.4 ± 0.6	99.3 ± 0.2
	C61–G37	99.6 ± 0.3	99.9 ± 0.0	99.9 ± 0.0
	A33–A66	69.4 ± 12.6	88.0 ± 7.5	97.1 ± 1.4
	G38–A66	57.5 ± 9.1	63.9 ± 3.4	80.2 ± 2.6
	G38–C60	97.8 ± 0.8	97.2 ± 0.5	97.8 ± 0.5
Gsw^{loop}	U34–A65	66.1 ± 11.9	16.0 ± 10.4	87.6 ± 10.1
	U61–A65	13.1 ± 11.1	28.7 ± 14.0	56.9 ± 11.1
	U61–A37	70.2 ± 7.9	95.9 ± 1.6	67.9 ± 6.1
	A33–A66	50.0 ± 15.0	1.5 ± 1.7	85.2 ± 10.1
	G38–A66	41.6 ± 15.1	54.8 ± 10.3	84.5 ± 7.7
	G38–C60	77.0 ± 10.0	98.4 ± 1.0	99.2 ± 0.4

[a] In %; mean ± SEM over three trajectories of 200 ns length in the absence of Mg^{2+} ions. The nucleotides 34, 37, 61, and 65 belong to the upper base quadruple (Fig. 1C and D) where the G37A/C61U mutation is located; the nucleotides 33, 38, 60, and 66 form the lower base quadruple (Fig. 1E and F).

Circular variances of dihedral angles were first calculated on a per-nucleotide basis and afterwards averaged over all nucleotides. The circular variance for a torsion angle θ was calculated following MacArthur and Thornton (1993) as:

$$\mathrm{Var}(\theta) = 1 - \overline{R} \tag{1}$$

where $\overline{R} = R/n$, n is the number of values, and R is given as

$$R^2 = \left(\sum\nolimits_{i=1}^{n} \cos\theta_i\right)^2 + \left(\sum\nolimits_{i=1}^{n} \sin\theta_i\right)^2 \tag{2}$$

Reported mean values (±standard error in the mean, SEM) were calculated over the independent simulations that were performed for each system. The overall SEM was calculated from the SEMs of the single trajectories according to the laws of error propagation. For testing the equality of two mean values a Welch two-sample *t*-test (Snedecor & Cochran, 1989; Welch, 1947) was applied, which does not require that the two samples have equal variances, as implemented in the program R.

3. RESULTS AND DISCUSSION

3.1. Placement of Mg^{2+} ions

Mg^{2+} ions have a strong influence on the structure and stability of RNA molecules (Draper et al., 2005; Woodson, 2005). However, only few MD simulations of RNA systems have been carried out using Mg^{2+} ions (Auffinger, 2012; Saini, Homeyer, Fulle, & Gohlke, 2013). This is a consequence of the challenges that persist in modeling these cations (see above), which result in poor sampling properties and the difficulty with current MD techniques to simulate losing of a ligand (e.g., a water or a nucleotide atom) from the inner-coordination sphere as a prerequisite to form a new inner-sphere complex (e.g., with a(nother) nucleotide atom) (Auffinger, 2012). This necessitates to pay particular attention to the initial placement of Mg^{2+} ions when setting up a simulation system, particularly if no direct experimental information on the location of Mg^{2+} is available as in the case of the aptamer domain of Gsw.

Applying the default placement procedure of the *leap* program (Schafmeister, Ross, & Romanovski, 1995) of the Amber software (Case et al., 2005) ions are added to favorable positions near the solute according to gas-phase electrostatics. For RNA, this is usually close to the negatively charged phosphate groups of the backbone, as shown for Gsw^{apt} in Fig. 2A for 20 Mg^{2+} ions after the thermalization step involving 500 ps of MD simulations. Already then all ions possess an inner-sphere hydration shell of at most five water molecules, i.e., they have at least one inner-sphere contact with an oxygen atom of the RNA backbone, likely due to the lack of a complete hydration shell around the ions after the placement. As a consequence, none of the ions moves away from its initial position during the first 100 ns of the MD simulations (Fig. 2B; note that a broader ion cloud particularly in the P1 region results from ion movements coupled to RNA motions).

A particularly favorable position exists if two phosphate groups are close together in the starting structure, as given for G27 in the P2 region and G68 in the P3 region (Fig. 2C). The Mg^{2+} ion placed there practically acts as a distance restraint between the two groups even over simulation times of 200 ns due to the slow exchange kinetics of inner-sphere ligands: the ion–phosphate oxygen interactions never break during that time. Such stable ion–phosphate oxygen interactions cause compaction of the surrounding structure (Fig. 2D). In addition, we observed that the connected nucleotides move in a correlated manner, which extends to sequentially neighboring nucleotides (data not shown). Even if placed as a hexahydrated Mg^{2+} bound initially as an outer-sphere complex at such a position, the Mg^{2+} ion can lose a water molecule from the inner sphere and directly chelate to the phosphate groups (Fig. 3A). Such an exchange is fostered by the force field bias toward inner-shell binding of Mg^{2+} ions in RNA simulations (Auffinger, 2012; Reblova et al., 2006, 2003).

As an alternative to the default placement, we developed a procedure in order to place Mg^{2+} ions with a first hydration shell by *leap*. Initially, we added dummy ions to the RNA with a charge equal to Mg^{2+} but a radius of 4 Å, which is close to the radius of a hexahydrated Mg^{2+} ion (Robinson et al., 2000). Subsequently, these dummy ions were replaced by Mg^{2+} ions with a first hydration shell of six water molecules, and the whole system was solvated in a box of TIP3P water. Harmonic restraints were then applied to the RNA, the Mg^{2+} ions, and the first hydration shell waters during the thermalization in order to counteract ion–RNA backbone interactions. As a result, it took at least 5 ns during the production runs after removal of all restraints before the first Mg^{2+} ion lost an inner-sphere water; further water losses of other ions were observed throughout the trajectory such that after 200 ns between 60% and 80% of the Mg^{2+} ions were still hexahydrated (Fig. 3A). If restraints on the ions and first hydration shell waters were omitted in the thermalization phase, occasionally Mg^{2+} ions lost one water

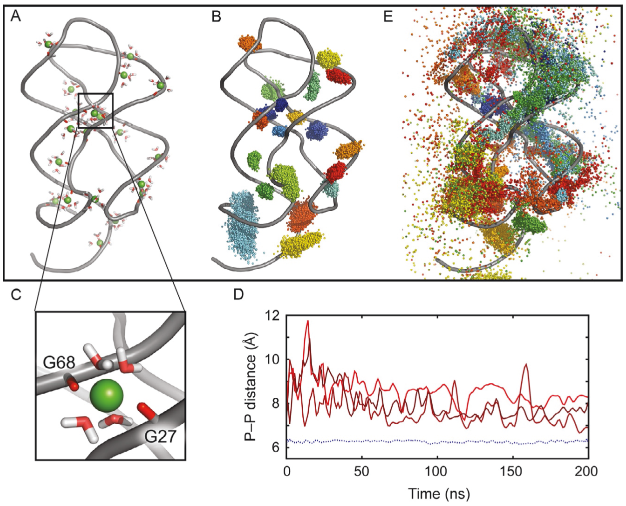


Figure 2 (A) Initial placement of 20 Mg^{2+} ions (green spheres) and surrounding water molecules (sticks) around the structure of the aptamer domain of the Gsw (with the backbone displayed as gray line). The ions were placed prior to the water molecules at electrostatically favorable locations as implemented in *leap* (Case et al., 2005) using the default procedure and are thus located very closely to the RNA backbone. (B and E) Positions of 20 Mg^{2+} ions during 100 ns of MD simulation. Different colors represent different Mg^{2+} ions. The ions were initially placed by *leap* either using the default procedure (B) or using the improved placement procedure (see Section 2) considering a first hydration shell (E). (C) Close up of a Mg^{2+} ion connecting the OP1 atoms of nucleotides G27 and G68 (red sticks connected to the RNA backbone in dark gray). Four water molecules (red and white sticks) form the hydration shell of the Mg^{2+} ion. (D) Distance calculated between the P atoms of nucleotides G27 and G68 shown in (C) during MD simulations of 200 ns length. The blue dotted line represents the distance in a simulation with the default placement of the ions; the three red solid lines, which were smoothed for better visibility, represent the distances in simulations with the improved placement procedure.

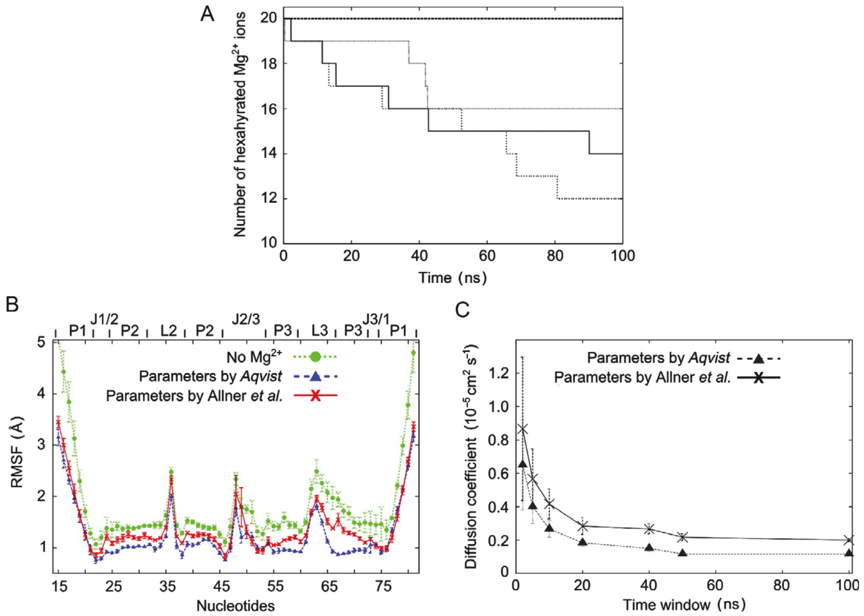


Figure 3 (A) Number of Mg^{2+} ions having six water molecules during Gsw^{apt} simulations with Mg^{2+} ions. The black dashed lines at 20 represent results for the three simulations using the Mg^{2+} parameters from Allner et al. (2012); the gray solid and dotted lines represent results for the three simulations using the Mg^{2+} parameters from Aqvist (1992). (B) Per-nucleotide RMSF values obtained from MD simulations of Gsw^{apt} without Mg^{2+} ions (green (light gray in the print version) dotted line with filled circle), with 20 Mg^{2+} ions with Mg^{2+} parameters from Aqvist (blue (dark gray in the print version) dashed line with filled triangle), or with 20 Mg^{2+} ions with Mg^{2+} parameters from Allner et al. (red (gray in the print version) solid line with "X" symbol); the mean ± SEM over three trajectories of 100 ns length is given. (C) Moving average of the diffusion coefficient (±SEM) over 20 Mg^{2+} ions calculated for varying time windows along three trajectories of 100 ns length of Gsw^{apt} simulations. The black solid line corresponds to the simulations using the Mg^{2+} parameters from Allner et al.; the gray dashed line corresponds to the simulations using the Mg^{2+} parameters from Aqvist.

molecule already there and formed an inner-sphere complex with the RNA (data not shown).

The hexahydrated Mg^{2+} ions show a dramatically different mobility during the MD simulations compared to the ions placed by the default procedure (Fig. 2B), with ions now preferentially and thoroughly sampling the space between two backbones, but also exploring almost all of the simulation box (Fig. 2E). This should allow the ions sufficient time to equilibrate prior to the formation of outer-sphere complexes with the RNA, considering fluctuation times of the bulk Mg^{2+} density of 5–8 ns, depending on the Mg^{2+} abundance, found by Hayes et al. in MD simulations of the SAM-I riboswitch (Hayes et al., 2012).

In line with the high mobility of hexahydrated Mg^{2+} ions, which show tumbling motions even when bound to RNA (Auffinger, Bielecki, & Westhof, 2003), a much less restricted structural dynamics of Gsw^{apt} is observed now in the region between G27 and G68: the distance between the two nucleotides is up to 1.5-fold larger than when restrained by inner-sphere binding of a Mg^{2+} ion and shows fluctuations up to 3 Å (Fig. 2D), resulting in noncorrelated movements of the two nucleotides (data not shown).

Thus, for all subsequent simulations, we used our optimized procedure for Mg^{2+} placement.

3.2. Parameters of Mg^{2+} ions

Describing Mg^{2+} ions by the nonbonded model, i.e., representing their interactions by Coulombic and Lennard-Jones terms, allows switching the coordination number and ligand exchange at the metal center. This and the simplicity of the model are reasons why this model is still widely used for representing Mg^{2+} ions in biomolecular simulations, despite it oversimplifying interactions between the ions and surrounding atoms and the existence of more sophisticated models (Li, Roberts, Chakravorty, & Merz, 2013). One only needs to determine Lennard-Jones parameters for the nonbonded model. Aqvist developed Lennard-Jones parameters for Mg^{2+} more than 20 years ago (Aqvist, 1992) using the hydration free energy as objective value; these parameters have been adopted for use in Amber force fields. However, these parameters lead to an ion–water exchange rate from the first solvation shell two orders of magnitude slower than the experimental rate (Allner et al., 2012). Hence, Allner et al. (2012) reparameterized the repulsive Lennard-Jones term starting from the CHARMM27 parameters for Mg^{2+} (Foloppe & MacKerell, 2000; MacKerell & Banavali, 2000) with the objective to reproduce this rate. Recently, Li et al. (2013) designed Lennard-Jones parameters for Mg^{2+} ions specifically for use in PME-MD simulations, with the objective to reproduce the hydration free energy and the ion–oxygen distance in the first hydration shell (CM set). As absolute hydration free energies for Mg^{2+} computed with the CM set deviate more strongly from the experimental value than those computed with the Amber force field adopted Aqvist parameters (Aqvist, 1992), we only tested the latter and the ones by Allner et al. (2012) with respect to their influence on the structural dynamics of Gsw^{apt}.

Atomic RMSF averaged at the nucleotide level (Fig. 3B) show that, in general, the MD simulations with 20 Mg^{2+} ions using either one of the two parameter sets are more similar to each other than to the MD simulations without Mg^{2+} ions; the latter simulations result in the highest RMSF, indicating the lack of a stabilizing effect on the RNA in the absence of Mg^{2+} ions. This result is at variance with findings by Reblova et al. (2006), which reported that a similar concentration of Mg^{2+} ions exerted surprisingly little effect on the dynamics of a long-duplex RNA, but in agreement with findings by Reblova et al. of stabilizing effects of Mg^{2+} in MD simulations of the 5*S* rRNA Loop E (Reblova et al., 2003). Comparing the two sets of MD simulations with Mg^{2+} ions, the one with parameters by Allner et al. shows slightly higher RMSF than the one with Aqvist parameters (Fig. 3B) and, hence, a less stabilizing effect by the Mg^{2+} ions. This may reflect findings by Allner et al. on *add* A-riboswitch MD simulations, where their new parameters resulted in, on average, less ions bound to the RNA via outer-sphere contacts than when using the CHARMM27 parameters, which result in similar ion–water characteristics as the Aqvist parameters (Allner et al., 2012). From a global perspective, the stabilizing influence of the Mg^{2+} ions is also reflected in the radius of gyration (R_g) as a measure of the compactness of Gsw^{apt}: $R_g = 16.5 \pm 0.2$ Å with the Mg^{2+} parameters from Allner et al. and 16.2 ± 0.1 Å with the Mg^{2+} parameters from Aqvist; for comparison, without Mg^{2+} ions, $R_g = 17.1 \pm 0.3$ Å.

Higher exchange rates have been found to be coupled to faster ion diffusion likely because the solvent structure is disrupted during the exchange (Moller, Rey, Masia, & Hynes, 2005). In line with this, our calculations of diffusion coefficients (D) over the 20 Mg^{2+} ions in the Gsw^{apt} system show 50–100% higher D values with the parameters of Allner et al. than with the ones of Aqvist for different time windows used in the moving average calculations (Fig. 3C). The computed D values for very short-time windows (2 ns; 0.86×10^{-5} $cm^2\ s^{-1}$ for the parameters of Allner et al.; 0.65×10^{-5} $cm^2\ s^{-1}$ for the parameters of Aqvist) are in very good agreement with experimental values reported for $MgSO_4$ (0.85×10^{-5} $cm^2\ s^{-1}$ at infinite dilution, Harned & Hudson, 1951) and the Mg^{2+} ion (0.71×10^{-5} $cm^2\ s^{-1}$, Mills & Lobo, 1989). This suggests that the majority of ions move essentially freely within this time window over the course of the MD simulations of 200 ns. For time windows larger than 50 ns, D drops by ~1/4 in the case of Aqvist parameters and ~1/3 in the case of parameters of Allner et al. with respect to the initial values (Fig. 3C). This is a result of a reduced ion mobility over these time windows due to the formation of inner-sphere contacts with Gsw$^{\text{apt}}$ in the case of Aqvist parameters (between 20% and 40% of all ions; Fig. 3A) and outer-sphere contacts formed by ions of both parameterizations (Fig. 3A).

While no high-resolution information on Mg^{2+} binding to the Gsw is available (Batey et al., 2004; Buck et al., 2011), it was unexpected that in neither of the three simulations with parameters of Allner et al. an inner-sphere contact was formed (Fig. 3A), taking into account that such contacts do occur in related purine-binding riboswitches (PDB ID 1Y26, Serganov et al., 2004 and 3LA5, Dixon et al., 2010). The larger drop in D in the case of parameters of Allner et al. may reflect that these parameters result in a Mg^{2+} hydration free energy that is ~15 kcal mol^{-1} less favorable (−417.3 kcal mol^{-1}, Allner et al., 2012) than that computed with Aqvist parameters (−432.6 kcal mol^{-1}, Li et al., 2013), employing the PME method in both cases. We speculate that this may result in a stronger coupling between the ions, observed for outer-sphere Mg^{2+} ions recently (Hayes et al., 2012), which may not have become visible during MD simulations of only 10 ns length by Allner et al.

From the above results, we find it difficult to discern that one of the two Mg^{2+} parameterizations is advantageous over the other with respect to MD simulations of the aptamer domain. We thus decided to use the Mg^{2+} parameters by Aqvist et al. for all subsequent simulations because of the longer experience with these parameters.

3.3. Force field influence

3.3.1 Formation of ladder-like structures

The irreversible formation of ladder-like RNA structures in long-MD simulations when using ff99 and ff99 + prmbsc0 has been described for a series of systems (Banas et al., 2010; Mlynsky et al., 2010; Sklenovsky et al., 2011), with different simulation times until the structural collapse was observed. In our case of 18 MD simulations of Gsw$^{\text{apt}}$ and Gsw$^{\text{loop}}$ with ff99 of 200 ns length, we observed the irreversible formation of a ladder-like structure of the aptamer domain only once at the end of a simulation in the terminal P1 region. When extending some of these simulations to 500 ns (data not shown), all aptamer domains were stable but one. This is similar to MD simulations of the preQ$_1$ riboswitch where a collapse occurred after ~1 μs (Banas, Sklenovsky, Wedekind, Sponer, & Otyepka, 2012). In the

18 MD simulations with ff99 + parmbsc0 of 200 ns length, two showed the sudden transition to a ladder-like structure after ~100 ns in the terminal P1 region. All trajectories with ladder-like structures were excluded from further analysis. Using ff10, which includes the parmχ_{OL3} refinement, neither of the 18 MD simulations showed a transition to a ladder-like structure.

3.3.2 Structural deviations of the aptamer domain

Data on the structural deviations of Gsw$^{\mathrm{apt}}$ and Gsw$^{\mathrm{loop}}$ with respect to the starting structures in terms of mean RMSD calculated over three independent simulations for each combination of force field (ff99, ff99 + parmbsc0, and ff10) and number of Mg^{2+} ions (0, 12, and 20) are shown in Table 2. The structural deviations are in general moderate ($\leq$3 Å for all but 4 out of the 18 cases, where the RMSD is up to 4 Å), which agrees with findings of MD simulations on related systems (Priyakumar & MacKerell,, 2010; Sharma, Bulusu, & Mitra, 2009; Villa, Wöhnert, & Stock, 2009), and experimental data according to which Gsw$^{\mathrm{apt}}$ with and without Mg^{2+}, and Gsw$^{\mathrm{loop}}$ with Mg^{2+}, have a stable tertiary structure (Buck et al., 2010). However, in simulations with ff99 and ff99 + parmbsc0, Gsw$^{\mathrm{loop}}$ shows in general larger RMSD values than Gsw$^{\mathrm{apt}}$, and this effect is most pronounced in the absence of Mg^{2+}. This result reflects the destabilized loop–loop interactions in Gsw$^{\mathrm{loop}}$ (Buck et al., 2010, 2011; see also below). In contrast, in simulations with ff10, Gsw$^{\mathrm{apt}}$, and Gsw$^{\mathrm{loop}}$ show very similar structural deviations at the respective Mg^{2+} concentrations.

3.3.3 Mobility of the aptamer domain

In order to investigate the influence of the force field on the dynamics of the RNA, we calculated atomic RMSF and averaged them per nucleotide (Fig. 4A–F). The general shapes of the curves obtained for each combination of force field and number of Mg^{2+} ions are very similar. The stem regions P2 and P3 are the least mobile, in contrast to high fluctuations observed for the aptamer terminal helix P1. Furthermore, the J2/3 region is found to be mobile, in agreement with SHAPE experiments (Stoddard et al., 2008) and the suggestion that the J2/3 region acts as an entry gate for the ligand to the binding site (Gilbert et al., 2006). Finally, pronounced mobility is observed in the L2 and L3 regions where tertiary interactions lead to the formation of two base quadruples.

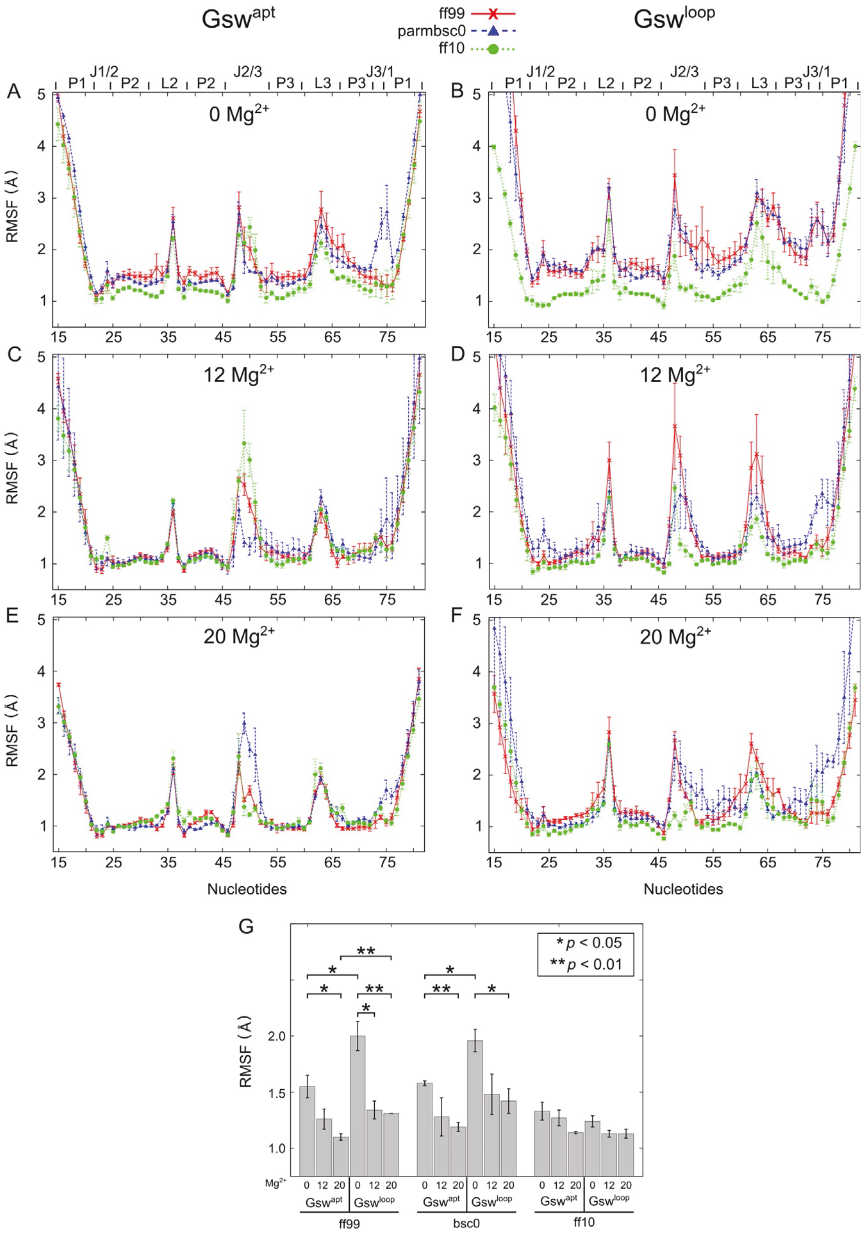


Figure 4 (A–F) Per-nucleotide RMSF obtained from MD simulations of Gsw^{apt} (A, C, and E) and Gsw^{loop} (B, D, and F) using different force fields and different numbers of Mg^{2+} ions; the mean ± SEM over three trajectories of 200 ns length is given. ff99: red (gray in the print version) solid line with "X" symbol; parmbsc0: blue (dark gray in the print version) dashed line with filled triangle; and ff10: green (light gray in the print version) dotted line with filled circle. For (A) and (B), absence of Mg^{2+} ions; (C) and (D), 12 Mg^{2+} ions per RNA molecule; and (E) and (F) 20 Mg^{2+} ions per RNA molecule. (G) Mean RMSF ± SEM of core nucleotides for the simulations shown in panels (A–F) grouped by the used force field. The stars indicate statistically significant differences between the mean values (*: $p < 0.05$; **: $p < 0.01$).

As to differences between the curves, qualitatively, MD simulations of Gsw^{apt} (Fig. 4A, C, and E) show in general lower RMSF of the aptamer domain than MD simulations of Gsw^{loop} (Fig. 4B, D, and F); within one system, the RMSF become the lower the more Mg^{2+} ions have been added. Considering mean RMSF values over the 80% of core nucleotides (see Section 2) of the aptamer domain, these are statistically significantly different in the case of ff99 and ff99 + parmbsc0 for Gsw^{apt} versus Gsw^{loop} without Mg^{2+}, as well as for Gsw^{apt} and Gsw^{loop} without versus with 20 Mg^{2+}, respectively (Fig. 4G). These findings match perfectly with experimental observations that Gsw^{loop} without Mg^{2+} does not show formation of tertiary interactions (Buck et al., 2010), with adverse effects on the structural stability of the aptamer, whereas Gsw^{apt} does; furthermore, addition of 20 Mg^{2+} results in a significant stabilization especially of the tertiary interactions in the case of Gsw^{apt} (Buck et al., 2010) or the formation of the tertiary interactions in the case of Gsw^{loop} (Buck et al., 2010).

Considering ff99 alone, two further differences are found significant (Fig. 4G): Between Gsw^{apt} and Gsw^{loop} with 20 Mg^{2+} ions and between Gsw^{loop} without and with 12 Mg^{2+} ions. The first case reflects that tertiary interactions in Gsw^{apt} are already formed without Mg^{2+} and are further stabilized upon Mg^{2+} addition (Buck et al., 2010); in contrast, in Gsw^{loop} the tertiary interactions start to form only at $[Mg^{2+}]:[RNA] > 18:1$ (Buck et al., 2010), suggesting that the Gsw^{loop} ensemble may still be more heterogeneous even in the presence of 20 Mg^{2+}. The second case agrees with the observation that addition of a few Mg^{2+} ions (ratio ~1:5–1:12, Buck et al., 2011) to Gsw^{loop} leads to only transiently formed tertiary interactions but an already ligand-binding competent aptamer ensemble (Buck et al., 2010), suggesting the beginning of structural compaction.

Likely the most unexpected observation is made for ff10 (Fig. 4G). First, MD simulations with ff10 result overall in the lowest mean RMSF when comparing to results for ff99 and ff99 + parmbsc0 and the different combinations of Gsw^{apt}/Gsw^{loop} and number of Mg^{2+} ions. Second, none of the differences between the simulated systems is significant in the case of ff10, in stark contrast to what can be expected with respect to experiment (see above). Thus, this suggests that MD simulations with ff10 lead to too strongly damped aptamer motions and/or to an overly stable aptamer domain.

In order to investigate to what extent the differences in the force fields contribute to the differences in the aptamer mobility, we computed frequency distributions of the torsion angles α, γ, and χ for Gsw^{loop} simulations in the absence of Mg^{2+}. As expected (Perez et al., 2007), ff99 + parmbsc0 and ff10 show much reduced populations of the g^+ region for α (Fig. 5A) and of the t region for γ (Fig. 5B) in comparison to ff99, and in turn higher populations of the g^- and g^+ regions, respectively. The locations of the respective distribution maxima remained largely unchanged (α: ff99: −75°; ff99 + parmbsc0: −80°; ff10: −75°), however. Likewise expected (Zgarbova et al., 2011), ff10 shows a much reduced population of the high-*anti* region of χ in comparison to ff99 and ff99 + parmbsc0 (Fig. 5C), which is a result of shifting the location of the distribution maximum by ~5° toward the *anti* region plus a narrowing of the distribution function of the main population. The population changes have a pronounced influence on the spread of the torsion angles, measured in terms of the circular variance (Eq. 1; Fig. 5D). Note that the circular variance is bounded in the interval [0,1], with 1 denoting that all torsion angle values are spread out evenly around a circle. Highly significant differences between circular variances are found for α and γ torsions between ff99 versus ff99 + parmbsc0 or ff10; similarly, highly significant differences are found for the χ torsion between ff99 or ff99 + parmbsc0 versus ff10 (Fig. 5D). These findings suggest that the modifications in the α and γ torsions are not the cause for the low-aptamer mobility observed with ff10, because the mean RMSF for Gsw^{loop} is not significantly ($p = 0.819$) different between ff99 and ff99 + parmsc0 (Fig. 4G). In contrast, a significantly reduced circular variance of the χ torsion is observed between those simulations of Gsw^{loop} that also show a significant reduction in the aptamer mobility (Fig. 4G; ff99 vs. ff10: $p = 0.005$; ff99 + parmbsc0 vs. ff10: $p = 0.003$).

Finally, we projected the per-nucleotide frequency with which a χ torsion is found in the high-*anti* region in MD simulations of Gsw^{loop} with ff99 and ff10 onto the aptamer domain (Fig. 5E and F). For some nucleotides, increased frequencies are found in both trajectories (U36, U47, U48, U61, G62, A64, and A65) with frequencies >80% for A64. Remarkably, most of the nucleotides with increased frequencies found in MD simulations with either force field also have a χ torsion in the high-*anti* region in one or both of the ligand bound Gsw^{apt} and Gsw^{loop} crystal structures (Fig. 5G; U47, G62, A64, and A65). The nucleotides with increased frequencies of their χ torsion in the high-*anti* region in MD simulations with either force field are found in those regions of the aptamer domain that are known to be the most mobile: The J2/3 region, which is suggested to act as an entry gate for the ligand to the binding site (Gilbert et al., 2006) and the loop regions (L2 and L3).

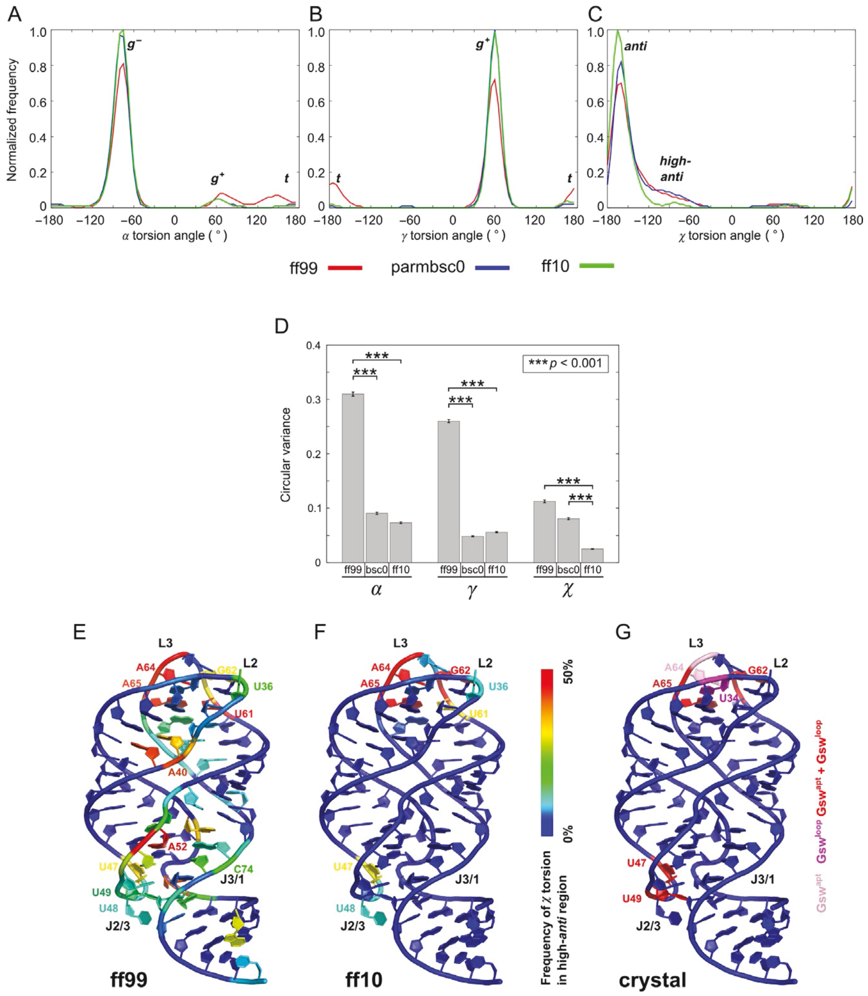


Figure 5 (A–C) Normalized histograms of torsion angles α (A), γ (B), and χ (C) in MD simulations of Gsw$^{\text{loop}}$ in the absence of Mg^{2+} ions using different force fields. Red (gray in the print version): ff99; blue (dark gray in the print version): parmbsc0; and green (light gray in the print version): ff10. g^+, g^-, and t regions are assigned according to Perez et al. (2007), the *anti* and high-*anti* regions are assigned according to Blackburn and Gait (1996). (D) Mean circular variance (Eq. 1) ±SEM in the torsion angles α, γ, and χ for the force fields ff99, parmbsc0 (bsc0), and ff10. The stars indicate statistically significant differences between the mean values (***: $p < 0.001$). (E and F) Per-nucleotide frequency of the χ torsion angle being in the high-*anti* region (between $-110°$ and $-60°$) for simulations using the ff99 (E) and the ff10 (F) force field. Frequency values between 0% (blue (dark gray in the print version)) and 50% (red (gray in the print version)) are mapped onto the structure of Gsw$^{\text{loop}}$ according to the color scale. Nucleotides A52 and A64 (G62, A64, and A65) showed frequencies >50% and up to 90% (90%) in the case of ff99 (ff10). (G) Occupation of the high-*anti* region of the χ torsion angle in the

(Continued)

Differences for nucleotides with increased frequencies of the χ torsion in the high-*anti* region between MD simulations with ff99 and ff10 are found in the extended J2/3 region around A52, in the J3/1 region around C74, which forms one side of the binding site and gets stabilized upon ligand binding, and to a lesser extent in one strand of the P3 stem adjacent to the very mobile L3 loop. Here, simulations with ff99 show frequencies of a χ torsion in the high-*anti* region of at most 25% (Fig. 5E). None of these regions show χ torsions in the high-*anti* state in the Gsw^{apt} and Gsw^{loop} crystal structures (Fig. 5G). Note, however, that the crystal structures represent ligand bound aptamer domains with stabilized binding sites, whereas in the MD simulations analyzed here without Mg^{2+} the aptamer domains are in the least stable states. In that respect, it appears unexpected that MD simulations with ff10 do agree so well with the crystal structures in that they show no χ torsions in the high-*anti* state in these regions either (Fig. 5F). Phrased differently, together with our above results, our analyses indicate that a small propensity for visiting high-*anti* states of χ torsions is required for appropriately modeling the dynamics of known mobile regions of the Gsw aptamer domain.

3.3.4 Loop–loop interactions

The occupancy of hydrogen bonds between bases of the two base quadruples in the L2/L3 region are reported in Table 3 for MD simulations of Gsw^{apt} and Gsw^{loop} in the absence of Mg^{2+} ions and using different force fields. The values are averages over all hydrogen bonds found between two bases. The force field dependence of the results is similar to what has been found for the force field influence on the mobility of the aptamer domain. First, for Gsw^{apt}, ff10 results in the highest occupancies, with almost permanent formation of hydrogen bonds between the bases of the upper (nucleotides 34, 37, 61, and 65; Fig. 1C and D) and lower (nucleotides 33, 38, 60, and 66; Fig. 1E and F) base quadruples, respectively. ff99+parmbsc0 shows equally high occupancies for the upper

Figure 5—Cont'd ligand bound crystal structures of Gsw^{apt} (PDB ID 4FE5, Stoddard et al., 2013) and Gsw^{loop} (PDB ID 3RKF, Buck et al., 2011). Nucleotides whose χ torsion angles are not in the high-*anti* region in either of the crystal structures are colored in blue (dark gray in the print version). The other residues are colored according to the crystal structure in which their χ torsion angle is in the high-*anti* region: Gsw^{apt} only (light pink (light gray in the print version)), Gsw^{loop} only (magenta (gray in the print version)), or both (red (dark gray in the print version)).

quadruple but ~10% to ~20% lower occupancies for the lower quadruple. The lowest occupancies are found with ff99, with values <90% for the upper quadruple and <70% for the lower quadruple. Note that in this case the SEM is as high as 12%; closer inspection revealed that one out of the three independent simulations then differs with respect to the other two, suggesting that longer simulation times will be required to achieve better converged results for the more mobile system. The exception to the above are hydrogen bonds between C61 and G37 as well as between G38 and C60, which have occupancies >97% irrespective of the force field; the two Watson–Crick base pairs have been described as the most crucial components of the tertiary interaction (Gilbert, Love, Edwards, & Batey, 2007).

Second, for Gsw^{loop}, a decrease in the occupancies with respect to Gsw^{apt} is found in general for each force field used. However, for ff10, only in the case of U61–A65 an occupancy below 60% is found, whereas such low occupancies are found for four and three base pairs in the case of ff99 + parmbsc0 and ff99, respectively. Moreover, only in the case of ff99 do we find significantly lower occupancies of ~70–77% for the two base pairs U61–A37 and G38–C60 that showed very persistent hydrogen bonds in Gsw^{apt} (see above). Finally, we find a destabilization of the lower base quadruple as a result of the mutation in the upper one only in the case of ff99 and ff99 + parmbsc0; in contrast, for ff10, we find that the occupancies of the lower base quadruple are rather comparable to the ones found for Gsw^{apt}.

The G37A/C61U mutation was introduced in Gsw^{loop} to disturb the native hydrogen bond network connecting the L2 and L3 loops (Noeske, Buck, et al., 2007). Our observations of a reduced stability of tertiary interactions in Gsw^{loop} for ff99 and ff99 + parmbsc0 are in line with this. Later, NMR experiments showed that the loop–loop interactions are not formed in the absence of Mg^{2+} ions in Gsw^{loop} (Buck et al., 2010), which is at variance with our nonzero occupancies. However, one needs to consider that we start from a Gsw^{loop} crystal structure that is almost indistinguishable in the L2/L3 region from Gsw^{apt} (Fig. 1C–F) and that our simulation times of 200 ns are still short with respect to movements of an RNA on a rugged energy landscape (Sponer et al., 2014). Finally, for Gsw^{apt}, it has been reported that, even if the tertiary interactions are already present in the absence of Mg^{2+}, addition of Mg^{2+} leads to a significant stabilization especially for the loop–loop interaction (Buck et al., 2010, 2011). Considering the hydrogen bond occupancies, it is only possible to observe such a stabilization when using ff99: hydrogen bond occupancies of the base quadruples in Gsw^{apt} raise to >90% in the presence of 20 Mg^{2+} then (except for A66 and G38, where the occupancies are low throughout all simulations).

3.4. Concluding remarks

In this study, we aimed at identifying conditions for MD simulations that allow to appropriately model experimentally demonstrated differences in the structural stability of Gsw^{apt} and Gsw^{loop}, including the pronounced influence of Mg^{2+} on the structural dynamics of the aptamer domain. The most outstanding result is the strong force field dependence of the structural dynamics, with the original Amber force field ff99 yielding the best agreement with experimental observations whereas the use of ff10, implementing the parmbsc0 and parmχ_{OL} modifications, results in strongly damped aptamer motions and overly stable tertiary loop–loop interactions; ff99 + parmbsc0 yields an intermediate behavior.

We investigated the aptamer domain of the Gsw from *B. subtilis* here. The function of riboswitches inherently depends on its ability to switch between the inactive and the active states, and it requires that the aptamer domain maintains ligand-binding ability also in the inactive state in order to allow for a quick response on the presence of ligands. Especially, transcriptionally acting riboswitches such as the one investigated here have only a limited timeframe to respond to ligand binding (Stoddard et al., 2010). In order to achieve this quick response, the aptamer domain must be marginally stable, such that it can either follow the inactive route if no ligand binds, or it can quickly bind a ligand and transmit the information to the expression platform. These properties, its intricate tertiary structure, yet sufficiently small size, and the wealth of available experimental information on structure, stability, and dynamics under various conditions (Batey, 2012) make this system ideally suited for such investigations.

Our conclusions are based on detailed investigations of differences in the mobility of nucleotides and occupancies of tertiary loop–loop interactions. Ideally, our simulation results should have been compared to more direct measures of RNA dynamics, e.g., from NMR relaxation measurements (Juneja, Villa, & Nilsson, 2014). Such information is not yet available, however, on a broad per-residue level for the unbound Gsw. Still, we believe that our conclusions are well grounded. First, we performed three independent simulations for each investigated system leading to an aggregate simulation time of >11 μs, which allows us to estimate the uncertainty in the simulation results; only for occupancies of tertiary interactions in the most mobile systems (Gsw^{apt} and Gsw^{loop} in the absence of Mg^{2+} and using ff99) did we find indications that our MD simulations have not yet converged. Second, we used thorough statistical analysis for evaluating the significance of differences in mean values. Third, we compared our results to detailed experimental information on the structural stability of the aptamer domain and the formation of the loop–loop interactions (Buck et al., 2010, 2011; Stoddard et al., 2008). Finally, and most importantly, we analyze *differences* in the characteristics of two related RNA systems (Gsw^{apt} and Gsw^{loop}) and under various Mg^{2+} concentrations, rather than absolute characteristics of different RNA systems. Our approach should thus inherently profit from error cancelation.

The analysis of the spread of torsions angles sampled by the three force fields revealed a significantly reduced circular variance of the χ torsion between those simulations of Gsw^{loop} that also show a significant reduction in the aptamer mobility. Together with the population analysis of the χ torsion, the reduced variance is a result of a ~5-fold lower population of the high-*anti* region in ff10 than in ff99 and ff99 + parmbsc0. This is a consequence of an increased slope of the χ torsion profile in the high-*anti* region in ff10 compared to ff99 and ff99 + parmbsc0 (Zgarbova et al., 2011), which was successfully introduced to suppress ladder-like RNA structures (Banas et al., 2012; Sklenovsky et al., 2011).

The study by Zgarbova et al. (2011) demonstrated already that overstabilization of the χ-*anti* region has adverse effects on the geometry of A-RNA. Compared with the original effort by Zgarbova et al. (2011) to suppress ladder-like RNA structures in long MD simulations, our results suggest that even the moderate stabilization of the χ-*anti* region in ff10 can have an unwanted damping effect on the functionally relevant structural dynamics of marginally stable RNA systems. Our suggestion is supported by the finding of a high coincidence of χ torsions with high-*anti* values in the Gsw^{apt} and Gsw^{loop} crystal structures in those regions that are known to be most mobile; in contrast, MD simulations with ff10 result in very low frequencies of occurrence of χ torsions with high-*anti* values in some of the mobile regions. From the perspective of future force field development, our results suggest to include systems such as the aptamer domain of the Gsw in such studies and to extent the objectives for force field optimization beyond criteria of structural closeness to the native structure or the stabilization of signature interactions seen in experimental structures (Banas et al., 2010; Perez et al., 2007; Sklenovsky et al., 2011; Zgarbova et al., 2011) toward agreement of (differences in) the structural dynamics with experiment.

Finally, from our investigations on the influence of the set up of Mg^{2+} ions on the structural dynamics of Gsw^{apt} we strongly recommend not to use the default procedure implemented in the *leap* program of the Amber suite for placing Mg^{2+} around a nucleic acid structure but rather to use *leap* to place larger dummy ions that are later replaced by hexahydrated Mg^{2+}. Regarding the use of different Mg^{2+} ion parameters, we observed only a small influence on the structural dynamics of Gsw^{apt}. Our observation of no inner-sphere contact formation in any of the MD simulations when using the recently developed Mg^{2+} parameters of Allner et al. was unexpected. However, further investigations including other RNA systems are required to confirm this finding.

ACKNOWLEDGMENTS

We gratefully acknowledge the computing time granted by the John von Neumann Institute for Computing (NIC) and provided on the supercomputer JUROPA at Jülich Supercomputing Center (JSC) (NIC project 4722). Additional computational support was provided by the "Center for Information and Media Technology" (ZIM) at the Heinrich-Heine-Universität Düsseldorf (Germany).

REFERENCES

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

Aqvist, J. (1990). Ion water interaction potentials derived from free-energy perturbation simulations. *Journal of Physical Chemistry*, *94*(21), 8021–8024.

Aqvist, J. (1992). Modeling of ion ligand interactions in solutions and biomolecules. *Journal of Molecular Structure: Theochem*, *88*, 135–152.

Auffinger, P. (2012). Ions in molecular dynamics simulations of RNA systems. In N. Leontis, & E. Westhof (Eds.), *RNA 3D structure analysis and prediction* (pp. 299–318). Berlin, Heidelberg: Springer.

Auffinger, P., Bielecki, L., & Westhof, E. (2003). The Mg^{2+} binding sites of the 5S rRNA loop E motif as investigated by molecular dynamics simulations. *Chemistry & Biology*, *10*(6), 551–561.

Banas, P., Hollas, D., Zgarbova, M., Jurecka, P., Orozco, M., Cheatham, T. E., III, et al. (2010). Performance of molecular mechanics force fields for RNA simulations: Stability of UUCG and GNRA hairpins. *Journal of Chemical Theory and Computation*, *6*(12), 3836–3849.

Banas, P., Sklenovsky, P., Wedekind, J. E., Sponer, J., & Otyepka, M. (2012). Molecular mechanism of $preQ_1$ riboswitch action: A molecular dynamics study. *Journal of Physical Chemistry B*, *116*(42), 12721–12734.

Batey, R. T. (2012). Structure and mechanism of purine-binding riboswitches. *Quarterly Reviews of Biophysics*, *45*(3), 345–381.

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

Besseova, I., Otyepka, M., Reblova, K., & Sponer, J. (2009). Dependence of A-RNA simulations on the choice of the force field and salt strength. *Physical Chemistry Chemical Physics*, *11*(45), 10701–10711.

Blackburn, G. M., & Gait, M. J. (1996). *Nucleic acids in chemistry and biology*. Oxford: Oxford University Press.

Buck, J., Noeske, J., Wöhnert, J., & Schwalbe, H. (2010). Dissecting the influence of Mg^{2+} on 3D architecture and ligand-binding of the guanine-sensing riboswitch aptamer domain. *Nucleic Acids Research*, *38*(12), 4143–4153.

Buck, J., Wacker, A., Warkentin, E., Wöhnert, J., Wirmer-Bartoschek, J., & Schwalbe, H. (2011). Influence of ground-state structure and Mg^{2+} binding on folding kinetics of the guanine-sensing riboswitch aptamer domain. *Nucleic Acids Research*, *39*(22), 9768–9778.

Case, D. A., Cheatham, T. E., III, Darden, T., Gohlke, H., Luo, R., Merz, K. M., et al. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, *26*(16), 1668–1688.

Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., et al. (2010). *AMBER 11*. San Francisco, CA: University of California.

Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., et al. (2012). *AMBER 13*. San Francisco, CA: University of California.

Cheatham, T. E., III, & Case, D. A. (2013). Twenty-five years of nucleic acid simulations. *Biopolymers*, *99*(12), 969–977.

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *Journal of Chemical Physics*, *98*(12), 10089–10092.

Dixon, N., Duncan, J. N., Geerlings, T., Dunstan, M. S., McCarthy, J. E. G., Leys, D., et al. (2010). Reengineering orthogonally selective riboswitches. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(7), 2830–2835.

Draper, D. E., Grilley, D., & Soto, A. M. (2005). Ions and RNA folding. *Annual Review of Biophysics and Biomolecular Structure*, *34*, 221–243.

Foloppe, N., & MacKerell, A. D., Jr. (2000). 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.

Gilbert, S. D., Love, C. E., Edwards, A. L., & Batey, R. T. (2007). Mutational analysis of the purine riboswitch aptamer domain. *Biochemistry*, *46*(46), 13297–13309.

Gilbert, S. D., Stoddard, C. D., Wise, S. J., & Batey, R. T. (2006). Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *Journal of Molecular Biology*, *359*(3), 754–768.

Gohlke, H., Kuhn, L. A., & Case, D. A. (2004). Change in protein flexibility upon complex formation: Analysis of Ras-Raf using molecular dynamics and a molecular framework approach. *Proteins: Structure, Function, and Bioinformatics*, *56*(2), 322–337.

Harned, H. S., & Hudson, R. M. (1951). The diffusion coefficient of magnesium sulfate in dilute aqueous solution at 25°. *Journal of the American Chemical Society*, *73*(12), 5880–5882.

Hayes, R. L., Noel, J. K., Mohanty, U., Whitford, P. C., Hennelly, S. P., Onuchic, J. N., et al. (2012). Magnesium fluctuations modulate RNA dynamics in the SAM-I riboswitch. *Journal of the American Chemical Society*, *134*(29), 12043–12053.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, *79*(2), 926–935.

Joung, I. S., & Cheatham, T. E., III. (2008). Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *Journal of Physical Chemistry B*, *112*(30), 9020–9041.

Juneja, A., Villa, A., & Nilsson, L. (2014). Elucidating the relation between internal motions and dihedral angles in an RNA hairpin using molecular dynamics. *Journal of Chemical Theory and Computation*, *10*(8), 3532–3540.

Li, P. F., Roberts, B. P., Chakravorty, D. K., & Merz, K. M. (2013). Rational design of particle mesh Ewald compatible Lennard-Jones parameters for +2 metal cations in explicit solvent. *Journal of Chemical Theory and Computation*, *9*(6), 2733–2748.

MacArthur, M. W., & Thornton, J. M. (1993). Conformational analysis of protein structures derived from NMR data. *Proteins*, *17*(3), 232–251.

MacKerell, A. D., Jr., & Banavali, N. K. (2000). 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.

Mills, R., & Lobo, V. M. M. (1989). *Self-diffusion in electrolyte solutions: A critical examination of data compiled from the literature*. Amsterdam: Elsevier.

Mlynsky, V., Banas, P., Hollas, D., Reblova, K., Walter, N. G., Sponer, J., et al. (2010). Extensive molecular dynamics simulations showing that canonical G8 and protonated A38H$^+$ forms are most consistent with crystal structures of hairpin ribozyme. *Journal of Physical Chemistry B*, *114*(19), 6642–6652.

Moller, K. B., Rey, R., Masia, M., & Hynes, J. T. (2005). On the coupling between molecular diffusion and solvation shell exchange. *Journal of Chemical Physics*, *122*(11), 114508.

Noeske, J., Buck, J., Fürtig, B., Nasiri, H. R., Schwalbe, H., & Wöhnert, J. (2007). Interplay of 'induced fit' and preorganization in the ligand induced folding of the aptamer domain of the guanine binding riboswitch. *Nucleic Acids Research*, *35*(2), 572–583.

Noeske, J., Schwalbe, H., & Wöhnert, J. (2007). Metal-ion binding and metal-ion induced folding of the adenine-sensing riboswitch aptamer domain. *Nucleic Acids Research*, *35*(15), 5262–5273.

Ohtaki, H., & Radnai, T. (1993). Structure and dynamics of hydrated ions. *Chemical Reviews*, *93*(3), 1157–1204.

Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., III, Laughton, C. A., et al. (2007). Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophysical Journal*, *92*(11), 3817–3829.

Priyakumar, U., & MacKerell, A. D., Jr. (2010). Role of the adenine ligand on the stabilization of the secondary and tertiary interactions in the adenine riboswitch. *Journal of Molecular Biology*, *396*(5), 1422–1438.

Pyle, A. M. (2002). Metal ions in the structure and function of RNA. *Journal of Biological Inorganic Chemistry*, 7(7–8), 679–690.

Reblova, K., Lankas, F., Razga, F., Krasovska, M. V., Koca, J., & Sponer, J. (2006). Structure, dynamics, and elasticity of free 16S rRNA helix 44 studied by molecular dynamics simulations. *Biopolymers*, *82*(5), 504–520.

Reblova, K., Spackova, N., Stefl, R., Csaszar, K., Koca, J., Leontis, N. B., et al. (2003). Non-Watson–Crick base pairing and hydration in RNA motifs: Molecular dynamics of 5S rRNA loop E. *Biophysical Journal*, *84*(6), 3564–3582.

Robinson, H., Gao, Y. G., Sanishvili, R., Joachimiak, A., & Wang, A. H. J. (2000). Hexahydrated magnesium ions bind in the deep major groove and at the outer mouth of A-form nucleic acid duplexes. *Nucleic Acids Research*, *28*(8), 1760–1766.

Roe, D. R., & Cheatham, T. E., III. (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, *9*(7), 3084–3095.

Ryckaert, J. P., Ciccotti, G., & Berendsen, H. J. C. (1977). Numerical integration of cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *Journal of Computational Physics*, *23*(3), 327–341.

Saini, J. S., Homeyer, N., Fulle, S., & Gohlke, H. (2013). Determinants of the species selectivity of oxazolidinone antibiotics targeting the large ribosomal subunit. *Biological Chemistry*, *394*(11), 1529–1541.

Salomon-Ferrer, R., Gotz, A. W., Poole, D., Le Grand, S., & Walker, R. C. (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *Journal of Chemical Theory and Computation*, *9*(9), 3878–3888.

Schafmeister, C. E. A. F., Ross, W. S., & Romanovski, V. (1995). *Leap*. San Francisco, CA: University of California.

Serganov, A., & Nudler, E. (2013). A decade of riboswitches. *Cell*, *152*(1–2), 17–24.

Serganov, A., Yuan, Y. R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A. T., et al. (2004). Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chemistry & Biology*, *11*(12), 1729–1741.

Sharma, M., Bulusu, G., & Mitra, A. (2009). 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.

Sklenovsky, P., Florova, P., Banas, P., Reblova, K., Lankas, F., Otyepka, M., et al. (2011). Understanding RNA flexibility using explicit solvent simulations: The ribosomal and group I intron reverse kink-turn motifs. *Journal of Chemical Theory and Computation*, *7*(9), 2963–2980.

Snedecor, G. W., & Cochran, W. G. (1989). *Statistical methods* (8th ed.). Ames: Iowa State University Press.

Sponer, J., Banas, P., Jurecka, P., Zgarbova, M., Kuhrova, P., Havrila, M., et al. (2014). Molecular dynamics simulations of nucleic acids. From tetranucleotides to the ribosome. *Journal of Physical Chemistry Letters*, *5*(10), 1771–1782.

Sponer, J., Cang, X. H., & Cheatham, T. E., III. (2012). Molecular dynamics simulations of G-DNA and perspectives on the simulation of nucleic acid structures. *Methods*, *57*(1), 25–39.

Stoddard, C. D., Gilbert, S. D., & Batey, R. T. (2008). Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA*, *14*(4), 675–684.

Stoddard, C. D., Montange, R. K., Hennelly, S. P., Rambo, R. P., Sanbonmatsu, K. Y., & Batey, R. T. (2010). Free state conformational sampling of the SAM-I riboswitch aptamer domain. *Structure*, *18*(7), 787–797.

Stoddard, C. D., Widmann, J., Trausch, J. J., Marcano-Velazquez, J. G., Knight, R., & Batey, R. T. (2013). Nucleotides adjacent to the ligand-binding pocket are linked to activity tuning in the purine riboswitch. *Journal of Molecular Biology*, *425*(10), 1596–1611.

Tucker, B. J., & Breaker, R. R. (2005). Riboswitches as versatile gene control elements. *Current Opinion in Structural Biology*, *15*(3), 342–348.

Villa, A., Wöhnert, J., & Stock, G. (2009). Molecular dynamics simulation study of the binding of purine bases to the aptamer domain of the guanine sensing riboswitch. *Nucleic Acids Research*, *37*(14), 4774–4786.

Wang, J. M., Cieplak, P., & Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, *21*(12), 1049–1074.

Welch, B. L. (1947). The generalization of students problem when several different population variances are involved. *Biometrika*, *34*(1–2), 28–35.

Woodson, S. A. (2005). Metal ions and RNA folding: A highly charged topic with a dynamic future. *Current Opinion in Chemical Biology*, *9*(2), 104–109.

Zgarbova, M., Otyepka, M., Sponer, J., Mladek, A., Banas, P., Cheatham, T. E., III, et al. (2011). Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *Journal of Chemical Theory and Computation*, 7(9), 2886–2902.

9.3 Publication II - Ligand-mediated and tertiary interactions cooperatively stabilize the P1 region in the guanine-sensing riboswitch

Hanke C.A., Gohlke, H.

PLOS ONE, 2017. **12**(6): p. e0179271.

PLOS | ONE

RESEARCH ARTICLE

Ligand-mediated and tertiary interactions cooperatively stabilize the P1 region in the guanine-sensing riboswitch

Christian A. Hanke, Holger Gohlke*

Mathematisch-Naturwissenschaftliche Fakultät, Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

* gohlke@uni-duesseldorf.de

Abstract

Riboswitches are genetic regulatory elements that control gene expression depending on ligand binding. The guanine-sensing riboswitch (Gsw) binds ligands at a three-way junction formed by paired regions P1, P2, and P3. Loops L2 and L3 cap the P2 and P3 helices and form tertiary interactions. Part of P1 belongs to the switching sequence dictating the fate of the mRNA. Previous studies revealed an intricate relationship between ligand binding and presence of the tertiary interactions, and between ligand binding and influence on the P1 region. However, no information is available on the interplay *among* these three main regions in Gsw. Here we show that stabilization of the L2-L3 region by tertiary interactions, and the ligand binding site by ligand binding, cooperatively influences the structural stability of terminal base pairs in the P1 region in the presence of Mg^{2+} ions. The results are based on molecular dynamics simulations with an aggregate simulation time of ~10 μs across multiple systems of the unbound state of the Gsw aptamer and a G37A/C61U mutant, and rigidity analyses. The results could explain why the three-way junction is a central structural element also in other riboswitches and how the cooperative effect could become contextual with respect to intracellular Mg^{2+} concentration. The results suggest that the transmission of allosteric information to P1 can be entropy-dominated.

Introduction

Riboswitches are genetic regulatory elements that regulate gene expression on the RNA level. Most riboswitches are found in the 5'-untranslated region of bacterial mRNA [1,2]. Riboswitches typically consist of an aptamer domain and an expression platform. The former binds ligands with high specificity [3]; the latter undergoes a conformational change upon binding of a ligand in the aptamer domain. Two prevalent modes of gene regulation by riboswitches were found in bacteria: translational and transcriptional gene regulation [4]. In transcriptionally acting riboswitches, the expression platform is involved in the formation of an intrinsic transcription terminator or antiterminator [5]. For this type of gene regulation to be effective, a decision at a branch point during transcription has to be made in favor of one of these two folding

OPEN ACCESS

Citation: Hanke CA, Gohlke H (2017) Ligand-mediated and tertiary interactions cooperatively stabilize the P1 region in the guanine-sensing riboswitch. PLoS ONE 12(6): e0179271. https://doi.org/10.1371/journal.pone.0179271

Editor: Paolo Carloni, Computational Biophysics, GERMANY

Received: December 18, 2016

Accepted: May 27, 2017

Published: June 22, 2017

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by institutional funds of Heinrich Heine University Düsseldorf. We gratefully acknowledge the computing time granted by the John von Neumann Institute for Computing (NIC) and provided on the supercomputer JUROPA at Jülich Supercomputing Center (JSC) (NIC project 4722, HDD08). Additional computational support was provided by the "Center for Information and Media Technology" (ZIM) at the

Heinrich-Heine-University of Düsseldorf (Germany). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

pathways [6,7]. Therefore, the unbound state has to maintain ligand-binding competence, but at the same time has to be able to follow the alternative pathway in the absence of the ligand [6].

The guanine-sensing riboswitch (Gsw) from the *xpt-pbuX* operon of *Bacillus subtilis* investigated here belongs to the class of purine binding riboswitches, binds guanine and hypoxanthine and acts transcriptionally [8,9]. The aptamer domain of this riboswitch consists of three paired regions P1, P2, and P3, two loops L2 and L3 capping the P2 and P3 helices, respectively, and three joining regions J1/2, J2/3 and J3/1 connecting the paired regions [8] (Fig 1A). Part of the P1 region belongs to the switching sequence, which is involved in the formation of the alternative structural element as a consequence of ligand binding [10]. Tertiary interactions between bases from L2 and L3 involve two base quadruples, forming a hydrogen bond network and keeping the P2 and P3 helices in close contact (Fig 1B and 1C). This arrangement of P2 and P3 together with P1 forms a three-way junction around J1/2, J2/3, and J3/1, which is a recurrent structural element in several RNA structures, including other riboswitches [11–13]. Ligands bind to Gsw's aptamer domain via hydrogen bond interactions with the joining regions J1/2, J2/3 and J3/1. C74 from the J3/1 region is responsible for ligand selectivity of this riboswitch [9]. The ligand is deeply buried in the binding site, which indicates that local flexibility is required for the ligand to bind [8]. J2/3, due to its comparatively high flexibility in the unbound state, is assumed to allow access of the ligand to the binding site [9,14,15]. In order to influence the gene regulation, the information about ligand binding has to be transmitted to the switching sequence in the P1 region. This communication may occur by base stacking interactions of the ligand with nucleotides of the P1 region [8] or by the stabilization of base triples between nucleotides of the J2/3 and P1 regions [16].

The tertiary L2-L3 interactions were found to play an important supporting role in facilitating ligand binding [16], especially of lower-affinity ligands such as hypoxanthine [8,19]. The replacement of the loops by stable UUCG tetraloops [8] resulted in a mutant that did not bind hypoxanthine. Although mutations outside the base quadruples were found to not significantly affect ligand binding [20], mutations in the base quadruples are less tolerated [20]. A G37A mutation resulted in an aptamer which did not bind hypoxanthine but was still able to bind guanine [16]. The additional mutation of C61U yielded a mutant G37A/C61U in which two of the seven hydrogen bonds in the respective wildtype base quadruple are missing. This double mutant showed a Mg^{2+} dependence of the ability to bind hypoxanthine [15]. It only regains the ligand binding ability in the presence of Mg^{2+} ions at a $[Mg^{2+}]$:[RNA] ratio $> 18:1$ [15,18]. This makes the double mutant an interesting test system where the stability of the tertiary interactions, and its influence on ligand binding, can be fine-tuned via the Mg^{2+} concentration. In the opposite direction, time-resolved NMR investigations using a photo-caged ligand showed that ligand binding stabilizes the L2-L3 interactions [14]. Furthermore, nonequilibrium MD simulations of temperature-induced energy flow in the Gsw reported a communication between the nucleotides at the ligand binding site and the distant L2-L3 region [21]. Additionally, umbrella sampling simulations using the distance between the loops as a reaction coordinate were applied to investigate the influence of the ligand on the formation of the loop-loop interactions in an adenine-sensing riboswitch [22,23]. Similarly, ligand-induced stabilization of the P1 region was investigated using steered molecular dynamics simulations, where the strands of the P1 region were separated by force [24]. An NMR study on an adenine-sensing riboswitch observed a coupling of the stability of the P1 region and the tertiary interactions with the ligand binding ability [25]. However, this effect was suggested to be smaller for the more stable guanine-sensing riboswitch aptamer [25].

Detailed structural information of the ligand-bound state of riboswitch aptamer domains is provided by crystal structures, including those of Gsw [8,26], and other experimental and theoretical investigations (reviewed in refs. [10,11,27,28]). In contrast, the unbound state is by far

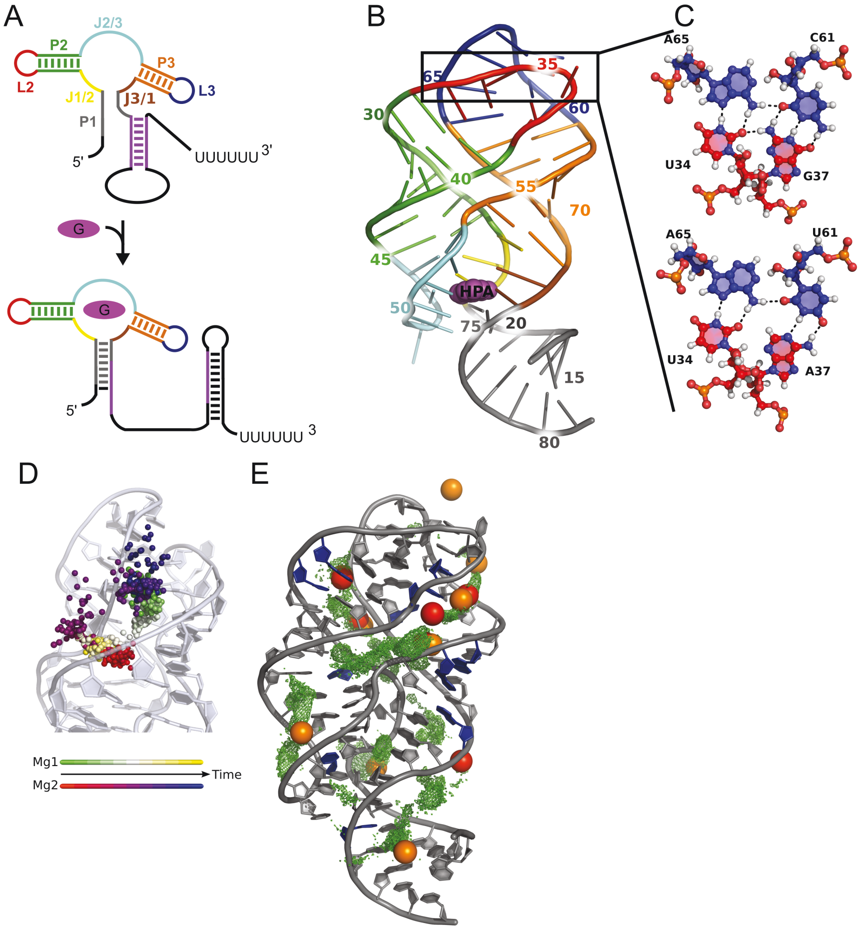


Fig 1. Structural features of the guanine-sensing riboswitch aptamer domain and behavior of Mg^{2+} ions. A: Schematic view of transcriptional regulation by the guanine-sensing riboswitch. In the unbound state (top), the switching sequence (purple) is involved in the formation of the anti-terminator. In the guanine (purple oval) bound state (bottom), the P1 region is stabilized, and part of the switching sequence is involved in the formation of the transcription terminator loop. Regions are assigned according to ref. [8]; grey: P1, green: P2, orange: P3, red: L2, blue: L3, yellow: J1/2, cyan: J2/3, brown: J3/1. B: Tertiary structure of the guanine-sensing riboswitch bound to hypoxanthine (HPA in purple) (PDB ID 4FE5) colored according to secondary structure elements as in panel A; the box marks the mutation site. C: Difference between Gsw^{apt} (top) and Gsw^{loop} (bottom) in the loop region. The G37A/C61U mutation results in a loss of two hydrogen bonds. Bases are colored as in panels A and B according to which loop they belong to. D: Exchange of two Mg^{2+} ions over a time of 8 ns. The positions of the two Mg^{2+} ions are shown as spheres and colored according to time with two different color scales. E: Comparison of preferred sites of occupancy of Mg^{2+} ions during 550 ns of MD simulations (green) to experimentally determined ion binding sites (red/orange: binding sites of [Co(NH$_3$)$_6$]$^{3+}$ ions in X-ray structures with PDB ID 4FE5 [17]/3RKF [18]) and nucleotides showing chemical shift changes upon magnesium titration in NMR experiments (blue) [18].

https://doi.org/10.1371/journal.pone.0179271.g001

less well characterized, especially at the atomistic level. Although crystal structures of unbound riboswitch aptamer domains other than from Gsw have been resolved [29], they only provide static information for what is considered a highly dynamic state [6,10,30]. In the case of Gsw, NMR investigations, in-line probing, and SHAPE experiments showed that the global fold of the unbound state of the aptamer domain is likely very similar to the bound state and is usually described as largely preorganized but locally disordered, with the P1, P2, and P3 regions preformed [31,32]; NMR studies revealed that this also holds even in the absence of Mg^{2+} ions [31]. In contrast, the ligand binding core, formed by J1/2, J2/3, and J3/1, is generally assumed to be disorganized [9,16,31,32]. The tertiary L2-L3 interactions are typically described as preformed in the unbound state even in the absence of Mg^{2+} ions [11,18,31]. However, the existence of a population of structures lacking the tertiary interactions is discussed [19,32]. A FRET study proposed that in the absence of Mg^{2+} and ligand the aptamer domain contains a highly dynamic P2 helix that moves with respect to the P3 helix, thereby allowing to form only transient L2-L3 interactions [19]. In the presence of Mg^{2+}, however, the studies agree that Gsw's L2-L3 interactions are stably formed [10]. For the G37A/C61U mutant, the loop-loop interactions were found to be formed only at a $[Mg^{2+}]$:[RNA] ratio $> 18:1$ [15,18].

From the above, an intricate picture emerges on the relationship between ligand binding to the aptamer domain of Gsw and presence of tertiary L2-L3 interactions, as well as between ligand binding and influence on the P1 region. However, to the best of our knowledge, no atom-level information is available on the interplay *among* these three main regions in Gsw, the L2-L3 region, the ligand binding core, and the P1 region.

Here, we investigate the interplay among these three regions by molecular dynamics (MD) simulations with an aggregate simulation time of ~10 μs of two variants of Gsw's aptamer domain: the wild type (Gsw^{apt}) and the G37A/C61U mutant (Gsw^{loop}). Based on a previous study by us [33], we pay particular attention to the force field influence and the influence of modeling Mg^{2+} on the structural dynamics of these RNAs. Finally, we perform rigidity analyses on the MD ensembles, which reveals that L2-L3 and ligand-mediated interactions cooperatively stabilize the terminal nucleotides in the P1 region, even in the absence of conformational changes in the binding site. We then discuss how our functional insights may provide an explanation as to why the three-way junction is a central structural element also in riboswitches, how the cooperative stabilization could become contextual with respect to the intracellular Mg^{2+} concentration, and what significance our findings have for the full aptamer's switching function.

Results

For unbound Gsw^{apt} and Gsw^{loop} with 0, 12, and 20 Mg^{2+} ions in the simulation box, three independent MD simulations of 550 ns length each were performed. Unless otherwise noted, results of these 18 MD simulations are reported below.

Mg^{2+} ions show high initial mobility and occupy binding sites in good agreement with experimentally determined ones

Mg^{2+} ions have a strong influence on structure and stability of RNA [34,35]. However, modeling these ions in MD simulations is considered difficult [36] because Mg^{2+} ions show a slow exchange kinetics of first shell ligands [37] and display very slow diffusion rates [38,39]. Hence, given the length of our MD simulations of 550 ns, it was important to carefully place the Mg^{2+} ions in the simulation box to prevent poor sampling. To account for this, we placed the Mg^{2+} ions as hexahydrated complexes at electrostatically favorable positions along the RNA. For details about this placement procedure and its advantages over placing non-hydrated Mg^{2+}, see ref. [33].

As a result, more than half of the Mg^{2+} ions showed a pronounced mobility during the first ~100 ns of the MD simulations (see Figure 2 in ref. [33]), and several ions remained mobile even after 500 ns. Additionally, we observed that several of the ions explore a large region around the aptamer in the MD simulations (shown exemplarily for three ions in S1 Fig). Furthermore, we occasionally observed that (temporarily) immobile hexahydrated Mg^{2+} ions swapped their positions during the MD simulations (Fig 1D). For other immobile Mg^{2+} ions, we observed that the ions lost water molecules from their first hydration shell and started to interact with the RNA (S2 Fig), despite the slow exchange kinetics. These observations are consistent with the known bias towards inner-shell binding of Mg^{2+} ions in RNA simulations using a non-polarizable force field [36,40,41]. S2 Fig also reveals that for both Gsw^{apt} and Gsw^{loop} with 12 Mg^{2+} and Gsw^{apt} with 20 Mg^{2+} the decrease in the number of hexahydrated Mg^{2+} is much slower in the second trajectory halves than the first, which suggests that these systems started to settle with respect to the binding of Mg^{2+} ions.

In order to investigate preferred occupation sites of Mg^{2+} ions in the MD simulations independent of their starting positions, an additional MD simulation was set up in which the Mg^{2+} ions were initially placed at least 10 Å from the RNA. As a result, preferred occupation sites between the RNA backbones were observed after 300 ns of MD simulations that are in very good agreement with binding sites of $[Co(NH_3)_6]^{3+}$ in crystal structures of Gsw aptamers [8,18] (Fig 1E): Four out of six experimental binding sites in PDB ID 4FE5 were occupied between 25% to 74% of the simulation time of 550 ns; two additional binding sites from PDB ID 3RKF were occupied ~57% of the simulation time (S1 Table). Similar occupation rates were observed in the 18 MD simulations for which the Mg^{2+} ions had been placed at electrostatically favorable positions along the RNA (S1 Table). Likewise, nucleotides that show chemical shift changes upon Mg^{2+} titration to Gsw^{loop} in NMR experiments [18] were found in a distance < 5 Å to the preferred occupation sites of Mg^{2+} ions in our MD simulation (Fig 1E). Note that one cannot expect a perfect agreement of computed occupation sites with experimentally determined ones. On the one hand, this is partly due to the conformational dynamics of the RNA during the MD simulations, which can lead to a shift of an occupation site with respect to the (static) crystal structures. On the other hand, the occupation of sites in the two crystal structures already differs quantitatively (shift between red and orange spheres in Fig 1E) and qualitatively (occupation in one crystal structure but not the other; Fig 1E). Yet, when comparing the occupancies (S1 Table) in terms of ratios of mole fractions of an occupied *versus* unoccupied binding site between different MD trajectories (e.g., for binding site 1, Gsw^{apt} 12 Mg^{2+}, and trajectories 1 and 2: (10/90) / (27 / 63)), for the overwhelming majority of the 96 cases (4 systems x 8 binding sites x 3 ratios), this ratio is close to one order of magnitude, and in only four cases it exceeds two orders of magnitude (the 16 cases where the occupancy was either 0 or 100% were not considered here); in our view, this observation displays an encouragingly high internal consistency of occupancies between different trajectories.

In summary, while these results do not prove convergence of the occupation of Mg^{2+} sites, they suggest that the respective systems sufficiently settled, and in a consistent manner, with respect to Mg^{2+} ion binding such that they can be compared to each other for investigating the influence of the concentration of Mg^{2+} ions on the structure and dynamics of Gsw^{apt} and Gsw^{loop}.

The aptamer domain remains globally stable during the MD simulations but shows local structural changes

Data on structural deviations of Gsw^{apt} and Gsw^{loop} are given in Table 1 in terms of mean root mean-square deviations (RMSD) over three trajectories after root mean-square fitting of the

Table 1. Root mean square deviations of Gsw^{apt} and Gsw^{loop} as a whole and for substructures[a].

Simulated system		Aptamer [b]	P1[c]	P2[c]	P3[c]	L2[c]	L3[c]
Gsw^{apt}	0 Mg^{2+}	3.0 ± 0.1	7.6 ± 0.5	2.4 ± 0.2	2.8 ± 0.2	3.1 ± 0.3	3.9 ± 0.3
	12 Mg^{2+}	2.9 ± 0.1	5.3 ± 0.4	2.0 ± 0.1	2.5 ± 0.1	2.1 ± 0.2	3.5 ± 0.2
	20 Mg^{2+}	2.3 ± 0.1	4.2 ± 0.3	1.6 ± 0.1	1.7 ± 0.1	2.1 ± 0.1	2.9 ± 0.2
Gsw^{loop}	0 Mg^{2+}	4.2 ± 0.2	9.9 ± 1.1	3.0 ± 0.2	3.7 ± 0.2	4.3 ± 0.3	5.4 ± 0.4
	12 Mg^{2+}	2.4 ± 0.1	4.6 ± 0.5	1.9 ± 0.1	1.9 ± 0.1	3.3 ± 0.2	3.8 ± 0.3
	20 Mg^{2+}	2.5 ± 0.1	4.4 ± 0.3	1.8 ± 0.1	2.1 ± 0.1	2.9 ± 0.2	3.6 ± 0.2
Simulated system		**J1/2[c]**	**J2/3[c]**	**J3/1[c]**			
Gsw^{apt}	0 Mg^{2+}	2.3 ± 0.1	3.4 ± 0.2	2.7 ± 0.2			
	12 Mg^{2+}	2.4 ± 0.1	4.6 ± 0.2	3.2 ± 0.2			
	20 Mg^{2+}	2.0 ± 0.1	3.8 ± 0.3	2.6 ± 0.2			
Gsw^{loop}	0 Mg^{2+}	3.3 ± 0.3	5.6 ± 0.3	5.2 ± 0.5			
	12 Mg^{2+}	1.9 ± 0.1	4.1 ± 0.2	2.3 ± 0.2			
	20 Mg^{2+}	2.0 ± 0.1	3.4 ± 0.2	2.8 ± 0.1			

[a] In Å; given is the mean ± SEM calculated over the three trajectories; the first 50 ns of each trajectory were omitted for the calculation. The RMSD was calculated after root mean-square fitting of the conformations on the initial conformation, considering for fitting only those 80% of the nucleotides that show the lowest RMSF (S2 Table).

[b] The complete RNA (nucleotides 15–81).

[c] RMSD of the respective substructure.

https://doi.org/10.1371/journal.pone.0179271.t001

conformations on the initial conformation, considering for fitting only those 80% of the nucleotides that show the lowest RMSF (S2 Table). All simulations show mean RMSD ≤ 3 Å for the whole aptamer structures (also exemplarily shown in S3 Fig as a function of simulation time for Gsw^{apt}), except for Gsw^{loop} in the absence of Mg^{2+} ions (4.2 Å). RMSD values of up to 3–4 Å have also been observed in other MD simulations of small riboswitches applying a similar simulation setup [42–44]. Accordingly, the comparison of the average conformation over the last 50 ns of MD simulations with the crystal structure shows only small overall structural deviations (S4 Fig). The ff99 force field used here has been reported to lead to irreversible transitions to unexpectedly open and under-twisted RNA structures resembling a ladder due to *anti* to high-*anti* shifts of the χ dihedral [43,45–47]. However, visual inspection of our MD trajectories neither revealed the formation of such ladder-like structures for Gsw^{apt} nor for Gsw^{loop} (S4 Fig). Furthermore, histograms of χ dihedrals using the ff99 force field on Gsw^{apt} and Gsw^{loop} in the presence of 20 Mg^{2+} ions are highly similar to those obtained for the ff10 force field, including the high-*anti* region (S5 Fig for the entire aptamer and S6 Fig for the P1 region only). We also note that some nucleotides in the bound crystal structures of Gsw^{apt} and Gsw^{loop} do occupy the χ high-*anti* region [33].

Mean RMSD for substructures of Gsw^{apt} and Gsw^{loop} were calculated after root mean-square fitting of those nucleotides showing the 80% smallest RMSF ("core nucleotides") (Table 1, S2 Table). Together with small RMSD values observed for all substructures but J2/3 when fitting the respective substructure onto itself (S3 Table), these RMSD values thus mostly report on relative motions of the substructure with respect to the core nucleotides. The largest values were observed for the P1 region, up to 6 Å for MD simulations in the presence of Mg^{2+} and 7.6 Å (9.9 Å) for Gsw^{apt} (Gsw^{loop}) in the absence of Mg^{2+}. Note, however, that these RMSD values result in part from the high mobility of the three terminal base pairs of P1. The P2 and P3 regions show overall the lowest mean RMSD values (≤ 2.8 Å, except for Gsw^{loop} in the absence of Mg^{2+} (3.0–3.7 Å)), as expected for paired regions in the center of the RNA. The larger structural deviations observed in the absence of Mg^{2+} result from relative motions of the P2 and P3 regions with respect to each other due to the lack of charge compensation by the ions. Regarding loops L2 and L3 of Gsw^{apt}, a similar trend is observed with mean RMSD values that are higher by about 1 Å in the absence of Mg^{2+} than in its presence; L2 and L3 of Gsw^{loop} show mean RMSD values that are yet higher by 1 Å than those of Gsw^{apt}. Finally, of the junction regions, J1/2 shows the lowest mean RMSD (~2 Å) in most simulations; the largest value (3.3 Å) is observed for Gsw^{loop} in the absence of Mg^{2+} ions. In contrast, the J2/3 region shows high mean RMSD values > 3 Å for all simulations as does the J3/1 region (~3 Å). Again, the only exception is Gsw^{loop} in the absence of Mg^{2+} ions where we observe mean RMSD values > 5 Å.

In summary, no gross conformational changes of the aptamer domain were observed during our MD simulations. On a local scale, the mutation in Gsw^{loop} induces local conformational rearrangements in the loop region as well as at the ligand binding site (J1/2, J2/3 and J3/1), manifested in higher RMSD values compared to the simulations of Gsw^{apt}.

The G37A/C61U mutation disturbs the native hydrogen bond network in the loop region

The tertiary interactions between the loops L2 and L3 are formed via a hydrogen bond network composed of two base quadruples. Each quadruple is formed by a canonical Watson-Crick base pair and a non-canonical base pair connected to it. The G37A/C61U mutation is located in the upper base quadruple, which is formed by the Watson-Crick base pair of nucleotides 37 and 61 and the two nucleotides 34 and 65 (Fig 1C). The lower base quadruple is identical in both Gsw variants and consists of the Watson-Crick base pair formed by nucleotides 38 and 60 and the two nucleotides 33 and 66 (S7 Fig).

We calculated average base pair occupancies for native base pairs in the loops L2 and L3 from the occupancies of the hydrogen bonds that are involved in the formation of the base pair (Table 2). For Gsw^{apt} in the absence of Mg^{2+} ions, almost all base pairs, except for

Table 2. Hydrogen bond occupancy in the L2/L3 loop region of Gsw^{apt} and Gsw^{loop} [a].

Gsw variant	Nucleotide pair	0 Mg^{2+} [b]	12 Mg^{2+} [b]	20 Mg^{2+} [b]
Gsw^{apt}	U34—G37	69.9 ± 6.6	92.8 ± 3.5	78.7 ± 4.8
	U34—A65	68.4 ± 4.2	94.6 ± 1.0	95.8 ± 1.0
	G37—C61	99.1 ± 0.2	99.8 ± 0.1	95.0 ± 2.7
	C61—A65	58.4 ± 7.5	97.1 ± 1.7	93.7 ± 3.9
	A33—A66	52.2 ± 5.8	96.9 ± 0.6	89.5 ± 2.6
	G38—C60	98.0 ± 0.3	96.6 ± 1.1	90.3 ± 3.4
	G38—A66	45.6 ± 5.1	79.2 ± 1.7	61.3 ± 3.2
Gsw^{loop}	U34—A37	-[c]	-[c]	-[c]
	U34—A65	54.6 ± 6.5	81.3 ± 3.5	68.7 ± 6.0
	A37—U61	86.4 ± 4.3	76.2 ± 5.3	66.9 ± 5.1
	U61—A65	4.9 ± 4.4	30.6 ± 7.3	24.1 ± 6.8
	A33—A66	28.2 ± 7.6	52.5 ± 5.6	53.6 ± 5.5
	G38—C60	70.1 ± 2.4	99.5 ± 0.1	70.7 ± 2.7
	G38—A66	22.3 ± 4.9	46.0 ± 4.5	64.4 ± 4.5

[a] In %; mean ± SEM over three trajectories of 550 ns length. The nucleotides 34, 37, 61, and 65 belong to the upper base quadruple where the G37A/C61U mutation is located; the nucleotides 33, 38, 60, and 66 form the lower base quadruple. The average value was calculated over the hydrogen bonds connecting the two nucleotides. The SEM was calculated by error propagation.

[b] Number of Mg^{2+} per RNA molecule.

[c] No base pairing for Gsw^{loop}.

https://doi.org/10.1371/journal.pone.0179271.t002

A66-G38 from the lower base quadruple, are formed for > 50% of the time. The most stable base pairs are the Watson-Crick base pairs C61-G37 (99%) and C60-G38 (98%), which are also described as the most crucial components of the tertiary interaction [20]. The presence of Mg^{2+} ions in the simulations of Gsw^{apt} results in a stabilization of almost all hydrogen-bonded base pairs and in occupancies for the hydrogen bonds of at least 78% except for the base pair A66-G38 (61%). The occupancies in the presence of Mg^{2+} ions are up to 44% higher than in the absence of Mg^{2+} ions.

For Gsw^{loop}, the loss of two hydrogen bonds in the upper base quadruple (Fig 1C) results in an overall destabilized hydrogen bond network in our MD simulations, manifested in lower occupancies for the hydrogen bonds in the loop region. In the absence of Mg^{2+} ions, the base pair between the nucleotides 61 and 65 in the upper base quadruple shows hydrogen bond occupancy of only ~5%, i.e., ~50% lower than for Gsw^{apt} in the absence of Mg^{2+}. The mutated Watson-Crick base pair A37-U61 yields a value of 86% in the absence of Mg^{2+}, ~10% lower than for the G37-C61 base pair in Gsw^{apt}. The last base pair in the upper base quadruple (A65-U34) also shows a reduced occupancy in the case of Gsw^{loop} compared to Gsw^{apt}. The lower base quadruple is influenced by the mutation as well. Here, the G38-C60 Watson-Crick base pair is again the most stable base pair from this quadruple (70%), whereas the other two base pairs show occupancies of 22% and 28% in the absence of Mg^{2+} ions. As in Gsw^{apt}, the presence of Mg^{2+} ions in simulations of Gsw^{loop} results in higher occupancy values for the hydrogen bonds of the hydrogen bond network, except for the interactions between the nucleotides 37 and 61.

In summary, we observe that the hydrogen bond network in the loop region of Gsw^{apt} is more stable than the network in Gsw^{loop}, as expected by the introduction of the G37A/C61U mutation in Gsw^{loop}. The G37A/C61U mutation does not only decrease the stability of the upper base quadruple in which the mutation is located, but also destabilizes the lower base quadruple. Furthermore, we found that the presence of Mg^{2+} ions results in more stable tertiary interactions for both Gsw^{apt} and Gsw^{loop}, in agreement with experimental findings [15,31].

The presence of Mg^{2+} ions maintains a more compact structure

In order to investigate the influence of Mg^{2+} ions and the G37A/C61U mutation on the compactness of the aptamer domain of Gsw, we calculated the radius of gyration (R_g) of the RNA molecules (Table 3). In the presence of Mg^{2+} ions, the mean R_g of both Gsw^{apt} (16.2 ± 0.3 Å) and Gsw^{loop} (16.4 ± 0.4 Å) agree with the value calculated for the ligand-bound crystal structures (16.1 Å and 16.3 Å, respectively), which contain $[Co(NH_3)_6]^{3+}$ ions in addition. Only minor differences were observed when comparing the mean R_g from MD simulations with 12 and 20 Mg^{2+} ions. The mean R_g in the absence of Mg^{2+} ions is increased by ~1 Å for both

Table 3. Radius of gyration of Gsw^{apt} and Gsw^{loop} [a].

Gsw variant	0 Mg^{2+} [b]	12 Mg^{2+} [b]	20 Mg^{2+} [b]	Crystal structure [c]
Gsw^{apt}	17.2 ± 0.1	16.3 ± 0.1	16.2 ± 0.1	16.1 [d]
Gsw^{loop}	17.4 ± 0.1	16.4 ± 0.1	16.4 ± 0.1	16.3 [e]

[a] In Å; given is the mean ± SEM calculated over three trajectories each. The radius of gyration was calculated omitting the P1 region.

[b] Number of Mg^{2+} ions per RNA molecule.

[c] The crystal structures are ligand-bound.

[d] Calculated for PDB ID 4FE5 [17]

[e] Calculated for PDB ID 3RKF [18]

https://doi.org/10.1371/journal.pone.0179271.t003

Gsw^{apt} and Gsw^{loop}. The more compact structure in the presence of Mg^{2+} ions is maintained due to the ions compensating the negative charges of the RNA backbone. For all three Mg^{2+} ion concentrations, the differences in mean R_g between Gsw^{apt} and Gsw^{loop} are only marginal ($\leq$ 0.2 Å), indicating only a minor influence of the mutation on the compactness of the structure.

In summary, we observe that the G37A/C61U mutation does not have an influence on the structural compactness of the RNA. The absence of Mg^{2+} ions in the MD simulations results in a less compact structure than when Mg^{2+} is present and in the ligand-bound crystal structures.

The G37A/C61U mutation increases, whereas Mg^{2+} ions decrease the structural dynamics of the aptamer

In order to gain insights into the influence of Mg^{2+} ions and the G37A/C61U mutation on the dynamics of the RNA, we calculated atomic root mean-square fluctuations (RMSF) and averaged them per nucleotide (Fig 2A and 2B). For all simulations, the stem regions P2 and P3 are the least mobile, while the terminal P1 region shows high RMSF. The low mobility of the P2 and P3 regions is in line with SHAPE experiments of the unbound state of a purine riboswitch characterizing these regions among the most stable ones [16].

In all MD simulations, the J2/3 region shows a pronounced mobility in agreement with experiments [16] and the assumption that the J2/3 region acts as an entry gate to the ligand binding site [9]. The fluctuations of nucleotides in the J1/2 and J3/1 regions are similar in magnitude to those in the P2 and P3 regions in all MD simulations, except for Gsw^{loop} in the absence of Mg^{2+} ions. Here, nucleotide 74 in the J3/1 region shows RMSF values that are ~2 Å higher than those of the other simulations. Pronounced fluctuations can also be observed in the L2 and L3 regions. In general, Gsw^{loop} shows a higher mobility than Gsw^{apt}, as seen from mean RMSF values calculated over the 80% least fluctuating nucleotides (S2 Table) that are up to ~0.6 Å higher in the former case (S4 Table). The presence of 20 Mg^{2+} ions results in mean RMSF values that are lower by ~0.5 Å (~0.8 Å) for Gsw^{apt} (Gsw^{loop}) than in the absence of Mg^{2+} ions.

In Fig 2C, 2D and 2E, differences in mean RMSF on a per-nucleotide level are projected onto the aptamer structure. Comparing the mean RMSF of Gsw^{apt} in the absence and presence of 20 Mg^{2+} ions (Fig 2C), we observe differences $\leq$ 1 Å at the ends of the terminal P1 region and in the loop region, which points to a subtle Mg^{2+}-induced decrease in structural dynamics of the loop region and the adjacent P3 region. Addressing the influence of the G37A/C61U mutation on the aptamer mobility surprisingly reveals that the largest differences in RMSF between Gsw^{loop} and Gsw^{apt} in the absence of Mg^{2+} ions (Fig 2D) are not found at the mutation site in the loop region, but at the ligand binding site ~30 Å away and at the even farther end of the terminal P1 region. At the ligand binding site, particularly J2/3 and J3/1 show higher mobility in Gsw^{loop}, with the largest difference calculated for nucleotide 74 (2.4 Å). These results provide the first evidence at the atomistic level that the G37A/C61U mutation has a long range effect in terms of increasing the structural dynamics at the ligand binding site, as has been inferred from the influence on the ligand binding ability of Gsw^{loop} [15,18], *and* the P1 region, which has not been considered so far.

As for Gsw^{apt} but much more marked, we observe a Mg^{2+}-induced decrease in structural dynamics of the aptamer for Gsw^{loop}: The Mg^{2+} ions now decrease the structural dynamics across the entire aptamer domain (Fig 2E). Large differences ($>$ 2 Å) are found for the ends of the terminal P1 region (between 2.2 and 4 Å) and for the nucleotides in J3/1. Especially nucleotide 74 shows a large decrease in mobility in the presence of Mg^{2+} ions (2.2 Å). In addition, the

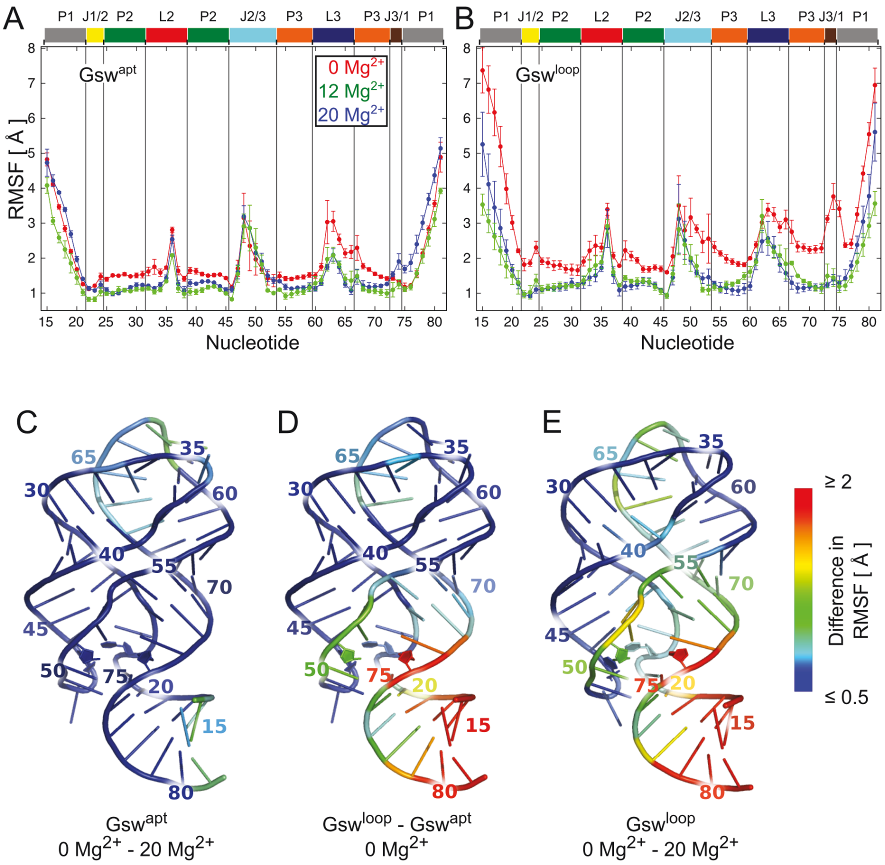


Fig 2. Atomic fluctuations calculated from MD simulations. A, B: Mean (± SEM) atomic fluctuations (RMSF) on a per nucleotide level for Gsw^{apt} (A) and Gsw^{loop} (B) over the three simulations for each system setup. Secondary structure regions are depicted above the plots and colored according to Fig 1A. Red: simulations in the absence of Mg^{2+} ions; green: simulations with 12 Mg^{2+} ions per RNA molecule; blue: simulations with 20 Mg^{2+} ions per RNA molecule. C, D, E: Differences in atomic fluctuations projected onto the RNA structure. Larger differences are colored red, smaller differences blue. Nucleotides responsible for ligand binding are shown as sticks. C: Difference in atomic fluctuations for Gsw^{apt} of 0 Mg^{2+}—20 Mg^{2+} ions per RNA molecule. D: Difference in atomic fluctuations in the absence of Mg^{2+} for Gsw^{loop}—Gsw^{apt}. E: Difference in atomic fluctuations for Gsw^{loop} of 0 Mg^{2+}—20 Mg^{2+} ions per RNA molecule.

https://doi.org/10.1371/journal.pone.0179271.g002

L3 loop, the P3 region, and the joining J2/3 region are affected by the Mg^{2+}-induced decrease in structural dynamics. Finally, the differences in RMSF between Gsw^{loop} and Gsw^{apt} in the presence of 20 Mg^{2+} ions (S8 Fig) are mostly small, with the largest difference (1.2 Å) found for nucleotide 63, which is part of the L3 loop, and other differences (< 1 Å) observed in the loop region.

In summary, we find that the G37A/C61U mutation results in increased structural dynamics of the aptamer domain of Gsw^{loop} compared to Gsw^{apt}. This is not only the case in the loop region where the mutation in Gsw^{loop} is located, but interestingly also at the distant ligand binding site *and* the P1 region. Mg^{2+} ions decrease the structural dynamics of the aptamer domains of both Gsw^{apt} and Gsw^{loop}. The Mg^{2+}-induced decrease of the structural dynamics is much more pronounced in the case of Gsw^{loop}, however, such that Gsw^{apt} and Gsw^{loop} behave similarly in terms of atomic mobility in the presence of 20 Mg^{2+} per RNA.

Particularly in the presence of Mg^{2+}, information on ligand binding is channeled to the P1 region

In order to gain further insights into the interplay of the L2-L3 region, the ligand binding core, and the P1 region as a function of ligand binding, we applied a graph-based rigidity analysis as implemented in the FIRST software [48]. The approach converts structural information of an input structure into a constraint network, where vertices represent atoms and edges (constraints) represent covalent and noncovalent interactions (hydrogen bonds, salt bridges, and hydrophobic interactions). Subsequently, the pebble game algorithm [49,50] is applied on the constraint network, which results in a decomposition into rigid regions and flexible links in between. Rigidity analyses have been successfully applied to investigate static properties (flexibility and rigidity characteristics) of proteins [48,51–54], also including analyses on the effect of ligand binding [52,55]. As to RNA, the approach was applied to the large ribosomal subunit [56–58] as well as smaller RNAs [59]. To do so, an RNA-specific parameterization of the constraint network has been developed [59], which is applied here. As suggested earlier [52,60,61], we applied the rigidity analysis to an ensemble of structures in order to improve the robustness of the results.

From the rigid cluster decomposition, we calculated the probability for each nucleotide to be in the largest rigid cluster ($p_{lrc}(i)$, Eq 1), as done previously [52], and averaged these results over the MD-derived ensembles of Gsw^{apt} and Gsw^{loop} *apo* structures. In order to simulate the presence of guanine in the binding site of Gsw, we added additional constraints to the network connecting the nucleotides U47, C74, and U51, which bind guanine in the crystal structure (Fig 3A and 3B). According to ref. [59], the number and distribution of constraints will result in the nucleobases of U47, C74, and U51 forming one rigid unit. To probe how this local rigidification will percolate through the aptamer domain, we also calculated $p_{lrc}(i)$ per nucleotide in the presence of the ligand constraints and, finally, the difference $\Delta p_{lrc}(i) = p_{lrc}(i)_{lig} - p_{lrc}(i)_{apo}$ (S9 Fig, Fig 3C, 3D, 3E and 3F). The results yield insights into the influence of ligand binding on the location of the largest rigid cluster in the aptamer domain. Note that by construction, any influence due to conformational changes of the RNA between *apo* and ligand-bound state is excluded.

In the absence of Mg^{2+} ions in the MD simulations (Fig 3C and 3D), the presence of the ligand constraints results in an increase of $p_{lrc}(i)$ for nucleotides in the J1/2, J2/3 and J3/1 regions, which is not surprising. Similarly, nucleotides in P1, P2, and P3 adjacent to the ligand binding site show an increased $p_{lrc}(i)$ in the presence of the ligand constraints, including nucleotides in the "P2 tune box" [17]. These results reveal a stabilization of the ligand binding site and its environment "above and below" due to the presence of ligand constraints.

In the presence of 20 Mg^{2+} ions (Fig 3E and 3F), we observe that the presence of the ligand constraints has only little effect on $p_{lrc}(i)$ of nucleotides in the P2, P3, L2, and L3 regions ($|\Delta p_{lrc}(i)| \leq 1\%$). In contrast, the presence of the ligand constraints result in a much larger increase in $p_{lrc}(i)$ for nucleotides in the J2/3 region (up to 42%) and in the J3/1 region (up to 20%). As for the absence of Mg^{2+} ions, we again observe an increase in $p_{lrc}(i)$ for all nucleotides in the P1 region by the presence of the ligand constraints, even though the values are higher in

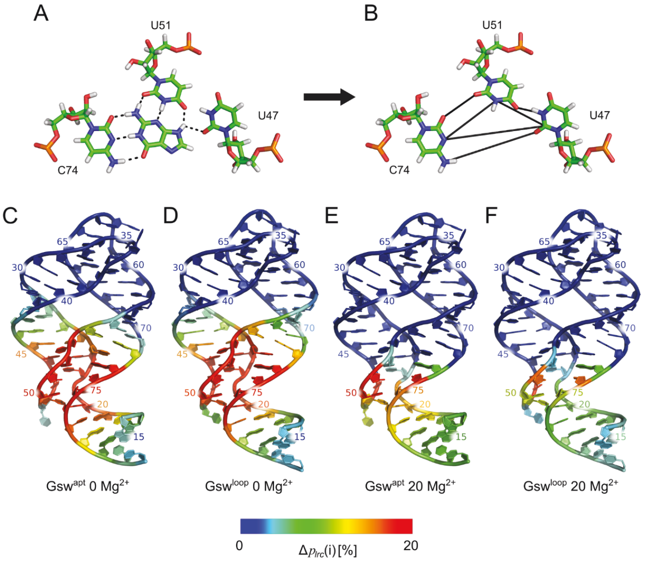


Fig 3. Rigidity analyses. A: Nucleotides involved in binding guanine in the binding site of the aptamer domain of Gsw (PDB ID 1Y27 [26]). Black dashed lines indicate hydrogen bonds. B: Constraints (black lines) added to conformations of the *apo* aptamer domain of Gsw to model the presence of guanine in the binding site for rigidity analyses. C, D, E, F: Nucleotide-wise differences in the probability to be in the largest rigid cluster ($\Delta p_{lrc}(i)$, Eq 1, E_{HB} = **−0.6 kcal/mol** for the rigidity analyses) between the ligand being present or absent in the aptamer domain, projected onto the aptamer domain of Gsw^{apt} (C, E) and Gsw^{loop} (D, F) in the absence of Mg^{2+} ions (C, D) and in the presence of 20 Mg^{2+} ions (E, F).

https://doi.org/10.1371/journal.pone.0179271.g003

the absence of Mg^{2+} ions. Thus, in the presence of 20 Mg^{2+} ions, the stabilizing effect of ligand constraints is more local and channeled towards the P1 region.

The detailed comparison of $\Delta p_{lrc}(i)$ between Gsw^{apt} and Gsw^{loop} in the absence of Mg^{2+} ions (Fig 3C and 3D) shows that the influence of the ligand constraints is similar by and large for the two variants, although differences up to ~5% in $\Delta p_{lrc}(i)$ are found. Nucleotides in the binding region show a stronger influence of the ligand constraints in Gsw^{loop}. This is likely due to the larger flexibility observed for Gsw^{loop} in the binding region in the *apo* state, which becomes more restricted by the presence of the ligand constraints. In contrast, the presence of the ligand constraints has a slightly larger influence on nucleotides in the P1 region for Gsw^{apt} (up to 3%) than for Gsw^{loop}. In the presence of 20 Mg^{2+} ions, the stabilizing effect of the ligand constraints towards the terminal ends in the P1 region is more pronounced for Gsw^{apt}, which shows $\Delta p_{lrc}(i)$ values for this region 2 to 7% higher than for Gsw^{loop}, with the higher values being closer to the ligand binding site.

In summary, we observe an increase in $p_{lrc}(i)$ in the presence of ligand constraints for nucleotides at the ligand binding site and in its environment. In Gsw^{apt} and when 20 Mg^{2+} ions are present, the rigidifying effect towards the terminal ends of P1 is largest. Considering that the tertiary loop interactions are already preformed in the unbound state of Gsw^{apt} even in the absence of Mg^{2+} ions [31,32], yet are further strengthened in the presence of Mg^{2+} [31,62], this

Table 4. ***Coop*(*i*) values for the terminal nucleotides in the P1 region[a].**

Nucleotide numbers	0 Mg^{2+}		20 Mg^{2+}	
	Coop(*i*) ± SEM [b]	*p* [c]	*Coop*(*i*) ± SEM [b]	*p* [c]
15	0.009 ± 0.038	ns	0.065 ± 0.021	**
16	0.003 ± 0.036	ns	0.060 ± 0.019	**
17	-0.020 ± 0.036	ns	0.061 ± 0.017	***
18	-0.043 ± 0.037	ns	0.060 ± 0.016	***
19	-0.090 ± 0.040	ns	0.062 ± 0.015	***
77	-0.136 ± 0.039	**	0.062 ± 0.015	***
78	-0.043 ± 0.037	ns	0.060 ± 0.016	***
79	-0.020 ± 0.036	ns	0.061 ± 0.017	***
80	0.001 ± 0.036	ns	0.060 ± 0.019	**
81	0.005 ± 0.039	ns	0.068 ± 0.021	**

[a] *Coop*(*i*) values were calculated based on Eq 2.

[b] The SEM was calculated via error propagation from the single terms of Eq 2.

[c] The *p* value was calculated using a two-sided one-sample *t*-test and the null hypothesis of equality to zero.

* $p < 0.05$;

** $p < 0.01$;

*** $p < 0.001$;

ns: $p \geq 0.05$

https://doi.org/10.1371/journal.pone.0179271.t004

finding suggests that both the L2-L3 region and ligand binding cooperatively stabilize the P1 region.

Tertiary interactions in the loop region and ligand interactions cooperatively stabilize the terminal nucleotides in the P1 region

In order to verify that the tertiary loop-loop interactions and the presence of the ligand constraints rigidify the terminal nucleotides in the P1 region cooperatively, we computed the cooperative effect *Coop*(*i*) on the stabilization of nucleotides *i* in the P1 region from respective $p_{lrc}(i)$ values according to Eq 2. If $Coop(i) > 0$, interactions at both sites together have a larger effect on nucleotide *i* than the sum of the effects of the separate interactions.

For the terminal nucleotides in the absence of Mg^{2+} ions, almost all of the nucleotides in the five terminal base pairs show *Coop*(*i*) values that are not significantly different from zero (Table 4), demonstrating that in this case the L2-L3 region and the ligand binding do not act cooperatively on P1. The only exception is nucleotide 77, which is close to the ligand binding site. In contrast, in the presence of 20 Mg^{2+} ions, all nucleotides in the five terminal base pairs show *Coop*(*i*) values that are positive and significantly or even highly significantly different from zero (Table 4, Fig 4A), demonstrating that with a structurally more stable aptamer domain, the L2-L3 region and the ligand binding act cooperatively on P1.

In summary, the stabilization of the L2-L3 region by tertiary interactions and the ligand binding site by ligand binding cooperatively influences the structural stability of the terminal base pairs in the P1 region in the presence of Mg^{2+} ions.

Discussion

We investigated the interplay of structural stability of the L2-L3 region, the ligand binding core, and the P1 region in the aptamer domain of Gsw by MD simulations and rigidity analyses of the wildtype structure and the G37A/C61U mutant. As the main result, we found that

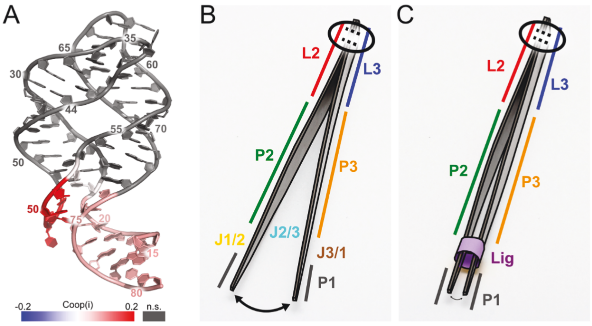


Fig 4. Cooperative influence on the P1 region. A: *Coop*(*i*) values mapped onto the aptamer. The values were calculated according to Eq 2 for the systems in the presence of 20 Mg^{2+} ions (see also Table 4). Grey nucleotides show *Coop*(*i*) values that are not significantly different from zero ($p > 0.05$). B, C: Model for the influence of tertiary interactions and ligand binding on the stability of the P1 region. The tweezer represents the aptamer domain of the guanine-sensing riboswitch. The secondary structure elements are indicated by colors as in Fig 1A. The tertiary interactions are shown as dotted lines at the top of the tweezers and are encircled. The flexibility of the P1 region is indicated by the differently sized arrows at the bottom. B: The ligand-unbound state; C: The ligand (purple) has a stabilizing influence on the P1 region, but more so if the tertiary interactions are present.

https://doi.org/10.1371/journal.pone.0179271.g004

stabilization of the L2-L3 region by tertiary interactions, and the ligand binding site by ligand binding, cooperatively influences the structural stability of the terminal base pairs in the P1 region in the presence of Mg^{2+} ions.

Previous, comprehensive studies had revealed intricate relationships between ligand binding to the aptamer domain of Gsw and presence of tertiary L2-L3 interactions [8,19] [16], as well as between ligand binding and influence on the P1 region [8] [16]. These studies stressed that subtle (structural) changes, such as the loss of two hydrogen bonds in the Watson-Crick base pair in the upper base quadruple in the L2-L3 region or the presence or absence of Mg^{2+} ions, profoundly affect the function of the aptamer domain, indicating that the overall structural stability of this domain in the absence of a ligand is marginal. While yielding valuable data to validate our simulation and modeling data with, these findings also provide a challenge in terms of appropriately parameterizing and setting up our computations.

As to the parameterization, contemporary force fields for RNA usually well describe canonical helices; however, they may show different successes on more complex RNAs [63–65]. To investigate to what extent a force field can bias insights gained from MD simulations, we had performed MD simulations with an aggregate simulation time of > 11 μs in previous work [33], probing the influence of the Amber ff99 force field [66] as well as of ff99 with parmbsc0 [67] and parmχOL [68] modifications on the aptamer's properties. There, ff99 yielded the best agreement with experimental observations, which provided the incentive for us to use ff99 also in the present study. Furthermore, results of rigidity analyses depend on the definition when to include a constraint into the network representation to model a noncovalent interaction [48]. In previous work, we had parameterized a constraint network representation for RNA structures [59]. The modifications were verified by comparing predictions from rigidity analysis to mobility information derived from crystallographic B-values and by predicting and comparing conformational variabilities of RNA structures to those of NMR ensembles [59]. This constraint network representation was used here. Finally, modeling Mg^{2+} ions in MD simulations is considered challenging [36,37,39,64]. Following our earlier findings [33], we used the

Mg^{2+} parameters developed by Åqvist [69] and placed them by a two-step procedure including the first hydration shell. This treatment resulted in a pronounced mobility of the Mg^{2+} ions during the first ~100 ns of the MD simulations, suggesting sufficient time for them to equilibrate around the aptamer domain. As a result, preferred occupation sites of the Mg^{2+} ions agree very well with experimentally determined ones from crystal structures and NMR experiments and with respect to different MD trajectories for one system. It is not surprising, though, that we do not obtain perfect agreement with the experimentally determined Mg^{2+} ion binding sites, as even two crystal structures do not yield exactly identical occupation of the binding sites (Fig 1E). Therefore, our setup and modeling of Mg^{2+} ions led to results that suggest that the respective systems sufficiently settled, and in a consistent manner, with respect to Mg^{2+} ion binding such that they can be compared to each other for investigating the Mg^{2+} dependence on the structure and dynamics of Gsw^{apt} and Gsw^{loop}.

Likely the largest challenge arose from modeling the structural dynamics of Gsw^{apt} and Gsw^{loop} in the *apo* state, starting from the conformations of the ligand-bound state extracted from crystal structures [8,18]. This challenge is reflected in that even for a less complex B-DNA structure, already ~5 μs of simulation time was required to achieve converged results (63). At present, no experimental structural information at the atomistic level of the unbound aptamer domain of Gsw is available. Yet, the facts that the aptamer domain of Gsw is one of the smallest riboswitch aptamer domains structurally known and that the global fold of the aptamer domain is already present in the unbound state [11,18,19,31,32] make the Gsw aptamer domain a seemingly well-suited system for our investigations. Experimental findings suggest that no large-scale conformational changes need to be simulated to reach the *apo* state [70]. Accordingly, the structural ensembles generated in our MD simulations display pronounced structural characteristics in very good agreement with experiment. At the global level, we found that the mean R_g of both Gsw^{apt} and Gsw^{loop} in the presence of 20 Mg^{2+} ions agree with respect to absolute and relative magnitudes with values calculated for the respective crystal structures (Table 3), despite local structural rearrangements in the loop region and at the ligand binding site induced by the double mutation in Gsw^{loop} (Table 1). Furthermore, we found that the mean R_g in the presence of Mg^{2+} ions is significantly ($p < 0.005$) reduced for both Gsw^{apt} and Gsw^{loop}, in agreement with experimental results on other riboswitch aptamer domains [71].

The tertiary interactions in the L2-L3 region play an important role for the structural stability and the ligand binding ability of the aptamer domain [8,16]. In experiments, the introduction of the G37A/C61U mutation caused a destabilization of the tertiary interactions and thereby influenced the ligand binding ability of Gsw^{loop} in a Mg^{2+}-dependent manner [31]. In our MD simulations, we did not observe a complete disruption of the tertiary interactions in Gsw^{loop} in the absence of Mg^{2+} ions, in contrast to NMR experiments where no tertiary interactions were found [18]; yet, hydrogen bond occupancies in the upper base quadruple reduced by up to 50% compared to Gsw^{apt} were found, as were reduced occupancies in the lower base quadruple (Table 2). Thus, while the first finding indicates that our simulation length is likely not long enough to reach an exact structural state of the loop L2-L3 region of *apo* Gsw^{loop} when starting from the crystal structure with formed tertiary interactions [18], the second finding suggests that our structural ensembles do reflect important differences between Gsw^{apt} and Gsw^{loop}. Furthermore, we observe a stabilization of the hydrogen bond network in the presence of Mg^{2+} ions, for both Gsw^{apt} and Gsw^{loop}, in agreement with experimental findings [15]. Hence, the ability to model the destabilized hydrogen bond network in the loop region in the case of Gsw^{loop} allows us to use Gsw^{loop} as a test system for investigating the influence of the absence of stable tertiary interactions on the structure and dynamics of the aptamer domain.

In general, we find that Gsw^{loop} shows higher atomic fluctuations than Gsw^{apt}, especially in the absence of Mg^{2+} ions (Fig 2A, 2B and 2D; S4 Table). The presence of Mg^{2+} ions results in a smaller difference in atomic fluctuations between Gsw^{apt} and Gsw^{loop}, reducing the effect of increased structural dynamics due to the mutation in Gsw^{loop} (Fig 2A and 2B; S8 Fig; S4 Table). In more detail, the largest mobility for all simulated systems occurs in the J2/3 region. This is in agreement with the fact that the J2/3 region was experimentally found to be the most flexible region in the unbound aptamer [16], and is in accordance with the assumption that J2/3 acts as an entry gate to the binding site [9]. Furthermore, our MD simulations show that the junction region shows decreased structural dynamics in the presence of Mg^{2+} ions compared to their absence, especially in Gsw^{loop} for which the stability of the tertiary interactions is dependent on Mg^{2+} ions as well. This is in agreement with results from chemical probing experiments that showed that the majority of the binding site experiences a significant reduction in conformational freedom upon formation of loop-loop interactions [16]. The J3/1 junction is another region that is influenced by the presence or absence of the tertiary interactions (Table 1, Fig 2). This region remained generally stable in our MD simulations, except for the simulations of Gsw^{loop} in the absence of Mg^{2+} ions, as seen from high RMSD and RMSF values in this case. The high mobility in the absence of Mg^{2+} ions in connection with the decreased mobility in the presence of Mg^{2+} ions could explain the experimental observation that Gsw^{loop} is not able to bind hypoxanthine in the absence of Mg^{2+} ions, while the hypoxanthine binding ability is restored in the presence of Mg^{2+} ions [18]: Nucleotide C74, located in J3/1 and hypothesized to be involved in the initial binding of the ligand [9,16], might exhibit too little stability for productive binding of the ligand in the absence of Mg^{2+}. This would be in line with, and extend, the atomistic understanding of the hypothesis that the formation of the initial encounter complex relies on a restricted conformational dynamics and a partial structural organization of the ligand binding core region [15]. The conformational restriction and partial organization of the ligand binding core has been hypothesized to be induced through stabilization of remote regions [15,16]. Indeed, we observed that the presence of Mg^{2+} ions in simulations of Gsw^{loop} results in a more stable hydrogen bond network in the loop region and a related stabilization of the ligand binding site.

In addition to the influence on the ligand binding core, our analyses of RMSF in the absence of Mg^{2+} reveal that the G37A/C61U mutation also increases the structural dynamics in the P1 region, which, to the best of our knowledge, has not been described before (Fig 2). The P1 region is about 40–50 Å away from the L2-L3 region and contains part of the switching sequence. The presence of Mg^{2+} reduces the structural dynamics particularly in that region, as indicated by a reduction in RMSF of > 2 Å (Fig 2), and this effect is stronger in Gsw^{loop} than in Gsw^{apt}. Together with the findings that the presence of Mg^{2+} restores the tertiary interactions in the L2-L3 region in Gsw^{loop} (Table 2), and further strengthens them in Gsw^{apt} [15] (Table 2), these results provide another indication that the tertiary interactions in the L2-L3 region exert a stabilizing effect on the P1 region.

The information about ligand binding and its effect has been shown or inferred to be transmitted to both the L2-L3 and P1 regions [8,14,16,21–23]. Our finding that the presence of the tertiary interactions in the L2-L3 region is transmitted to the P1 region then leads to the hypothesis that these three regions do not only “communicate” in a pairwise manner but that there is a cooperative interplay among all three. The cooperativity likely, at least in part, originates from dynamic allostery [72], i.e. due to changes in the structural dynamics upon ligand binding rather than by macromolecular conformational changes. To probe this interplay, we performed graph theory-based rigidity analyses [48,49] on the MD-derived structural ensembles of Gsw^{apt} and Gsw^{loop} in the *apo* state, modeling the presence of a ligand in the binding site by adding constraints to the network representation of the aptamer domains. In contrast to comparing the influence of a ligand in “true” *holo* ensembles of Gsw^{apt} and Gsw^{loop}, our procedure has the advantage that only the influence of additional interactions due to the ligand are detected, whereas potential ligand-induced conformational changes of the aptamer domain are excluded by definition. Obviously, the result of our ligand modeling will depend on the strength with which we couple the ligand to the aptamer network. It is important to note that we used a “weak” coupling scheme as we only provide constraints for hydrogen bonds that a ligand would form with and introduce between the three nucleobases U47, U52, and C47, according to information from crystal structures [8,26]. In particular, we do not include additional constraints for base stacking interactions between the ligand and nucleotides of the P1 region, nor those between the ligand and the base triple located above [8,26]. Such interactions would only increase effects due to ligand binding on the L2-L3 and/or P1 regions.

As a direct measure of the ligand influence on the structural stability, we evaluated the nucleotide-wise probability of finding a nucleotide in the largest rigid cluster across the structural ensemble, $p_{lrc}(i)$. This measure had been used successfully to analyze stabilizing effects on the binding partners in the context of protein-protein complex formation [52]. Notably, while in all cases an increase in $p_{lrc}(i)$ in the presence of ligand constraints was found for nucleotides at the ligand binding site and in its environment, the rigidifying effect towards the terminal ends of P1 was largest for Gsw^{apt} and when 20 Mg^{2+} ions were present (Fig 3). Computation of the cooperative effect *Coop*(i) then confirmed that stabilization of the L2-L3 region by tertiary interactions, and the ligand binding site by ligand binding, cooperatively influences the structural stability of the terminal base pairs in the P1 region in the presence of Mg^{2+} ions (Table 4). Hence, our computations reveal for the first time that there is a structurally stabilizing interplay among the L2-L3 region, the ligand binding core, and the P1 region in the aptamer domain of Gsw^{apt}. Accordingly, both the L2-L3 region and the ligand binding contribute to a more stable terminal part of the P1 region, the consequence of which could be that the adjacent switching region is not available for the formation of the antiterminator and, therefore, the transcription terminator loop can form [8,16]. A simple mechanical analog to this allosteric mechanism are tweezers (Fig 4B and 4C), where the mechanically connected end corresponds to the tertiary interactions, which are inherently necessary for the functional stability. The variability in the terminal ends of the tweezers, equivalent to P1, can be restrained by a sleeve (purple in Fig 4C) corresponding to the ligand at the ligand binding site. (Note that this analogy does not take into account that the ligand also has a stabilizing effect on the L2/L3 region.) Our functional insights may provide an explanation as to why the three-way junction is a central structural element also in other purine riboswitches and in the TPP riboswitch [12].

As to the role of Mg^{2+}, the Mg^{2+} concentration in *Bacillus subtilis* cells is estimated to be in the millimolar range [73], similar to the intracellular Mg^{2+} concentrations in the range of ~0.5 to 2 mM reported for other bacteria [74]. For intracellular riboswitch concentration, values between 1 to 100 nM have been estimated [75,76]. Together, this would result in [Mg^{2+}]:[RNA] ratios that embrace the critical [Mg^{2+}]:[RNA] ratio of ~20, above which the rigidifying effect towards the terminal ends of P1 is largest. Therefore, the cooperative effect detected here could become contextual [77] in that the same system of molecular components, Gsw and its ligands, can display different properties depending on the [Mg^{2+}]:[RNA] ratio.

Regarding the significance of our findings for the full aptamer’s switching function, by definition any conformational changes of the aptamer domain were excluded upon modeling the presence of the ligand for the rigidity analyses. In the absence of a conformational change, the observed stability changes relate to changes in the width of the states [78], such that the transmission of allosteric information from the ligand binding core to the P1 region can be entropy-dominated, as suggested by Cooper and Dryden [72] and Tsai *et al.* [79]. This result emphasizes the role of changes in the aptamer’s dynamics for the allosteric communication

[72]. Further confirmation for this entropy-dominated allosteric communication could be obtained from rigorous entropy calculations [80], which are beyond the scope of this study, however, or from NMR relaxation studies of dynamics in RNA. The proposed allosteric mechanism is in line with models that favor a high degree of preorganization of the unbound state of the aptamer domain [31,32,62], because according to this mechanism no large structural changes are required to transmit information about a bound ligand to the terminal base pairs of the P1 region. Our finding would also be supported by the fact that the ligand binding time (~0.04–0.25 s as estimated from on-rates for guanine binding of 0.2–1.5 * 10^5 M^{-1} s^{-1} for Gsw variants [9,17] and a guanine concentration of ~190 μM in *E. coli* [81]) is already comparable to the time required to transcribe the full aptamer (considering a transcription rate of the RNA polymerase of ~50 nt/s [17] and a length of the expression platform of ~100 nucleotides [8,82]) such that a quick stabilization of the P1 region is required in order to drive the formation of the transcription terminator in time for the subsequent regulation [83].

In our study, we only investigated the aptamer domain of the two Gsw variants Gsw^{apt} and Gsw^{loop}. A recent NMR study of the structurally similar adenine-sensing riboswitch uncovered a three-state behavior of the complete riboswitch including the expression platform [75]: A ligand bound state, a similar ligand-unbound yet binding-competent state, and a newly found ligand-unbound state incapable of ligand binding. Thus, ideally further studies should include the expression platform as well. However, due to the scarcity of structural information for the expression platform and the long time-scales for folding of RNA structures compared to the time-scales of standard MD simulations, the inclusion of the expression platform in all-atom MD simulations is difficult.

Methods

Molecular dynamics simulations

The setup of the molecular dynamics (MD) simulations in the present study is similar to that in our previous work [33]. However, the MD simulations performed here are 2.5 times longer than before. In detail: MD simulations were performed with the Amber 11 suite of programs [84,85] and the force field ff99 [86]. The starting structures for the MD simulations were obtained from ligand bound crystal structures. The ligands were removed from these, and the sequence of the wild type structure (PDB ID: 4FE5 [17]) was adapted to match the sequence of the G37A/C61U mutant (PDB ID: 3RKF [18]). For this, the coordinates of the bases of the mutant were copied after superimposing the wild type structure. As a result, wild type and mutant only differed in nucleotides 37 and 61. The modified wild type will be referred to as Gsw^{apt}, and the G37A/C61U mutant will be referred to as Gsw^{loop}, as introduced by Schwalbe *et al.* [18]. For both structures, three simulation systems with different Mg^{2+} concentrations (0, 12 and 20 Mg^{2+} ions per RNA molecule) were set up according to experimental findings on the Mg^{2+} dependence of Gsw^{loop} properties [15,18,31]. Due to slow exchange times of first shell ligands of the Mg^{2+} ions [37], the Mg^{2+} ions were added to the RNA system with a first hydration shell of six water molecules. This was done to prevent an early "sticking" of the Mg^{2+} ions to the RNA during the equilibration phase. To prevent clashes between the hydration shell and the RNA during the placement step, dummy ions with equal charge but larger radius of 4 Å, which is close to the radius of a hexahydrated Mg^{2+} ion [87], were placed first at electrostatically favorable locations with *leap* of Amber 11; these dummy ions were then replaced with a hexahydrated Mg^{2+} ion [69]. For details about the influence of the initial placement and the choice of the Mg^{2+} parameters see [33]. The system was afterwards placed in an octahedral box of TIP3P water molecules [88]. The distance between the edge of the water box and the closest RNA atom was at least 11 Å. Na^+ ions [89] were added to neutralize the system net charge, resulting in a final system size of ~50,000 atoms.

Each system was then prepared based on a protocol described earlier [90]. In detail, each system was minimized by 200 steps of steepest descent minimization, followed by 50 steps of conjugate gradient minimization. The particle mesh Ewald method [91] was used to treat long range electrostatic interactions, and the SHAKE algorithm [92] was used to constrain bond lengths involving bonds to hydrogen atoms. The time step for all MD simulations was 2 fs, with a direct-space nonbonded cutoff of 9 Å. We carried out canonical ensemble (NVT)-MD simulations for 50 ps, during which the system was heated from 100 to 300K. During this step, harmonic restraints with force constants of 5 kcal mol^{-1} $Å^{-2}$ to all solute atoms were applied. In addition, harmonic restraints were applied to the Mg^{2+} ions and the water molecules in their first hydration shell. Subsequently, isothermal isobaric ensemble (NPT)-MD simulations were used for 50 ps to adjust the solvent density. Finally, the force constants of the harmonic restraints on positions of RNA, Mg^{2+}, and first hydration shell waters were gradually reduced to 1 kcal mol^{-1} $Å^{-2}$ during 250 ps of NVT-MD simulations. This was followed by 50 ps of NVT-MD simulations without applying positional restraints. From the following 550 ns of NVT-MD simulations at 300 K performed by the GPU version of *pmemd* [93], conformations were extracted every 20 ps.

For each of the six system setups (Gsw^{apt} and Gsw^{loop} with three Mg^{2+} concentrations each), three independent MD simulations of 550 ns length each were performed summing up to a total simulation time of ~10 μs.

Trajectory analyses

The MD simulations were analyzed using *cpptraj* [94] of the AmberTools 13 program suite [95]. The initial 50 ns of each trajectory were discarded because the largest increase in overall RMSD was observed during this period (S3 Fig). The *radius of gyration* (ROG), a measure of the compactness of a structure, was calculated omitting the P1 region. This was done for conformations generated by MD simulations as well as for the crystal structures. Prior to the calculation of *root mean-square atomic fluctuations* (RMSF), global translational and rotational differences between the structures along the trajectory need to be removed by least-squares fitting. In order to reduce the influence of very mobile regions on the picture of internal motions [52,96–98], conformations were root mean-square fitted on those 80% of the nucleotides with the lowest fluctuations (S2 Table). These 54 nucleotides were chosen from an initial calculation of the mean RMSF over the three simulations of Gsw^{apt} in the absence of Mg^{2+} ions. These nucleotides were also used in the case of Gsw^{loop} then. *Root mean-square deviations* (RMSD) of atomic positions were calculated as a measure for structural similarity with respect to the initial structures. The RMSD was calculated after aligning the conformations onto the initial structure, using those 80% of the nucleotides with the lowest RMSF. The RMSD was then calculated for all atoms of the RNA as well as for all atoms of substructural parts of the aptamer (S5 Table). In order to investigate the change within a substructural part of the aptamer, RMSD values in S3 Table were calculated for the substructural parts after root mean-square fitting on the respective part in the initial conformation. Mean values were calculated over all three MD trajectories. Preferred *Mg^{2+} ion occupancy sites* were determined by the *cpptraj* program using the *grid* command using a grid spacing of 0.4 Å. The *occupancy for hydrogen bonds* between bases in the L2/L3 base quadruples were determined using default geometrical parameters (distance: 3.5 Å; angle: 120˚). Block averaging over blocks of 10 ns length was used to determine the statistical uncertainty within one trajectory. The values reported in Table 2 were averaged over all hydrogen bonds between a pair of bases. The hexahydration of Mg^{2+} ions was determined by "following" the hydration shell around the ions over the simulation. Here, "following" means that we calculated the number of water molecules around each Mg^{2+} ion within a distance of 3.5 Å. Changes in the number of water molecules in the first hydration shell were subsequently confirmed by visual inspection. Furthermore, the direct chelation of Mg^{2+} ions to the RNA was determined by calculating the distance of each Mg^{2+} ion to the RNA. We considered a distance below 3.5 Å between a Mg^{2+} ion and an RNA atom indicative that the hydration shell of this Mg^{2+} ion is incomplete. The direct chelation was then confirmed by visual inspection of the trajectory.

We assessed the convergence of our simulations in two ways: First, we performed a principal component analysis for the phosphorus atoms of the aptamer over all trajectories. Then, we projected each trajectory onto the combinations of principal components (PC) PC1/PC2, PC1/PC3, and PC2/PC3. As shown in S10–S12 Figs exemplarily for Gsw^{apt} in the presence of 20 Mg^{2+} ions and Gsw^{loop} in the absence of Mg^{2+}, all projections markedly overlap when comparing MD simulations of the same system, and when comparing the first and the second half of the trajectories. For Gsw^{loop} in the absence of Mg^{2+} ions, the projections onto PC1 show a larger spread between the MD simulations; however, the projections onto PC2 and PC3 overlap well. These results demonstrate that the three independent MD simulations per system sample similar regions in PC space, and the sampling agrees between the first and second halves of the trajectories. Additionally, we calculated the RMSD average correlation (RAC) [94,99] as implemented in *cpptraj* [94]. The RAC is a measure for the time scales on which structural changes happen in MD simulations. The RAC curves for all our simulations are shown in S13 Fig. From the bumps in the curves, we can estimate that structural changes happen on time scales of ~50–100 ns for Gsw^{loop} in the absence of Mg^{2+} ions (S13B Fig), whereas for the other systems, such bumps are found at shorter time intervals. For time intervals > 300 ns, the curves appear smooth, which suggests that there are no large structural changes happening on these time scales. Note that we calculated the RAC with respect to the first structure omitting the first 50 ns of each MD trajectory. As an alternative, the RAC can be calculated with respect to the average structure of an MD trajectory, which would result in smaller values [99]. Overall, these analyses are indicative to what extent the conformational ensembles of sampled aptamer structures are converged.

Rigidity analyses

In order to investigate the influence of the presence of a ligand in the binding site on the rigidity characteristics of the aptamer domain, we applied the FIRST software (version 6.2) [48]. FIRST converts structural information of an input structure into a constraint network representation, where vertices represent the atoms and edges represent interatomic interactions, e.g. covalent bonds, hydrogen bonds, or hydrophobic interactions. FIRST then uses the pebble game algorithm [49,50] to decompose the network into rigid regions and flexible links.

FIRST has not only been successfully applied to protein structures, but also proved successful in investigating and predicting flexibility and rigidity characteristics of RNA structures [56–59]. Here, we use an RNA-specific set of parameters developed by Fulle *et al.* [59] for modeling constraints due to non-covalent interactions, using -0.6 kcal mol^{-1} as an energy cutoff E_{cut} for hydrogen bonds (Fig 3C, 3D, 3E and 3F; S9 Fig). A value of $E_{cut} = -1.0\ kcal/mol$ yielded very similar results (S14 Fig), as found earlier [59].

From the rigid cluster decomposition, we calculate the probability $p_{lrc}(i)$ that atom i belongs to the largest rigid cluster according to $p_{lrc}(i) = \frac{n_1(i)}{N}$, where $n_1(i)$ is the number of occurrences of atom i as part of the largest rigid cluster, determined over all N snapshots extracted from the trajectory [52]. $p_{lrc}(i)$ was calculated in the absence (*apo*) and presence (*lig*) of a ligand, as was the difference

$$\Delta p_{lrc}(i) = p_{lrc}(i)_{lig} - p_{lrc}(i)_{apo} \tag{1}$$

The average of $\Delta p_{lrc}(i)$ over the three independent MD simulations for each of the six system setups is shown in S9 Fig, and projected onto the aptamer domain in Fig 3C, 3D, 3E and 3F. For the visualization, the value calculated for the N1 atom of each nucleotide's base is projected onto the whole nucleotide.

Averaging over an ensemble of snapshots from an MD simulation increases the robustness of the results compared to using a single input structure only [52,60,61]. Snapshots were extracted from the MD trajectories, considering the RNA and, if present, Mg^{2+} ions and their first hydration shell water molecules as proposed by Fulle *et al.* [59]. In detail, interactions between Mg^{2+} ions and their first hydration shell water molecules, or Mg^{2+} ions and the RNA, were modeled as covalent bonds, whereas interactions between water and RNA were modeled as hydrogen bonds. These snapshots were subsequently converted using the *ambpdb* program from the AmberTools suite of programs to FIRSTdataset files. In order to model the presence of the ligand guanine for the rigidity analyses by FIRST, we added constraints to the FIRSTdataset file connecting nucleotides U47, C74, and U51 in a pair-wise manner; these nucleotides interact with the ligand in the crystal structure (Fig 3A and 3B). To validate that the added constraints represent the bound ligand in the binding site, we additionally performed the rigidity analysis on the minimized X-ray structure of Gsw^{loop} (i) without a ligand in the binding site, (ii) with guanine at the position of the original ligand, and (iii) with the added constraints representing the ligand, and investigated the rigid cluster decomposition of the binding nucleotides (U47, U51, and C74) (S15 Fig). Without guanine or the ligand-representing constraints, the binding nucleotides do not belong to one rigid cluster (S15A Fig). The presence of guanine at the ligand binding site results in a rigid cluster that comprises all three binding nucleobases (S15B Fig). Similarly, when adding the ligand-representing constraints to the network, all three binding nucleobases are part of one rigid cluster (S15C Fig). The difference in the spread of the rigid cluster visible between S15B and S15C Fig relates to using a "weak" coupling scheme for the ligand-representing constraints, as discussed above.

In order to investigate the interplay between the presence of tertiary interactions in the loop region and the presence of a ligand in the binding site, we evaluate cooperative effects on the P1 region. To do so, we computed $p_{lrc}(i)$ values for the aptamer domain without ligand constraints ($p_{lrc}(i)_{apo}^{Gsw^{apt}}$ and $p_{lrc}(i)_{apo}^{Gsw^{loop}}$ for Gsw^{apt} and Gsw^{loop}, respectively) and with ligand constraints ($p_{lrc}(i)_{lig}^{Gsw^{apt}}$ and $p_{lrc}(i)_{lig}^{Gsw^{loop}}$ for Gsw^{apt} and Gsw^{loop}, respectively). $p_{lrc}(i)_{apo}^{Gsw^{loop}}$ was chosen as the reference because it lacks interactions at both the loop region and the binding site. From this, we calculated the cooperative effect on the stabilization of the P1 region due to interactions at the two sites according to

$$Coop(i) = ln\left(\frac{p_{lrc}(i)_{lig}^{Gsw^{apt}}}{p_{lrc}(i)_{apo}^{Gsw^{loop}}}\right) - \left[ln\left(\frac{p_{lrc}(i)_{apo}^{Gsw^{apt}}}{p_{lrc}(i)_{apo}^{Gsw^{loop}}}\right) + ln\left(\frac{p_{lrc}(i)_{lig}^{Gsw^{loop}}}{p_{lrc}(i)_{apo}^{Gsw^{loop}}}\right)\right] \tag{2}$$

where the term

$$ln\left(\frac{p_{lrc}(i)_{lig}^{Gsw^{apt}}}{p_{lrc}(i)_{apo}^{Gsw^{loop}}}\right) \tag{3}$$

describes the influence of the tertiary interactions together with the ligand,

$$ln\left(\frac{p_{lrc}(i)_{apo}^{Gsw^{apt}}}{p_{lrc}(i)_{apo}^{Gsw^{loop}}}\right) \tag{4}$$

describes the influence of the tertiary interactions alone, and

$$ln\left(\frac{p_{lrc}(i)_{lig}^{Gsw^{loop}}}{p_{lrc}(i)_{apo}^{Gsw^{loop}}}\right) \tag{5}$$

describes the influence of the ligand alone.

If $Coop(i) > 0$, interactions at both sites together have a larger effect on nucleotide i than the sum of the effects of the separate interactions. We are mainly interested in the effect on the terminal nucleotides in the P1 region, and thus in $Coop(i)$ for $i \in \{15, 16, 17, 18, 19, 77, 78, 79, 80, 81\}$.

Supporting information

S1 Fig. Positions of three Mg^{2+} ions over a simulation time of 550 ns. The ions (colored in blue, green, and orange) explore the space around the Gsw aptamer (black) in a MD simulation of Gsw^{apt} with 20 Mg^{2+} ions.
(TIFF)

S2 Fig. Number of hexahydrated Mg^{2+} over the simulation time. Results are shown for three simulations of Gsw^{apt} (three blue, solid lines) and Gsw^{loop} (three red, dashed lines). A: Simulations with 12 Mg^{2+} ions; B: Simulations with 20 Mg^{2+} ions.
(TIFF)

S3 Fig. RMSD of all nucleotides of the aptamer with respect to the first conformation after fitting of the 80% least fluctuating nucleotides (S2 Table). The RMSD values for the three independent simulations are shown in different grey colors for Gsw^{apt} in the absence of Mg^{2+} (A) and the presence of 20 Mg^{2+} (B).
(TIFF)

S4 Fig. Average conformation calculated over the last 50 ns of the MD simulations. The average conformations for Gsw^{apt} in the absence of Mg^{2+} (A) and in the presence of 20 Mg^{2+} ions (B) and for Gsw^{loop} in the absence of Mg^{2+} (C) and in the presence of 20 Mg^{2+} ions (D) (colored in different grey tones for the three independent simulations) is overlaid with the crystal structure of Gsw^{apt} (PDB ID 4FE5 [17]) for (A) and (B) and of Gsw^{loop} (PDB ID 3RKF [18]) for (C) and (D) (colored in blue).
(TIFF)

S5 Fig. Distribution of χ dihedral for different force fields. Normalized histogram for simulations of simulations of Gsw^{apt} (A) and Gsw^{loop} (B) in the presence of 20 Mg^{2+} ions using the ff99 force field (blue) and the ff10 force field (red) for comparison. Different line types correspond to three independent simulations. Grey boxes correspond to χ dihedrals in the crystal structures with PDB ID 4FE5 [17] and 3RKF [18]. The simulations using the ff10 force field were extended from our previous study [33] to a simulation length 550 ns each.
(TIFF)

S6 Fig. Distribution of χ dihedrals in the P1 region of the aptamer for different force fields. Normalized histogram for simulations of Gsw^{apt} (A) and Gsw^{loop} (B) in the presence of 20

Mg^{2+} ions using the ff99 force field (blue) and the ff10 force field (red) for comparison. Different line types correspond to three independent simulations. Grey boxes correspond to χ dihedrals in the crystal structures with PDB ID 4FE5 (17) and 3RKF (18). The simulations using the ff10 force field were extended from our previous study [33] to a simulation length 550 ns each.
(TIFF)

S7 Fig. Lower base quadruple formed by the L2 and L3 loops. A: Gsw^{apt}; B: Gsw^{loop}. Hydrogen bonds are shown as black dashed lines. Bases are colored as in Fig 1A according to which loop they belong to.
(TIFF)

S8 Fig. Differences in RMSF in the presence of 20 Mg^{2+} ions for Gsw^{loop}—Gsw^{apt} projected onto the aptamer structure. Larger differences are colored red, smaller differences blue. Nucleotides responsible for ligand binding are shown as sticks.
(TIFF)

S9 Fig. Average difference (±SEM) in the probability ($\Delta p_{lrc}(i)$, Eq 1) of each nucleotide to be in the largest rigid cluster between the ligand being present or absent in the aptamer domain, using a value of $E_{HB} = -0.6$ kcal/mol for the rigidity analyses. The differences are shown in the absence of Mg^{2+} ions (red) or in the presence of 12 (green) and 20 (blue) Mg^{2+} ions for Gsw^{apt} (A) and Gsw^{loop} (B). The colored boxes on the top depict substructures of the riboswitch.
(TIFF)

S10 Fig. Projection of snapshots onto PC1 and PC2. The principal component analysis was performed over all MD trajectories. The trajectories for Gsw^{apt} with 20 Mg^{2+} ions (A, C) and Gsw^{loop} in the absence of Mg^{2+} ions (B, D) are projected onto the first (PC1) and second (PC2) eigenvector. The projections for the first (A, B) and second half (C, D) of the trajectories are shown separately. Histograms are shown on top and on the right side of the projections. Different blue and red colors correspond to the three independent MD simulations of each system.
(TIFF)

S11 Fig. Projection of snapshots onto PC1 and PC3. The principal component analysis was performed over all MD trajectories. The trajectories for Gsw^{apt} with 20 Mg^{2+} ions (A, C) and Gsw^{loop} in the absence of Mg^{2+} ions (B, D) are projected onto the first (PC1) and third (PC3) eigenvector. The projections for the first (A, B) and second half (C, D) of the trajectories are shown separately. Histograms are shown on top and on the right side of the projections. Different blue and red colors correspond to the three independent MD simulations of each system.
(TIFF)

S12 Fig. Projection of snapshots onto PC2 and PC3. The principal component analysis was performed over all MD trajectories. The trajectories for Gsw^{apt} with 20 Mg^{2+} ions (A, C) and Gsw^{loop} in the absence of Mg^{2+} ions (B, D) are projected onto the second (PC2) and third (PC3) eigenvector. The projections for the first (A, B) and second half (C, D) of the trajectories are shown separately. Histograms are shown on top and on the right side of the projections. Different blue and red colors correspond to the three independent MD simulations of each system.
(TIFF)

S13 Fig. RMSD average correlation (RAC). The RAC for trajectories of Gsw^{apt} in the absence (A), with 12 Mg^{2+} (C), and 20 Mg^{2+} ions (E), and of Gsw^{loop} in the absence (B), with 12 Mg^{2+}

(D), and 20 Mg^{2+} ions (F) is shown for time intervals up to 500 ns. The RAC was calculated with respect to the first conformation, omitting the first 50 ns of each trajectory.
(TIFF)

S14 Fig. Differences in the probability of each nucleotide to be in the largest rigid cluster between the ligand being present or absent in the aptamer domain, using a value of E_{HB} = −1.0 kcal/mol for the rigidity analyses. A, B: Average difference (±SEM). The differences are shown in the absence of Mg^{2+} ions (red) or in the presence of 12 (green) and 20 (blue) Mg^{2+} ions for Gsw^{apt} (A) and Gsw^{loop} (B). The colored boxes on the top depict substructures of the riboswitch. C, D, E, F: Average difference in the probability ($\Delta p_{lrc}(i)$, Eq 1) of each nucleotide to be in the largest rigid cluster between the ligand being present or absent in the aptamer domain, projected onto the aptamer domain of Gsw^{apt} (C, E) and Gsw^{loop} (D, F) in the absence of Mg^{2+} ions (C, D) and in the presence of 20 Mg^{2+} ions (E, F).
(TIFF)

S15 Fig. Rigid cluster decomposition of the ligand binding site (nucleotides U47, U51, C74). The blue sticks represent one rigid cluster obtained from the rigid cluster decomposition using the FIRST software. For this analysis, the X-ray structure of Gsw^{loop} was subjected to an energy minimization before the rigidity analysis. Here, only the binding nucleotides are shown. A: The binding site of Gsw^{loop} is empty. B: Guanine is present in the binding pocket. C: Constraints representing the bound ligand (c.f. Methods section "Rigidity analyses") were placed between the binding nucleotides.
(TIFF)

S1 Table. Occupation of experimentally determined Mg^{2+} binding sites.
(PDF)

S2 Table. Core nucleotides taken into account for analysis.
(PDF)

S3 Table. Root mean square deviations of Gsw^{apt} and Gsw^{loop} for substructural parts of the aptamer.
(PDF)

S4 Table. Root mean square fluctuations for Gsw^{apt} and Gsw^{loop} for the 80% least fluctuating nucleotides ("core nucleotides").
(PDF)

S5 Table. Assignment of nucleotides to substructural parts of the Gsw aptamer domain.
(PDF)

Acknowledgments

We gratefully acknowledge the computing time granted by the John von Neumann Institute for Computing (NIC) and provided on the supercomputer JUROPA at Jülich Supercomputing Center (JSC) (NIC project 4722, HDD08). Additional computational support was provided by the "Center for Information and Media Technology" (ZIM) at the Heinrich-Heine-University of Düsseldorf (Germany).

Author Contributions

Conceptualization: HG.

Data curation: CAH.

Formal analysis: CAH.

Funding acquisition: HG.

Investigation: CAH.

Methodology: CAH HG.

Project administration: HG.

Software: CAH.

Supervision: HG.

Validation: CAH HG.

Visualization: CAH.

Writing – original draft: CAH HG.

Writing – review & editing: CAH HG.

References

1. Mandal M, Breaker RR (2004) Gene regulation by riboswitches. Nature Reviews Molecular Cell Biology 5: 451–463. https://doi.org/10.1038/nrm1403 PMID: 15173824
2. Winkler WC, Breaker RR (2005) Regulation of bacterial gene expression by riboswitches. Annual Review of Microbiology 59: 487–517. https://doi.org/10.1146/annurev.micro.59.030804.121336 PMID: 16153177
3. Breaker RR (2011) Prospects for riboswitch discovery and analysis. Molecular Cell 43: 867–879. https://doi.org/10.1016/j.molcel.2011.08.024 PMID: 21925376
4. Serganov A, Patel DJ (2012) Metabolite recognition principles and molecular mechanisms underlying riboswitch function. Annual Review of Biophysics 41: 343–370. https://doi.org/10.1146/annurev-biophys-101211-113224 PMID: 22577823
5. Tucker BJ, Breaker RR (2005) Riboswitches as versatile gene control elements. Current Opinion in Structural Biology 15: 342–348. https://doi.org/10.1016/j.sbi.2005.05.003 PMID: 15919195
6. Stoddard CD, Montange RK, Hennelly SP, Rambo RP, Sanbonmatsu KY, Batey RT (2010) Free state conformational sampling of the SAM-I riboswitch aptamer domain. Structure 18: 787–797. https://doi.org/10.1016/j.str.2010.04.006 PMID: 20637415
7. Garst AD, Batey RT (2009) A switch in time: detailing the life of a riboswitch. Biochimica et Biophysica Acta 1789: 584–591. https://doi.org/10.1016/j.bbagrm.2009.06.004 PMID: 19595806
8. Batey RT, Gilbert SD, Montange RK (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. Nature 432: 411–415. https://doi.org/10.1038/nature03037 PMID: 15549109
9. Gilbert SD, Stoddard CD, Wise SJ, Batey RT (2006) Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. Journal of Molecular Biology 359: 754–768. https://doi.org/10.1016/j.jmb.2006.04.003 PMID: 16650860
10. Batey RT (2012) Structure and mechanism of purine-binding riboswitches. Quarterly Reviews of Biophysics 45: 345–381. https://doi.org/10.1017/S0033583512000078 PMID: 22850604
11. Porter EB, Marcano-Velazquez JG, Batey RT (2014) The purine riboswitch as a model system for exploring RNA biology and chemistry. Biochimica Et Biophysica Acta, Gene Regulatory Mechanisms 1839: 919–930.
12. de la Pena M, Dufour D, Gallego J (2009) Three-way RNA junctions with remote tertiary contacts: A recurrent and highly versatile fold. RNA 15: 1949–1964. https://doi.org/10.1261/rna.1889509 PMID: 19741022
13. Lescoute A, Westhof E (2006) Topology of three-way junctions in folded RNAs. RNA 12: 83–93. https://doi.org/10.1261/rna.2208106 PMID: 16373494
14. Buck J, Fürtig B, Noeske J, Wöhnert J, Schwalbe H (2007) Time-resolved NMR methods resolving ligand-induced RNA folding at atomic resolution. Proceedings of the National Academy of Sciences of the United States of America 104: 15699–15704. https://doi.org/10.1073/pnas.0703182104 PMID: 17895388

15. Buck J, Noeske J, Wöhnert J, Schwalbe H (2010) Dissecting the influence of Mg^{2+} on 3D architecture and ligand-binding of the guanine-sensing riboswitch aptamer domain. Nucleic Acids Research 38: 4143–4153. https://doi.org/10.1093/nar/gkq138 PMID: 20200045

16. Stoddard CD, Gilbert SD, Batey RT (2008) Ligand-dependent folding of the three-way junction in the purine riboswitch. RNA 14: 675–684. https://doi.org/10.1261/rna.736908 PMID: 18268025

17. Stoddard CD, Widmann J, Trausch JJ, Marcano-Velazquez JG, Knight R, Batey RT (2013) Nucleotides adjacent to the ligand-binding pocket are linked to activity tuning in the purine riboswitch. Journal of Molecular Biology 425: 1596–1611. https://doi.org/10.1016/j.jmb.2013.02.023 PMID: 23485418

18. Buck J, Wacker A, Warkentin E, Wöhnert J, Wirmer-Bartoschek J, Schwalbe H (2011) Influence of ground-state structure and Mg^{2+} binding on folding kinetics of the guanine-sensing riboswitch aptamer domain. Nucleic Acids Research 39: 9768–9778. https://doi.org/10.1093/nar/gkr664 PMID: 21890900

19. Brenner MD, Scanlan MS, Nahas MK, Ha T, Silverman SK (2010) Multivector fluorescence analysis of the xpt guanine riboswitch aptamer domain and the conformational role of guanine. Biochemistry 49: 1596–1605. https://doi.org/10.1021/bi9019912 PMID: 20108980

20. Gilbert SD, Love CE, Edwards AL, Batey RT (2007) Mutational analysis of the purine riboswitch aptamer domain. Biochemistry 46: 13297–13309. https://doi.org/10.1021/bi700410g PMID: 17960911

21. Nguyen PH, Derreumaux P, Stock G (2009) Energy flow and long-range correlations in guanine-binding riboswitch: A nonequilibrium molecular dynamics study. The Journal of Physical Chemistry B 113: 9340–9347. https://doi.org/10.1021/jp902013s PMID: 19569726

22. Allner O, Nilsson L, Villa A (2013) Loop-loop interaction in an adenine-sensing riboswitch: A molecular dynamics study. RNA 19: 916–926. https://doi.org/10.1261/rna.037549.112 PMID: 23716711

23. Di Palma F, Bottaro S, Bussi G (2015) Kissing loop interaction in adenine riboswitch: insights from umbrella sampling simulations. BMC Bioinformatics 16.

24. Di Palma F, Colizzi F, Bussi G (2013) Ligand-induced stabilization of the aptamer terminal helix in the add adenine riboswitch. RNA 19: 1517–1524. https://doi.org/10.1261/rna.040493.113 PMID: 24051105

25. Nozinovic S, Reining A, Kim YB, Noeske J, Schlepckow K, Wöhnert J, et al. (2014) The importance of helix P1 stability for structural pre-organization and ligand binding affinity of the adenine riboswitch aptamer domain. RNA Biology 11: 655–666. https://doi.org/10.4161/rna.29439 PMID: 24921630

26. Serganov A, Yuan YR, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, et al. (2004) Structural basis for discriminative regulation of gene expression by adenine-and guanine-sensing mRNAs. Chemistry & Biology 11: 1729–1741.

27. Serganov A, Nudler E (2013) A decade of riboswitches. Cell 152: 17–24. https://doi.org/10.1016/j.cell.2012.12.024 PMID: 23332744

28. Sanbonmatsu KY (2014) Dynamics of riboswitches: Molecular simulations. Biochimica Et Biophysica Acta, Gene Regulatory Mechanisms 1839: 1046–1050.

29. Liberman JA, Wedekind JE (2012) Riboswitch structure in the ligand-free state. Wiley interdisciplinary reviews RNA 3: 369–384. https://doi.org/10.1002/wrna.114 PMID: 21957061

30. Prychyna O, Dahabieh MS, Chao J, O'Neill MA (2009) Sequence-dependent folding and unfolding of ligand-bound purine riboswitches. Biopolymers 91: 953–965. https://doi.org/10.1002/bip.21283 PMID: 19603494

31. Noeske J, Buck J, Fürtig B, Nasiri HR, Schwalbe H, Wöhnert J (2007) Interplay of 'induced fit'and preorganization in the ligand induced folding of the aptamer domain of the guanine binding riboswitch. Nucleic Acids Research 35: 572–583. https://doi.org/10.1093/nar/gkl1094 PMID: 17175531

32. Ottink OM, Rampersad SM, Tessari M, Zaman GJR, Heus HA, Wijmenga SS (2007) Ligand-induced folding of the guanine-sensing riboswitch is controlled by a combined predetermined–induced fit mechanism. RNA 13: 2202–2212. https://doi.org/10.1261/rna.635307 PMID: 17959930

33. Hanke CA, Gohlke H (2015) Force Field Dependence of Riboswitch Dynamics. In: Chen S-J, Burke-Aguero DH, editors. Computational methods for understanding riboswitches. Burlington: Academic Press. pp. 163–191.

34. Woodson SA (2005) Metal ions and RNA folding: a highly charged topic with a dynamic future. Current Opinion in Chemical Biology 9: 104–109. https://doi.org/10.1016/j.cbpa.2005.02.004 PMID: 15811793

35. Draper DE, Grilley D, Soto AM (2005) Ions and RNA folding. Annual Review of Biophysics and Biomolecular Structure 34: 221–243. https://doi.org/10.1146/annurev.biophys.34.040204.144511 PMID: 15869389

36. Auffinger P (2012) Ions in molecular dynamics simulations of RNA systems. In: Leontis N, Westhof E, editors. RNA 3D Structure Analysis and Prediction. Berlin, Heidelberg: Springer. pp. 299–318.

37. Ohtaki H, Radnai T (1993) Structure and Dynamics of Hydrated Ions. Chemical Reviews 93: 1157–1204.

38. Harned HS, Hudson RM (1951) The Diffusion Coefficient of Magnesium Sulfate in Dilute Aqueous Solution at 25˚. Journal of the American Chemical Society 73: 5880–5882.

39. Mills R, Lobo VMM (1989) Self-diffusion in electrolyte solutions: a critical examination of data compiled from the literature. Amsterdam: Elsevier.

40. Reblova K, Spackova N, Stefl R, Csaszar K, Koca J, Leontis NB, et al. (2003) Non-Watson-Crick base-pairing and hydration in RNA motifs: Molecular dynamics of 5S rRNA loop E. Biophysical Journal 84: 3564–3582. https://doi.org/10.1016/S0006-3495(03)75089-9 PMID: 12770867

41. Reblova K, Lankas F, Razga F, Krasovska MV, Koca J, Sponer J (2006) Structure, dynamics, and elasticity of free 16S rRNA helix 44 studied by molecular dynamics simulations. Biopolymers 82: 504–520. https://doi.org/10.1002/bip.20503 PMID: 16538608

42. Villa A, Wöhnert J, Stock G (2009) Molecular dynamics simulation study of the binding of purine bases to the aptamer domain of the guanine sensing riboswitch. Nucleic Acids Research 37: 4774–4786. https://doi.org/10.1093/nar/gkp486 PMID: 19515936

43. Banas P, Sklenovsky P, Wedekind JE, Sponer J, Otyepka M (2012) Molecular mechanism of preQ$_1$ riboswitch action: A molecular dynamics study. Journal of Physical Chemistry B 116: 12721–12734.

44. Sharma M, Bulusu G, Mitra A (2009) 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: 1673–1692. https://doi.org/10.1261/rna.1675809 PMID: 19625387

45. Banas P, Hollas D, Zgarbova M, Jurecka P, Orozco M, Cheatham TE III, et al. (2010) Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. Journal of Chemical Theory and Computation 6: 3836–3849.

46. Mlynsky V, Banas P, Hollas D, Reblova K, Walter NG, Sponer J, et al. (2010) Extensive molecular dynamics simulations showing that canonical G8 and protonated A38H$^+$ forms are most consistent with crystal structures of hairpin ribozyme. Journal of Physical Chemistry B 114: 6642–6652.

47. Sklenovsky P, Florova P, Banas P, Reblova K, Lankas F, Otyepka M, et al. (2011) Understanding RNA Flexibility Using Explicit Solvent Simulations: The Ribosomal and Group I Intron Reverse Kink-Turn Motifs. Journal of Chemical Theory and Computation 7: 2963–2980. https://doi.org/10.1021/ct200204t PMID: 26605485

48. Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF (2001) Protein flexibility predictions using graph theory. Proteins: Structure, Function, and Genetics 44: 150–165.

49. Jacobs DJ, Thorpe MF (1995) Generic rigidity percolation: The pebble game. Physical Review Letters 75: 4051–4054. https://doi.org/10.1103/PhysRevLett.75.4051 PMID: 10059802

50. Jacobs DJ, Hendrickson B (1997) An algorithm for two-dimensional rigidity percolation: The pebble game. Journal of Computational Physics 137: 346–365.

51. Rader AJ, Hespenheide BM, Kuhn LA, Thorpe MF (2002) Protein unfolding: Rigidity lost. Proceedings of the National Academy of Sciences of the United States of America 99: 3540–3545. https://doi.org/10.1073/pnas.062492699 PMID: 11891336

52. Gohlke H, Kuhn LA, Case DA (2004) Change in protein flexibility upon complex formation: Analysis of Ras-Raf using molecular dynamics and a molecular framework approach. Proteins: Structure, Function, and Bioinformatics 56: 322–337.

53. Rader AJ (2010) Thermostability in rubredoxin and its relationship to mechanical rigidity. Physical Biology 7.

54. Pfleger C, Rathi PC, Klein DL, Radestock S, Gohlke H (2013) Constraint Network Analysis (CNA): a Python software package for efficiently linking biomacromolecular structure, flexibility, (thermo-)stability, and function. Journal of Chemical Information and Modeling 53: 1007–1015. https://doi.org/10.1021/ci400044m PMID: 23517329

55. Fulle S, Christ NA, Kestner E, Gohlke H (2010) HIV-1 TAR RNA spontaneously undergoes relevant apo-to-holo conformational transitions in molecular dynamics and constrained geometrical simulations. Journal of Chemical Information and Modeling 50: 1489–1501. https://doi.org/10.1021/ci100101w PMID: 20726603

56. Wang YM, Rader AJ, Bahar I, Jernigan RL (2004) Global ribosome motions revealed with elastic network model. Journal of Structural Biology 147: 302–314. https://doi.org/10.1016/j.jsb.2004.01.005 PMID: 15450299

57. Fulle S, Gohlke H (2009) Constraint counting on RNA structures: Linking flexibility and function. Methods 49: 181–188. https://doi.org/10.1016/j.ymeth.2009.04.004 PMID: 19398009

58. Fulle S, Gohlke H (2009) Statics of the Ribosomal Exit Tunnel: Implications for Cotranslational Peptide Folding, Elongation Regulation, and Antibiotics Binding. Journal of Molecular Biology 387: 502–517. https://doi.org/10.1016/j.jmb.2009.01.037 PMID: 19356596

59. Fulle S, Gohlke H (2008) Analyzing the flexibility of RNA structures by constraint counting. Biophysical Journal 94: 4202–4219. https://doi.org/10.1529/biophysj.107.113415 PMID: 18281388

60. Mamonova T, Hespenheide B, Straub R, Thorpe MF, Kurnikova M (2005) Protein flexibility using constraints from molecular dynamics simulations. Physical Biology 2: S137–S147. https://doi.org/10.1088/1478-3975/2/4/S08 PMID: 16280619

61. Rathi PC, Radestock S, Gohlke H (2012) Thermostabilizing mutations preferentially occur at structural weak spots with a high mutation ratio. Journal of Biotechnology 159: 135–144. https://doi.org/10.1016/j.jbiotec.2012.01.027 PMID: 22326626

62. Noeske J, Buck J, Wöhnert J, Schwalbe H (2009) Ligand Binding and Conformational Changes in the Purine-Binding Riboswitch Aptamer Domains. In: Walter NG, editor. Non-Protein Coding RNAs.

63. Cheatham TE III, Case DA (2013) Twenty-Five Years of Nucleic Acid Simulations. Biopolymers 99: 969–977. https://doi.org/10.1002/bip.22331 PMID: 23784813

64. Sponer J, Banas P, Jurecka P, Zgarbova M, Kuhrova P, Havrila M, et al. (2014) Molecular Dynamics Simulations of Nucleic Acids. From Tetranucleotides to the Ribosome. Journal of Physical Chemistry Letters 5: 1771–1782. https://doi.org/10.1021/jz500557y PMID: 26270382

65. Sponer J, Cang XH, Cheatham TE III (2012) Molecular dynamics simulations of G-DNA and perspectives on the simulation of nucleic acid structures. Methods 57: 25–39. https://doi.org/10.1016/j.ymeth.2012.04.005 PMID: 22525788

66. Wang JM, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? Journal of Computational Chemistry 21: 1049–1074.

67. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE III, Laughton CA, et al. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. Biophysical Journal 92: 3817–3829. https://doi.org/10.1529/biophysj.106.097782 PMID: 17351000

68. Zgarbova M, Otyepka M, Sponer J, Mladek A, Banas P, Cheatham TE III, et al. (2011) Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. Journal of Chemical Theory and Computation 7: 2886–2902. https://doi.org/10.1021/ct200162x PMID: 21921995

69. Aqvist J (1992) Modeling of Ion Ligand Interactions in Solutions and Biomolecules. Theochem-Journal of Molecular Structure 88: 135–152.

70. Chen SJ (2008) RNA folding: Conformational statistics, folding kinetics, and ion electrostatics. Annual Review of Biophysics 37: 197–214. https://doi.org/10.1146/annurev.biophys.37.032807.125957 PMID: 18573079

71. Baird NJ, Ferre-D'Amare AR (2010) Idiosyncratically tuned switching behavior of riboswitch aptamer domains revealed by comparative small-angle X-ray scattering analysis. RNA 16: 598–609. https://doi.org/10.1261/rna.1852310 PMID: 20106958

72. Cooper A, Dryden DTF (1984) Allostery without Conformational Change—a Plausible Model. European Biophysics Journal with Biophysics Letters 11: 103–109. PMID: 6544679

73. Dann CE, Wakeman CA, Sieling CL, Baker SC, Irnov I, Winkler WC (2007) Structure and mechanism of a metal-sensing regulatory RNA. Cell 130: 878–892. https://doi.org/10.1016/j.cell.2007.06.051 PMID: 17803910

74. Wakeman CA, Goodson JR, Zacharia VM, Winkler WC (2014) Assessment of the Requirements for Magnesium Transporters in Bacillus subtilis. Journal of Bacteriology 196: 1206–1214. https://doi.org/10.1128/JB.01238-13 PMID: 24415722

75. Reining A, Nozinovic S, Schlepckow K, Buhr F, Furtig B, Schwalbe H (2013) Three-state mechanism couples ligand and temperature sensing in riboswitches. Nature 499: 355–359. https://doi.org/10.1038/nature12378 PMID: 23842498

76. Furtig B, Nozinovic S, Reining A, Schwalbe H (2015) Multiple conformational states of riboswitches fine-tune gene regulation. Current Opinion in Structural Biology 30: 112–124. https://doi.org/10.1016/j.sbi.2015.02.007 PMID: 25727496

77. Whitty A (2008) Cooperativity and biological complexity. Nature Chemical Biology 4: 435–439. https://doi.org/10.1038/nchembio0808-435 PMID: 18641616

78. Gohlke H, Ben-Shalom IY, Kopitz H, Pfeiffer-Marek S, Baringhaus KH (2017) Rigidity Theory-Based Approximation of Vibrational Entropy Changes upon Binding to Biomolecules. Journal of Chemical Theory and Computation 13: 1495–1502. https://doi.org/10.1021/acs.jctc.7b00014 PMID: 28355485

79. Tsai CJ, del Sol A, Nussinov R (2008) Allostery: Absence of a change in shape does not imply that allostery is not at play. Journal of Molecular Biology 378: 1–11. https://doi.org/10.1016/j.jmb.2008.02.034 PMID: 18353365

80. van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, et al. (2006) Biomolecular modeling: Goals, problems, perspectives. Angewandte Chemie International Edition 45: 4064–4092. https://doi.org/10.1002/anie.200502655 PMID: 16761306

81. Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nature Chemical Biology 5: 593–599. https://doi.org/10.1038/nchembio.186 PMID: 19561621

82. Ottink OM, Westerweele IM, Tesssari M, Nelissen FHT, Heus HA, Wijmenga SS (2010) ^{1}H and ^{13}C resonance assignments of a guanine sensing riboswitch's terminator hairpin. Biomolecular NMR Assignments 4: 89–91. https://doi.org/10.1007/s12104-010-9215-z PMID: 20204714

83. Quarta G, Sin K, Schlick T (2012) Dynamic energy landscapes of riboswitches help interpret conformational rearrangements and function. PLoS Computational Biology 8: e1002368. https://doi.org/10.1371/journal.pcbi.1002368 PMID: 22359488

84. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM, et al. (2005) The Amber biomolecular simulation programs. Journal of Computational Chemistry 26: 1668–1688. https://doi.org/10.1002/jcc.20290 PMID: 16200636

85. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, et al. (2010) AMBER 11. University of California, San Francisco.

86. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins: Structure, Function, and Bioinformatics 65: 712–725.

87. Robinson H, Gao YG, Sanishvili R, Joachimiak A, Wang AHJ (2000) Hexahydrated magnesium ions bind in the deep major groove and at the outer mouth of A-form nucleic acid duplexes. Nucleic Acids Research 28: 1760–1766. PMID: 10734195

88. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. Journal of Chemical Physics 79: 926–935.

89. Aqvist J (1990) Ion Water Interaction Potentials Derived from Free-Energy Perturbation Simulations. Journal of Physical Chemistry 94: 8021–8024.

90. Gohlke H, Kiel C, Case DA (2003) Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. Journal of Molecular Biology 330: 891–913. PMID: 12850155

91. Darden T, York D, Pedersen L (1993) Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. Journal of Chemical Physics 98: 10089–10092.

92. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical Integration of Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. Journal of Computational Physics 23: 327–341.

93. Salomon-Ferrer R, Gotz AW, Poole D, Le Grand S, Walker RC (2013) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. Journal of Chemical Theory and Computation 9: 3878–3888. https://doi.org/10.1021/ct400314y PMID: 26592383

94. Roe DR, Cheatham TE III (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. Journal of Chemical Theory and Computation 9: 3084–3095. https://doi.org/10.1021/ct400341p PMID: 26583988

95. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, et al. (2012) AMBER 13. University of California, San Francisco.

96. Abseher R, Nilges M (1998) Are there non-trivial dynamic cross-correlations in proteins? Journal of Molecular Biology 279: 911–920. https://doi.org/10.1006/jmbi.1998.1807 PMID: 9654442

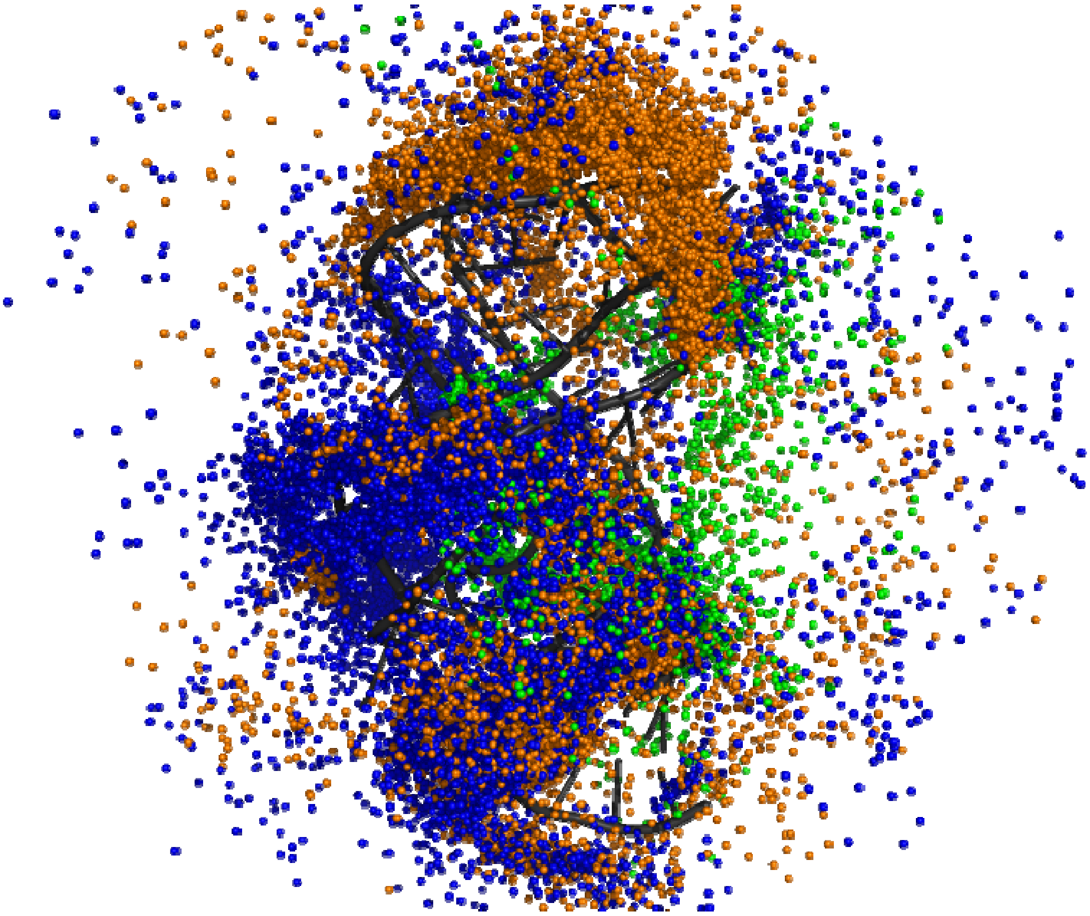
97. Hunenberger PH, Mark AE, van Gunsteren WF (1995) Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. Journal of Molecular Biology 252: 492–503. https://doi.org/10.1006/jmbi.1995.0514 PMID: 7563068

98. Karplus M, Ichiye T (1996) Comment on a "fluctuation and cross correlation analysis of protein motions observed in nanosecond molecular dynamics simulations". Journal of Molecular Biology 263: 120–122. https://doi.org/10.1006/jmbi.1996.0562 PMID: 8913294

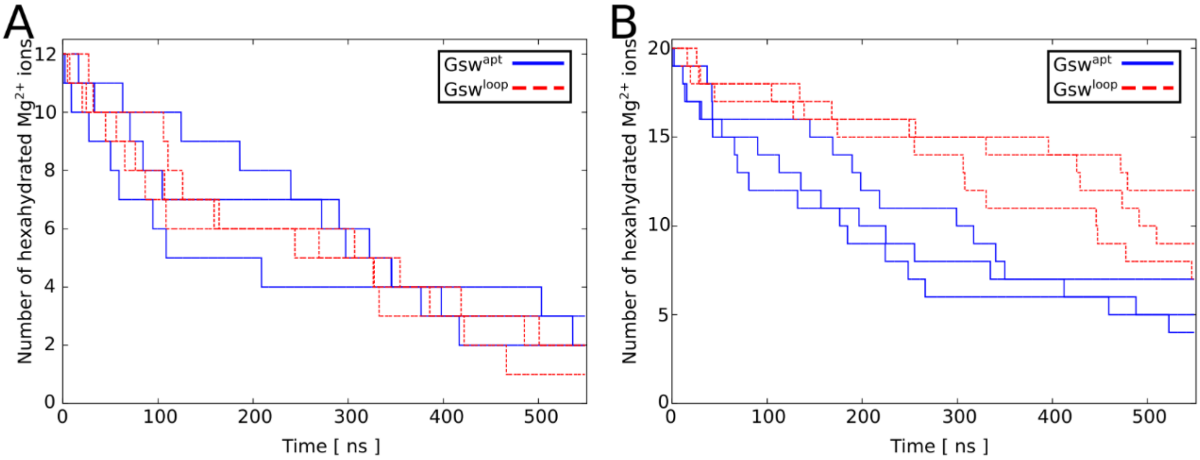
99. Galindo-Murillo R, Roe DR, Cheatham TE (2015) Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAACGAACGAACGC). Biochimica Et Biophysica Acta 1850: 1041–1058. https://doi.org/10.1016/j.bbagen.2014.09.007 PMID: 25219455

Supporting information

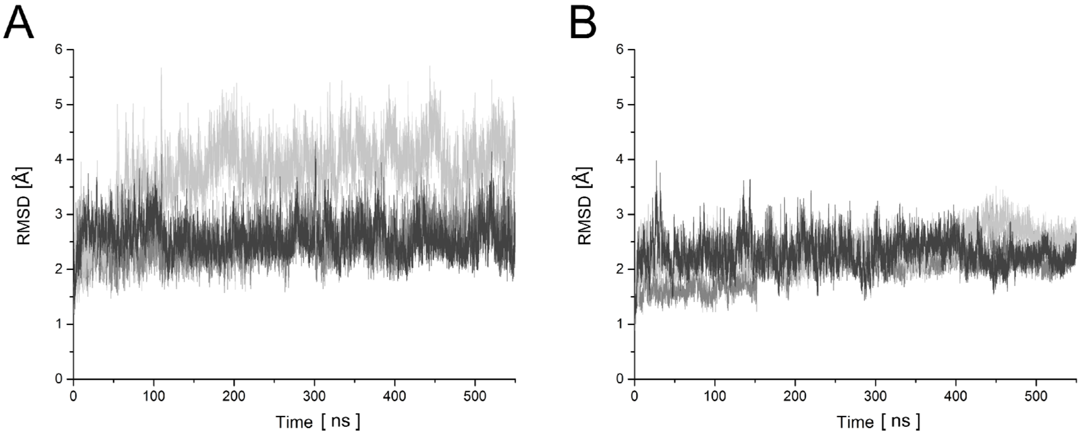
S1 Fig.



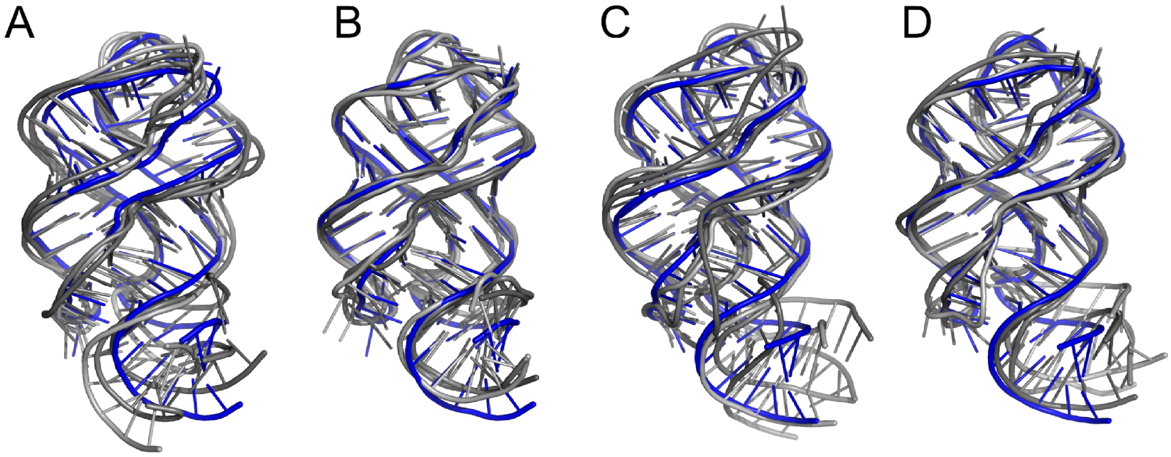
S2 Fig.



S3 Fig.

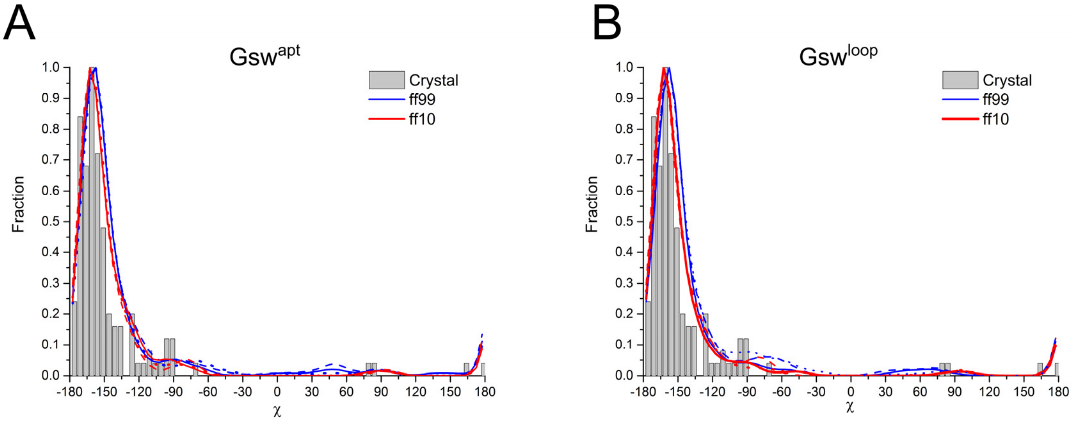


S4 Fig.

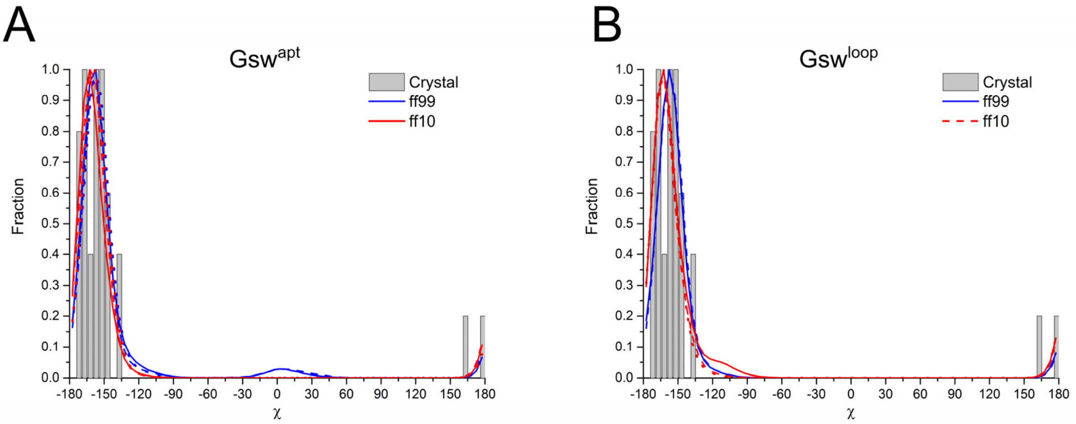


Gsw^{apt} 0 Mg^{2+} Gsw^{apt} 20 Mg^{2+} Gsw^{loop} 0 Mg^{2+} Gsw^{loop} 20 Mg^{2+}

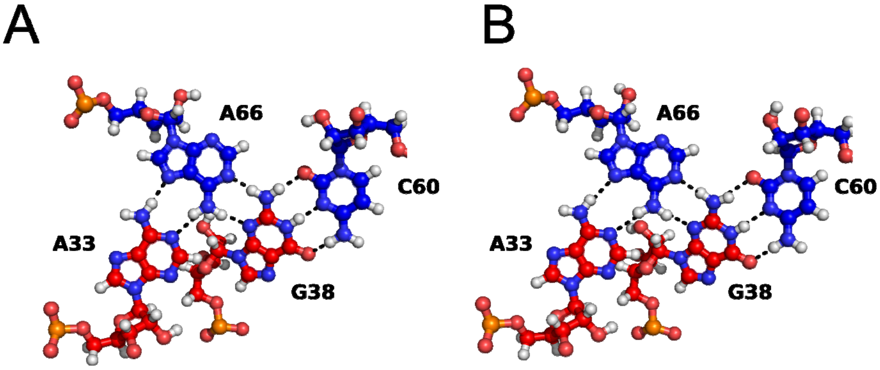
S5 Fig.



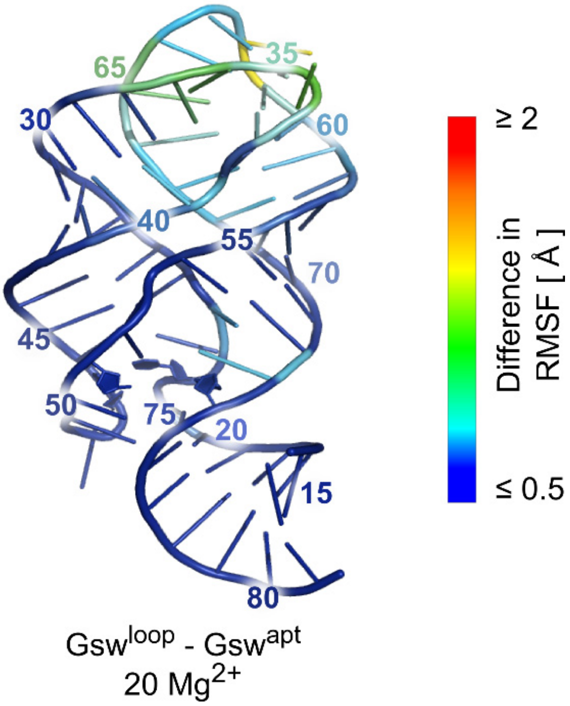
S6 Fig.



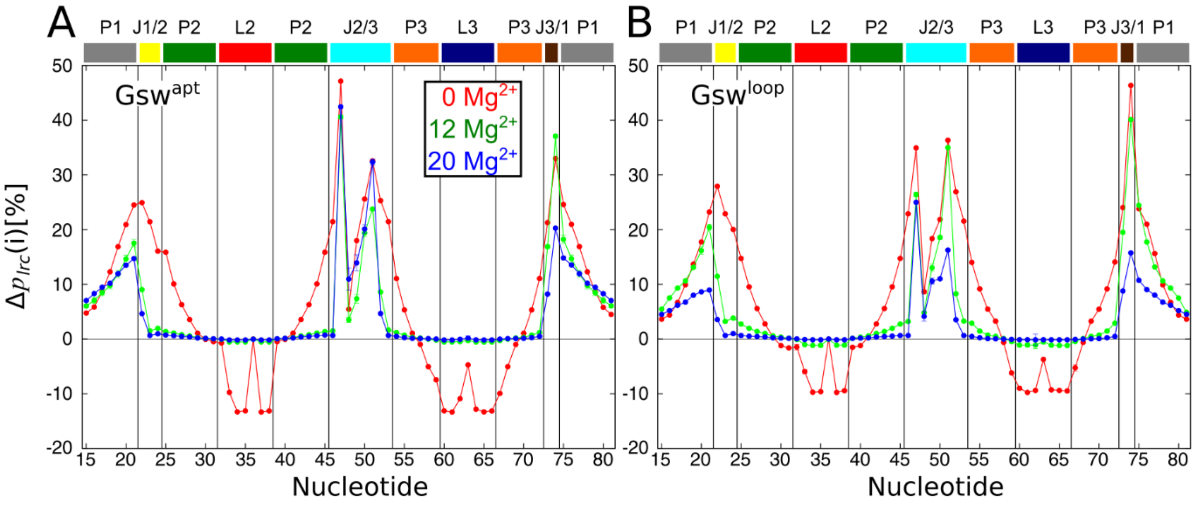
S7 Fig.



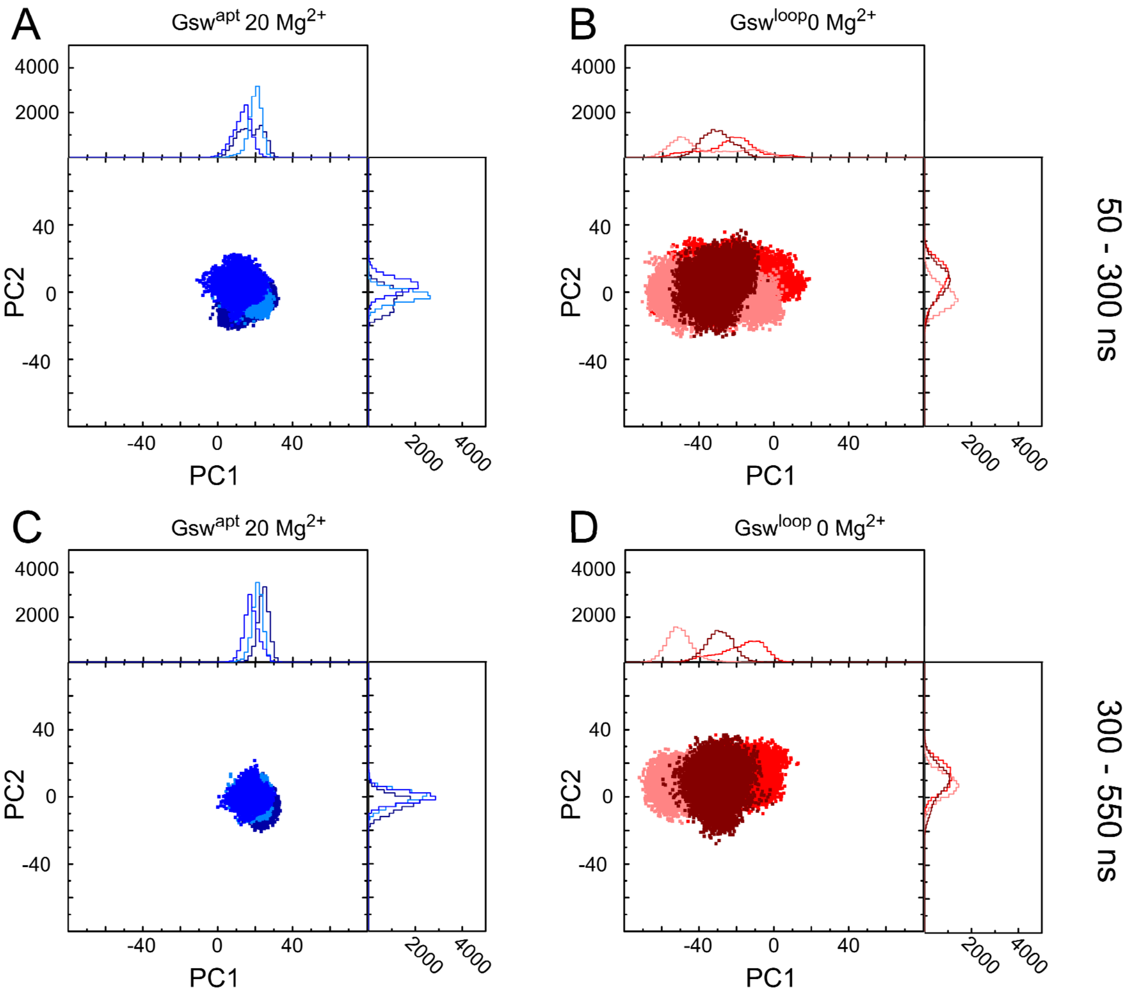
S8 Fig.



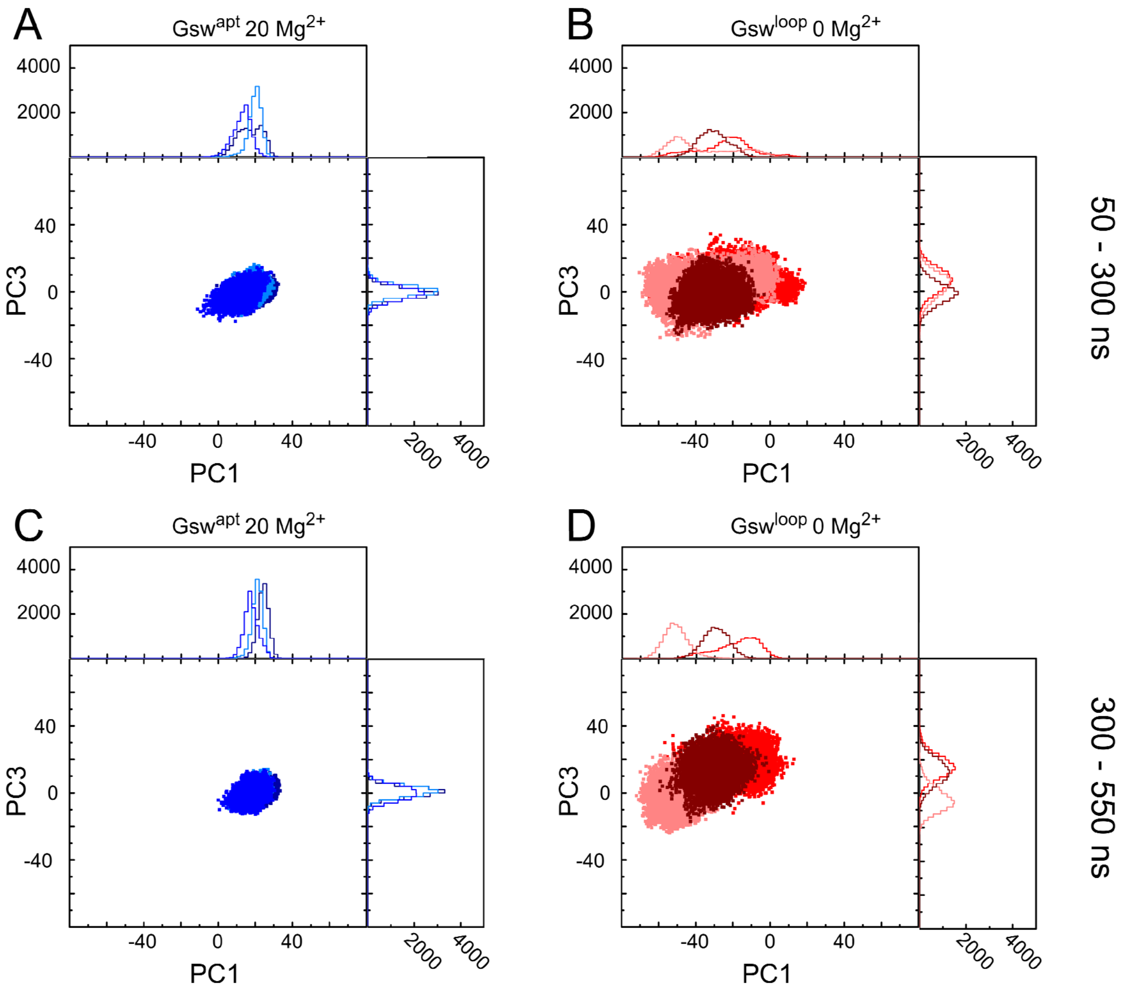
S9 Fig.



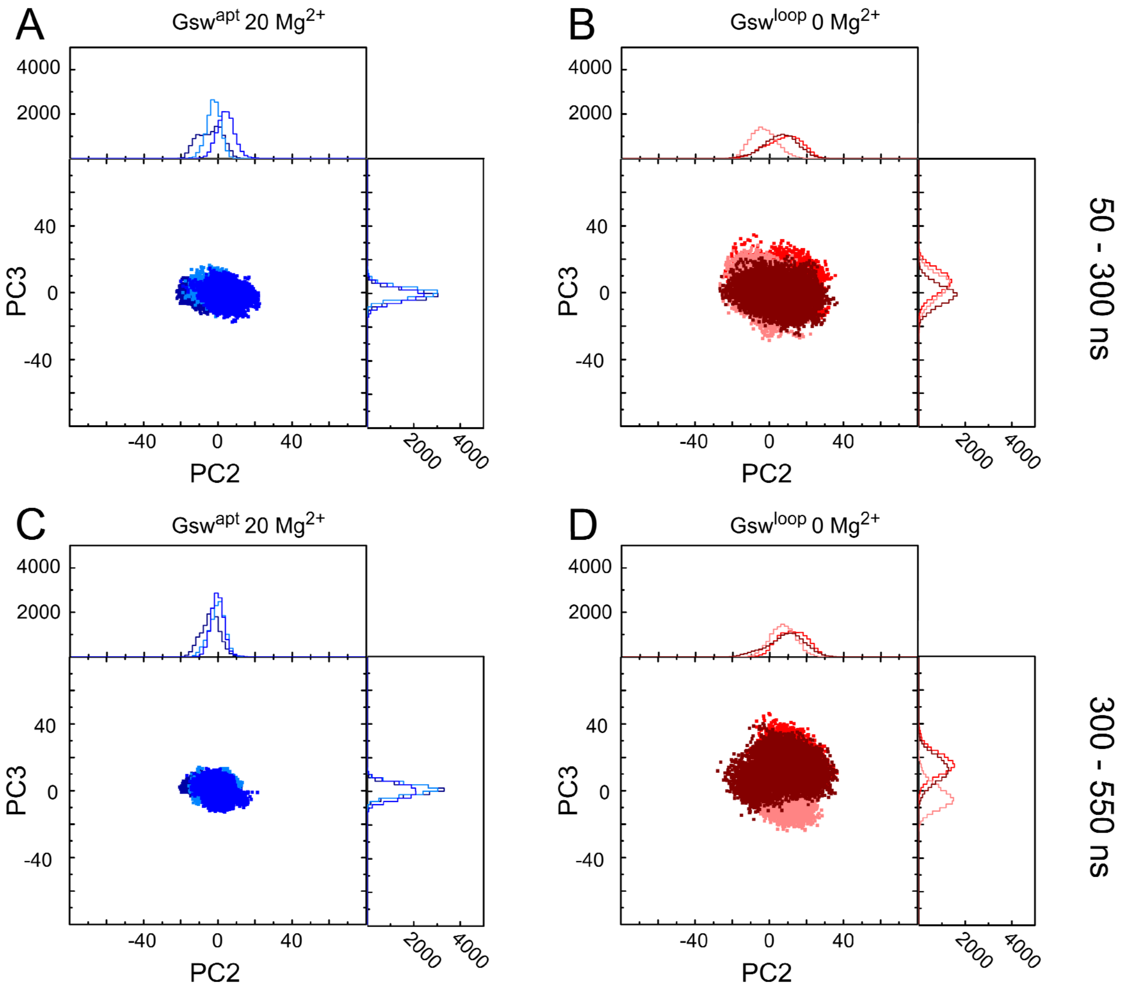
S10 Fig.



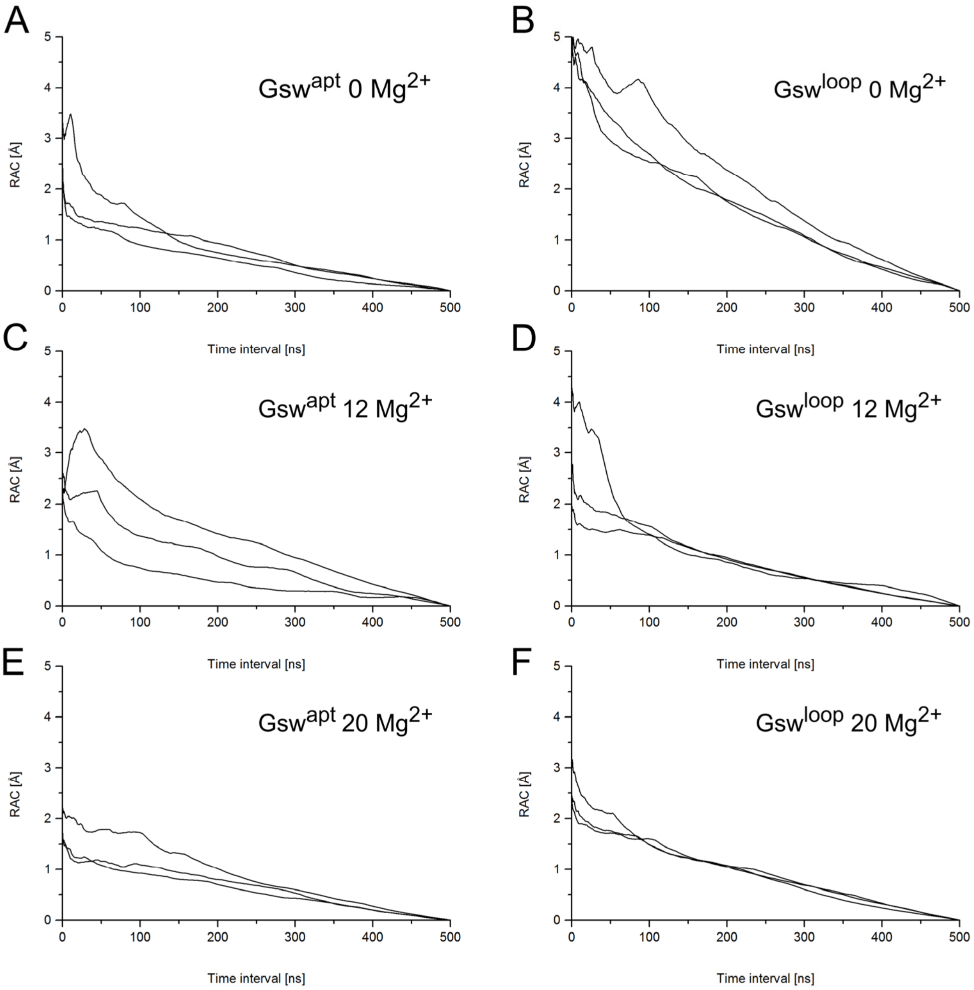
S11 Fig.


A
Gsw^apt 20 Mg2+
B
Gsw^loop 0 Mg2+
C
Gsw^apt 20 Mg2+
D
Gsw^loop 0 Mg2+
4000
2000
40
0
-40
PC3
PC1
2000
4000
50 - 300 ns
300 - 550 ns

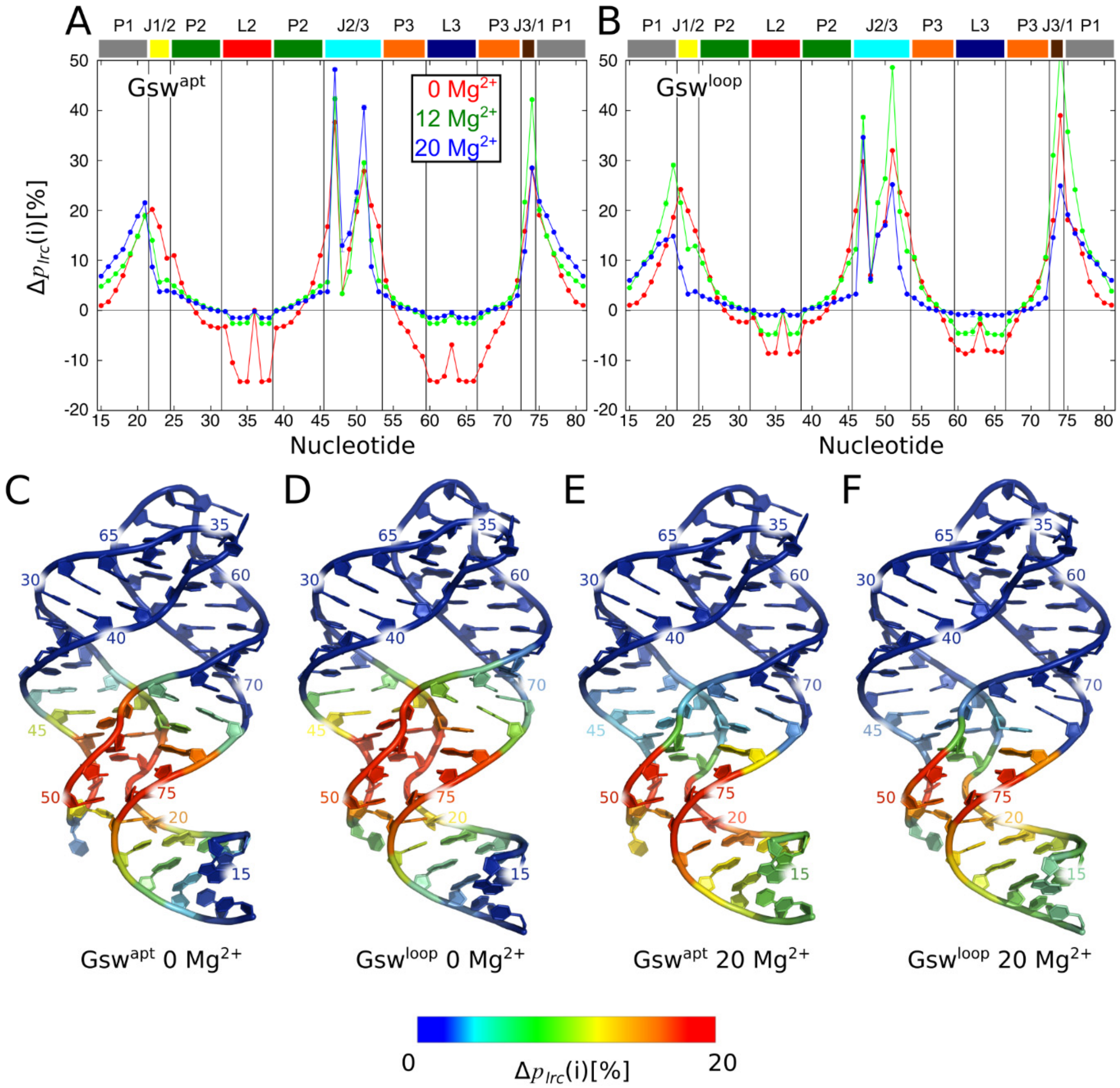
S12 Fig.


A
Gsw^{apt} 20 Mg^{2+}
B
Gsw^{loop} 0 Mg^{2+}
C
Gsw^{apt} 20 Mg^{2+}
D
Gsw^{loop} 0 Mg^{2+}
4000
2000
40
0
-40
PC3
PC2
2000
4000
50 - 300 ns
300 - 550 ns

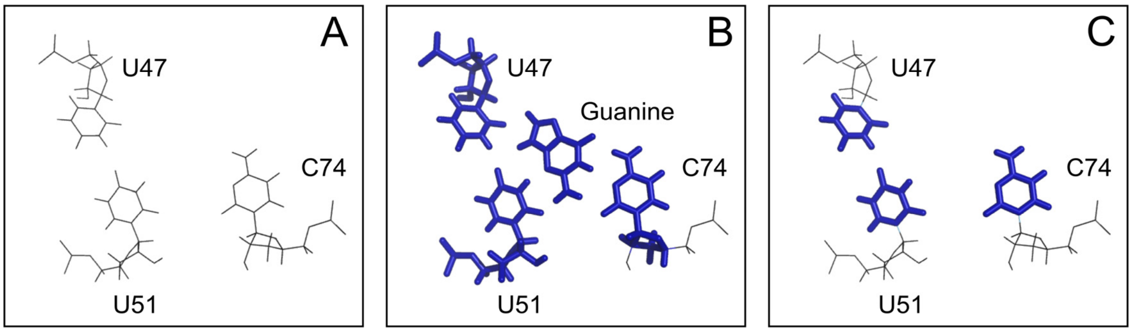
S13 Fig.


A
B
C
D
E
F
Gsw^{apt} 0 Mg^{2+}
Gsw^{loop} 0 Mg^{2+}
Gsw^{apt} 12 Mg^{2+}
Gsw^{loop} 12 Mg^{2+}
Gsw^{apt} 20 Mg^{2+}
Gsw^{loop} 20 Mg^{2+}
RAC [Å]
Time interval [ns]
0
1
2
3
4
5
0
100
200
300
400
500

S14 Fig.



S15 Fig.



S1 Table: Occupation of experimentally determined Mg^{2+} binding sites[a]

Simulation system		Mg^{2+} binding sites							
		1[b]	2[b]	3[b]	4[b]	5[b]	6[b]	7[c]	8[c]
Gsw^{apt} 12 Mg^{2+} random initial position[d]		9	29	74	25	40	17	56	57
Gsw^{apt} 12 Mg^{2+}	1	10	27	99	28	88	47	75	55
	2	27	40	99	99	77	31	97	0
	3	94	74	89	73	95	51	5	8
Gsw^{apt} 20 Mg^{2+}	1	60	97	100	97	96	79	95	9
	2	69	100	97	88	94	30	78	14
	3	27	65	82	98	75	56	99	15
Gsw^{loop} 12 Mg^{2+}	1	32	13	99	93	5	74	100	5
	2	7	100	100	33	89	6	25	3
	3	25	10	80	14	16	4	85	5
Gsw^{loop} 20 Mg^{2+}	1	31	67	100	62	35	61	98	16
	2	54	73	100	80	74	90	63	57
	3	100	55	99	88	82	25	98	19

[a] In %. A Mg^{2+} ion occupied a binding site if it was closer than 5 Å to the binding site.

[b] $[Co(NH_3)_6]^{3+}$ position in crystal structures with PDB IDs 4FE5 (1) and 3RKF (2).

[c] Additional $[Co(NH_3)_6]^{3+}$ position found in crystal structure with PDB ID 3RKF (2).

[d] Mg^{2+} ions were randomly placed at least 10 Å away from the RNA.

References

1. Stoddard, C.D., Widmann, J., Trausch, J.J., Marcano-Velazquez, J.G., Knight, R. and Batey, R.T. (2013) Nucleotides adjacent to the ligand-binding pocket are linked to activity tuning in the purine riboswitch. *J. Mol. Biol.*, **425**, 1596-1611.
2. Buck, J., Wacker, A., Warkentin, E., Wöhnert, J., Wirmer-Bartoschek, J. and Schwalbe, H. (2011) Influence of ground-state structure and Mg^{2+} binding on folding kinetics of the guanine-sensing riboswitch aptamer domain. *Nucleic Acids Res.*, **39**, 9768-9778.

S2 Table: Core nucleotides taken into account for analysis[a]

Assignment	**Nucleotide numbers**
80% least fluctuating nucleotides	19-35, 37-47, 50-61, 65-78

[a] Determined from the MD simulations of Gsw^{apt} in the absence of Mg^{2+} ions as the 80% least fluctuating nucleotides. The core nucleotides were used for fitting the structures prior to the calculation of the RMSD (Table 1) and the RMSF (Fig. 2, S8 Fig), and for calculating the average RMSF in S4 Table.

S3 Table: Root mean square deviations of Gsw$^{\text{apt}}$ and Gsw$^{\text{loop}}$ [a]

Simulated system		Aptamer [b]	P1[c]	P2[c]	P3[c]	L2[c]	L3[c]
Gsw$^{\text{apt}}$	0 Mg^{2+}	3.0 ± 0.1	2.2 ± 0.2	1.4 ± 0.1	1.6 ± 0.1	1.5 ± 0.1	2.3 ± 0.2
	12 Mg^{2+}	2.9 ± 0.1	1.7 ± 0.1	1.1 ± 0.1	1.4 ± 0.1	1.1 ± 0.1	2.0 ± 0.1
	20 Mg^{2+}	2.3 ± 0.1	1.6 ± 0.1	1.2 ± 0.1	1.1 ± 0.1	1.3 ± 0.1	1.8 ± 0.1
Gsw$^{\text{loop}}$	0 Mg^{2+}	4.2 ± 0.2	2.5 ± 0.2	1.5 ± 0.1	2.0 ± 0.1	2.5 ± 0.1	2.8 ± 0.1
	12 Mg^{2+}	2.4 ± 0.1	2.1 ± 0.2	1.3 ± 0.1	1.2 ± 0.1	2.3 ± 0.1	2.4 ± 0.1
	20 Mg^{2+}	2.5 ± 0.1	1.8 ± 0.1	1.3 ± 0.1	1.2 ± 0.1	1.5 ± 0.1	2.5 ± 0.2

Simulated system		J1/2[c]	J2/3[c]	J3/1[c]
Gsw$^{\text{apt}}$	0 Mg^{2+}	1.3 ± 0.1	2.7 ± 0.1	1.2 ± 0.1
	12 Mg^{2+}	1.2 ± 0.1	3.2 ± 0.2	1.5 ± 0.2
	20 Mg^{2+}	1.2 ± 0.1	2.7 ± 0.2	0.8 ± 0.1
Gsw$^{\text{loop}}$	0 Mg^{2+}	1.6 ± 0.1	3.4 ± 0.2	2.0 ± 0.2
	12 Mg^{2+}	1.3 ± 0.0	2.8 ± 0.2	1.3 ± 0.1
	20 Mg^{2+}	1.2 ± 0.1	2.5 ± 0.1	1.0 ± 0.1

[a] In Å; given is the mean ± SEM calculated over the three trajectories for each simulated system; the first 50 ns of each trajectory were omitted for the calculations.

[b] The complete RNA (nucleotides 15-81). The RMSD was calculated after root mean-square fitting of the conformations on the initial conformation, considering only those 80% of the nucleotides ("core nucleotides", S2 Table) that show the lowest RMSF.

[c] RMSD of the respective substructure after root mean-square fitting of the conformations on the initial conformation, considering only the nucleotides of the substructure.

S4 Table: Root mean square fluctuations for Gsw^{apt} and Gsw^{loop} [a]

Gsw variant	**0 Mg^{2+} [b]**	**12 Mg^{2+} [b]**	**20 Mg^{2+} [b]**
Gsw^{apt}	1.70 ± 0.02	1.39 ± 0.01	1.18 ± 0.02
Gsw^{loop}	2.27 ± 0.03	1.34 ± 0.02	1.39 ± 0.02

[a] In Å. First, the mean ± SEM RMSF for each of the 80% least fluctuating nucleotides ("core nucleotides", Table S2) was calculated over three trajectories of 550 ns, after fitting onto the first structure taking into account only the core nucleotides; the first 50 ns of each trajectory were omitted. The values reported here are mean values ± SEM over these core nucleotides.

[b] Number of Mg^{2+} ions per RNA molecule.

S5 Table: Assignment of nucleotides to substructural parts of the Gsw aptamer domain[a]

Substructural part	**Nucleotide numbers**
P1	15-21, 75-81
P2	25-31, 39-45
P3	54-59, 67-72
L2	32-38
L3	60-66
J1/2	22-24
J2/3	46-53
J3/1	73,74

[a] According to Batey *et al.* (1).

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

9.4 Publication III - Tertiary interactions in the unbound guanine-sensing riboswitch focus functional conformational variability to the binding site

Hanke C.A., Gohlke, H.

Submitted (2017).

Tertiary interactions in the unbound guanine-sensing riboswitch focus functional conformational variability on the binding site

Christian A. Hanke[1], Holger Gohlke[1,2]*

[1]Mathematisch-Naturwissenschaftliche Fakultät, Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany
[2]John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC) & Institute for Complex Systems - Structural Biochemistry (ICS 6), Forschungszentrum Jülich, 52425 Jülich, Germany

Running title: Role of tertiary interactions in the unbound guanine-sensing riboswitch

[*]**Address:** Universitätsstr. 1, 40225 Düsseldorf, Germany.
Phone: (+49) 211 81 13662; Fax: (+49) 211 81 13847
E-mail: gohlke@uni-duesseldorf.de , h.gohlke@fz-juelich.de

Keywords: Guanine-sensing riboswitch; unbound state; tertiary interactions; REMD; free energy landscape; potential of mean force

Abstract

Riboswitches are genetic regulatory elements mainly found in bacteria, which regulate gene expression based on the availability of a ligand. Purine-sensing riboswitches, including the guanine-sensing riboswitch (Gsw), possess tertiary interactions connecting the L2 and L3 loops. These interactions are important for ligand binding to the aptamer. However, atomic-level structural knowledge about the unbound state and how the tertiary interactions influence the conformational heterogeneity of the aptamer is still scarce. We performed replica exchange molecular dynamics simulations of the aptamer domain of wild type Gsw and a G37A/C61U mutant, which exhibits destabilized tertiary interactions, at different Mg^{2+} concentrations with an aggregate simulation time of ~16 μs, and subsequently obtained free energy landscapes. Our data provide evidence that suggests that the unbound state of wild type Gsw is conformationally rather homogeneous from a global view point, yet the ligand binding site shows functionally necessary mobility required for ligand binding. For the mutant, the data suggest a heterogeneous ensemble, in particular without Mg^{2+}. Hence, the tertiary interactions focus functional conformational variability on the binding site region of wild type Gsw. Our data allows speculating that already the weakening of the tertiary interactions by two hydrogen bonds shifts the kinetics of folding from downhill folding without traps or intermediate states for wild type Gsw to a folding including intermediates and misfolded structures for the mutant. A slowed-down folding of the aptamer might favor a decision during transcriptional regulation for the off-path, even if the ligand binds.

Introduction

Riboswitches are an elegant mechanism of gene regulation mostly occurring in the 5'-untranslated region of bacterial mRNA, because they regulate gene expression depending on ligand binding without the need for proteins or other components. A riboswitch typically consists of two domains, an aptamer domain which binds the ligand, and an expression platform, which undergoes a conformational change upon ligand binding[1-3]. The expression platform often contains part of a transcription terminator in the case of transcriptionally acting riboswitches. For these riboswitches, the aptamer has to fold and bind a ligand before the RNA polymerase passes the expression platform. Otherwise, the activity of the RNA polymerase cannot be regulated by the riboswitch.

Even though the unbound state of riboswitches plays a key role in the regulatory decision[4], its understanding on a structural level is still limited. This might be partially due to the marginally stable character of riboswitch aptamers, which, on the one hand, is required to allow for fast switching, but, on the other hand, hampers the resolution of structural properties in the unbound state. The so far resolved crystal structures of ligand-free riboswitch aptamers show only very little structural differences to the bound state[5-7]. It was suggested that this might be due to the crystal environment, which can select minor states from an ensemble of structures[7,8]. A recent study resolved structures of an unbound adenine-sensing riboswitch using a femtosecond X-ray free electron laser[9], which show differences to the bound state around the ligand binding site and at the region forming the connection to the expression platform.

Purine-sensing riboswitches, such as the guanine- and adenine-sensing riboswitches, are among the most studied riboswitches and were subject to a variety of experimental and computational studies [8,10]. The guanine- and adenine-sensing riboswitch aptamer domain consists of three paired helical regions (P1, P2, and P3), two loops (L2 and L3), which form tertiary interactions, and three joining regions (J1/2, J2/3, and J3/1), which connect the helical regions and form the ligand binding site [11] (Figure 1A). Ligands bound to the respective aptamer domain are deeply buried in the binding site (Figure 1E), which indicates that the binding site must be mobile for the ligand to bind [11]. Between L2 and L3, tertiary interactions are formed (Figure 1B-D), which differ in their stability between the adenine- and the guanine-sensing riboswitch; the interactions are more stable for the latter [10]. The tertiary interactions were found to play an important supporting role for ligand binding [12]. This finding is corroborated by mutation studies, which showed that the replacement of the loops by stable UUCG-tetraloops in the guanine-sensing riboswitch is detrimental to the riboswitch’s ligand binding ability [11].

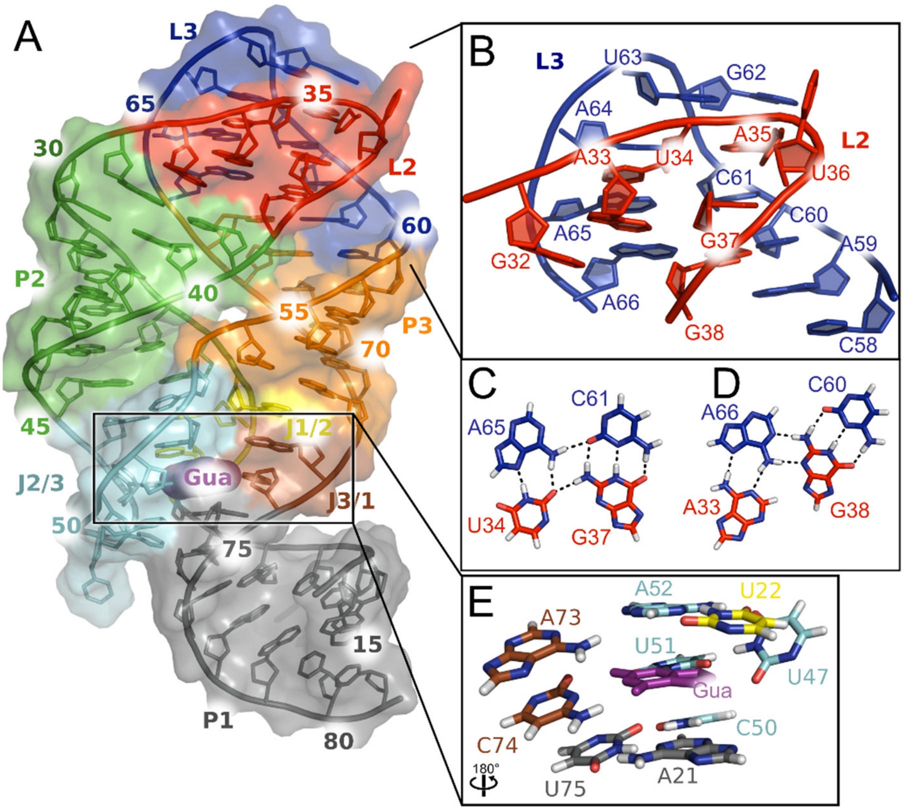


Figure 1. A: Structure of the aptamer domain of the wildtype guanine-sensing riboswitch bound to guanine (purple) [11, 60]. Secondary structure elements were assigned according to ref. [11]. The different secondary structure elements are color-coded. B: Loops L2 and L3. C: Upper base quadruple formed between L2 and L3. Only the nucleobases are shown. Carbon atoms are colored according to panel A. D: Lower base quadruple formed between L2 and L3. Only the nucleobases are shown. Carbon atoms are colored according to panel A. The base quadruples shown in panels C and D were rotated by 90° relative to panel B. E: Nucleobases forming the ligand-binding site. The colors of the carbon atoms are according to panel A. Three nucleotides (U47, U51, and C74) bind the guanine directly. Furthermore, one base triple (U22-A52-A73) closes the binding site to the top, and one base triple (A21-C50-U75) closes it to the bottom.

While for the adenine-sensing riboswitch a heterogeneous ensemble of structures is suggested for the unbound state [13,14], the guanine-sensing riboswitch aptamer (Gsw) is assumed to be more ordered [13]. This is supported by NMR investigations, in-line probing, and SHAPE experiments of Gsw that showed that the global fold of the aptamer domain is already present in the unbound state, with the P1, P2, and P3 regions and the interactions between the L2 and L3 loops preformed [15,16]. In contrast, the ligand binding core is generally assumed to be disorganized, even though the degree of the (dis)organization of the joining regions is discussed[12,15-17]. Overall, these results suggest that the stability of the tertiary interactions influences the conformational ensemble of the unbound state. However, insights at the atomistic level as to the characteristics of the unbound state of Gsw in the presence and absence of tertiary interactions are largely missing.

To investigate the influence of the tertiary interactions on the accessible conformational ensemble in the unbound state, here, we used two variants of the Gsw aptamer domain: The wildtype (Gsw^{apt}) from the *xpt-pbuX* operon of *Bacillus subtilis*, which possesses stable tertiary interactions in the loop region, and the G37A/C61U variant (Gsw^{loop}) [18, 19], in which the tertiary interactions are destabilized. Experimentally, Gsw^{loop} shows a Mg^{2+} dependence of the formation of its tertiary interactions and of its ligand binding ability, while these properties of the wildtype are mostly Mg^{2+} independent [18]. Thus, Gsw^{loop} is a very well suited model system as the stability of the tertiary interactions, and hence its structural and functional influence, can be tuned via the Mg^{2+} concentration [18, 19]. We performed replica exchange molecular dynamics (REMD) simulations for Gsw^{apt} and Gsw^{loop} in the absence and presence of Mg^{2+} ions with an aggregate simulation time of ~16.5 μs in order to overcome the limited sampling capability of standard molecular dynamics (MD) simulations in the context of slow RNA folding kinetics [20]. From the simulations, we computed the first free energy landscapes of the unbound Gsw aptamer at the atomistic level. Our results reveal that the tertiary interactions focus functional conformational variability on the binding site region. Our findings suggest that the destabilized loop region might slow down folding of the aptamer, which in turn could influence the regulatory decision.

Methods

Replica exchange molecular dynamics simulations

In order to overcome the limited sampling capabilities of standard molecular dynamics (MD) simulations, and to generate a conformational ensemble of the unbound state of the guanine-sensing riboswitch aptamer domain, we performed temperature Replica Exchange MD (REMD) simulations in explicit solvent using the Amber 11 suite of programs [43, 44] in conjunction with the force field ff99 [45]. We chose this force field based on our previous findings that the ff99 force field best reproduces the experimental data for our system [21].

In temperature REMD, copies of a simulations system are simulated in parallel at different temperatures. At regular intervals, exchanges between neighboring replica are attempted. An exchange between replica i and replica j is accepted with probability

$$\mathrm{P} = \min\{1,\ \exp[(\beta_\mathrm{i} - \beta_\mathrm{j})(\mathrm{U_i} - \mathrm{U_j})]\}$$

where U_i is the potential energy of replica i and temperature T_i, and $\beta_i = \frac{1}{k_B T_i}$ with k_B the Boltzmann constant [24, 46]. If an exchange attempt is successful, the momenta for the replica are scaled according to

$$p^{(i)\prime} = \sqrt{\frac{T_{new}}{T_{old}}}\, p^{(i)}$$

$p^{(i)\prime}$ are the new momenta after the exchange, $p^{(i)}$ are the momenta before the exchange, T_{old} is the temperature of the system before the exchange, and T_{new} is the new temperature after the exchange [24].

The range of temperatures used for the REMD simulations was determined considering experimental unfolding temperatures determined for Gsw$^{\mathrm{apt}}$ and Gsw$^{\mathrm{loop}}$ [18]. The highest temperature for the REMD simulations was chosen to be higher by ~10 K than the melting temperature determined for secondary structure elements in Gsw$^{\mathrm{apt}}$ [18]. The range of temperatures hence was 273.0 to 363.0 K, with the temperatures being distributed in a geometric progression, i.e., the same ratio was used to scale each temperature from the one below it, according to ref. [47] (273.0, 274.2, 275.5, 276.7, 278.0, 279.2, 280.5, 281.8, 283.1, 284.3, 285.6, 286.9, 288.2, 289.5, 290.8, 292.2, 293.5, 294.8, 296.2, 297.5, 298.8, 300.2, 301.6, 302.9, 304.3, 305.7, 307.1, 308.5, 309.9, 311.3, 312.7, 314.1, 315.5, 316.9, 318.4, 319.8, 321.3, 322.7, 324.2, 325.7, 327.1, 328.6, 330.1, 331.6, 333.1, 334.6, 336.1, 337.7, 339.2, 340.7, 342.3, 343.8, 345.4, 346.9, 348.5, 350.1, 351.7, 353.3, 354.9, 356.5, 358.1, 359.7, 361.4, and 363.0 K). A test simulation (data not shown) of 100 ns length of Gsw$^{\mathrm{apt}}$ in the absence of Mg^{2+} at

363.0 K showed that the overall fold of the aptamer domain did not remain stable, while most WC-base pairs and the tertiary interactions in the loop region remained formed.

The number of replica for the REMD simulations was chosen to ensure an average exchange rate between neighboring replica ~20-30% [23, 24]. Initial trials showed that 64 replica yielded a sufficiently high exchange rate between the replica (Table S1). Furthermore, at the given number of replica and the given temperature spacing, the potential energy distributions of neighboring replica overlap well (Figure S1, Table S2), which allows for successful exchanges.

Simulated systems and preparation of REMD simulations

The simulation system for the REMD simulations were prepared following the protocol used earlier [21, 22]. In detail, we set up two variants of the guanine-sensing riboswitch aptamer domains (Gsw): The wild type system, also referred to as Gsw^{apt}, and a G37A/C61U mutant, referred to as Gsw^{loop}. The starting structures for these two Gsw variants were obtained from ligand bound crystal structures. From these, the ligands were removed and the sequence of the wild type structure (PDB ID 1U8D/4FE5 [11]) was adapted to match the sequence of the G37A/C61U mutant (PDB ID 3RKF [19]), except for the nucleotides 37 and 61.

For both variants, two simulation systems were set up as described earlier [22] using different Mg^{2+} concentrations (0 and 20 Mg^{2+} ions per RNA molecule, respectively) based on experimental findings on the Mg^{2+} dependence of Gsw^{loop} stability [18, 19]. Following our earlier findings on the importance of initial Mg^{2+} ion placement, Mg^{2+} ions were added with a first hydration shell to allow for initial mobility of the ions, using the protocol described earlier [21]. Subsequently, the system was placed in an octahedral box of TIP3P water molecules [48], such that the distance between the edges of the boxes and the closest solute molecule was at least 11 Å. Na^+ ions were added as counter ions in order to neutralize the net charge of the simulation system, yielding a final system size of ~50,000 atoms.

The thermalization of the generated systems was done based on a protocol applied earlier [21, 22, 49]. First, the generated system was minimized by 200 steps of steepest descent minimization followed by 50 steps of conjugate gradient minimization. Subsequently, the system was initially heated from 100 K to 300 K performing canonical ensemble (NVT)-MD simulations for 50 ps. Then, isothermal-isobaric ensemble (NPT)-MD simulations were used for 50 ps to adjust the solvent density. During these first steps, harmonic restraints with a force constant of 5 kcal mol^{-1} $Å^{-2}$ were applied to all atoms of the RNA, as well as to the Mg^{2+} ions and their first hydration shell of water molecules. Next, this system was copied in order to serve as a starting point for all 64 replica. These 64 replica were then heated over 50 ps in NVT-MD simulations to their respective simulation temperatures. The harmonic restraints applied before were kept during this step. The harmonic restraints were then gradually reduced to 1 kcal mol^{-1} Å^{-2} over 250 ps, followed by 50 ps of unrestrained NVT-MD simulation. Each replica was then allowed to further equilibrate at its respective temperature for 5 ns, during which no REMD exchanges were attempted.

During the following 60 ns of production REMD simulations, exchanges between neighboring replica were attempted every 1 ps, and conformations were extracted every 10 ps. The REMD simulations were performed with the sander module of the Amber11 suite of molecular simulation programs [43, 44].

Trajectory analyses

The ff99 force field was found to yield ladder-like structures in long MD simulations of RNA systems [20, 32]. Since these structures represent non-physiological structures, the generated trajectories were screened for the occurrence of ladder-like structures (see also the Discussion for an assessment of the frequency of such structures). Conformations with ladder-like characteristics were excluded from further structural analyses.

In order to investigate structural ensembles found in local basins of the free energy landscape, we performed cluster analysis using the *cpptraj* program, which is part of the AmberTools program suite [50, 51]. We performed a clustering based on the RMSD of the phosphorus atoms of the nucleotides excluding the P1 region, where two structures were assigned to the same cluster if their pairwise RMSD was smaller than 2 Å. The P1 region was excluded due to its larger variability in order to focus on conformational changes occurring within the aptamer domain.

Hydrogen bond occupancies were calculated for conformational ensembles by the *cpptraj* program using a geometric criterion with a distance cutoff of 3.5 Å and an angle cutoff of 120°.

Q values

In order to describe the unfolding of biomolecules, an often used measure is the number of native contacts. Here, we followed the definition by Best et al. [25, 26] for Q as a measure of the number of native contacts

$$Q_s = N_s^{-1} \sum_{(i,j)} \frac{1}{1+exp\left(\gamma\left(r_{ij}-\lambda\, r_{ij}^0\right)\right)} \qquad \text{eq. 1}$$

with $\mathrm{N_s}$ the number of pairs (i,j) of native atomic contacts in a reference structure s, $\mathrm{r_{ij}}$ is the distance between the atoms i and j, and r_{ij}^0 is the distance of the contact in structure s.

$\gamma = 5$ Å^{-1}, and $\lambda = 1.5$ were chosen according to Best et al. [25]. In our case, the reference structure was the crystal structure. We calculated Q_{all} for all nucleotides of the aptamer, but also $Q_{loopinteractions}$ for the nucleotides forming the tertiary interactions (nucleotides 33, 34, 61, and 65 of the upper base quadruple and nucleotides 37, 38, 60, and 66 of the lower base quadruple Figure 1B, C, D), and $Q_{bindingsite}$ for the nucleotides forming the binding site (nucleotides 21, 22, 47, 50, 51, 52, 73, 74, 75; Figure 1E). Atom pairs are only considered if they are closer than a certain threshold in structure s and if the residues they belong to are separated by more than two in sequence. The thresholds for the atom pairs to be considered were 4.5 Å for Q_{all} and $Q_{bindingsite}$, and 3.5 Å for $Q_{loopinteractions}$ to increase the sensitivity for detecting a disturbance in the native hydrogen bond network in the latter case.

Melting curves

Melting curves for each REMD simulation were computed by collecting the structural ensembles at each temperature. For each structure in the ensemble, Q_{all} values were computed. A structure was considered folded if $Q_{all} > 0.75$. The fraction of folded structures was calculated as $\frac{N_f}{N}$ with N_f being the number of folded structures, and N being the total number of structures in the ensemble. We determined the melting temperature for the aptamer as the temperature at which the Q_{all} value reaches 50% of its initial value.

Two-dimensional potential of mean force

In order to make use of all the data obtained from the REMD simulations, i.e. from all replica simulated at different temperatures, a reweighting scheme can be applied to obtain unbiased results. That way an overall conformational ensemble is obtained that relates to one temperature only. Such a reweighting can be done with the weighted histogram analysis method (WHAM)[52,53] or the multistate Bennet acceptance ratio (MBAR) method [28, 54]. We used the MBAR method implemented in the *Pymbar* software[28] to reweight the conformational ensembles related to each replica. From the overall conformational ensemble, a two-dimensional potential of mean force (2D-PMF) was computed using the Pymbar software and $Q_{loopinteractions}$ and $Q_{bindingsite}$ as reaction coordinates, yielding a relative free energy ΔG for every bin of the 2D-PMF [55-58]. The uncertainties for the calculated 2D-PMFs were estimated with the *Pymbar* software[28]. We performed a worst-case estimation for the errors, for which only uncorrelated snapshots from the trajectories were considered. In this worst-case estimation, the *Pymbar* software yields errors for the free energy range investigated here (up to

5 kcal mol^{-1}) of below 2 kcal mol^{-1}. The errors computed by the *Pymbar* software for the data used in the 2D-PMFs was in general below 1 kcal mol^{-1}, and for the investigated free energy range below 0.7 kcal mol^{-1}.

Convergence of the REMD simulations

We calculated free energy landscapes from the REMD simulation results using a reweighting scheme implemented in the Pymbar software [28]. In order to investigate how well our REMD simulations are converged and which parts of the free energy landscape are reliably sampled, we calculated the free energy landscapes also for only the first and second halves of the trajectories (Figure S2). Subsequently, we visually investigated those areas that have been sampled in both parts of the trajectories; in particular, the basins that we investigate in Figure 2 have been sampled in both halves of the trajectories.

Definition of free energy basins

For the closer investigation of the energetic basins on the free energy landscape, we binned the relative free energies in steps of *kT* with *T* = 300 K. These steps have also been used to color the height of the relative free energy in Figure 2. A free energy basin was identified from the bin having the smallest free energy and, recursively applied, all neighboring pixels that belong to the same free energy bin. The free energy basins will be referred to in the form MX_Y^Z, where X is the number of the basin, Y is the number of Mg^{2+} ions in the simulation system, and Z is the Gsw variant (apt or loop) in the simulation system. The free energy basins shown in Figure 2 will thus be referred to as $\mathrm{M1}_0^{\mathrm{apt}}$, $\mathrm{M1}_{20}^{\mathrm{apt}}$, $\mathrm{M1}_0^{\mathrm{loop}}$, $\mathrm{M2}_0^{\mathrm{loop}}$, $\mathrm{M3}_0^{\mathrm{loop}}$, $\mathrm{M1}_{20}^{\mathrm{loop}}$, and $\mathrm{M1}_{20}^{\mathrm{loop}}$.

Intactness of the ligand binding site

In order to investigate structural deviations of the ligand binding site that may preclude ligand binding, we first calculated the RMSD of the nucleotides that bind the ligand in the native structure, i.e. nucleotides 22, 47, and 74. Furthermore, in order to estimate whether the ligand binding site would be large enough to accommodate a ligand, we calculated a volume-like measure. For this, we used the FPS tool [59], which computes the geometrically accessible volume (AV), given a starting point and an upper distance [59]. We set the starting point to be in the center of the ligand binding site, using the center of mass of nucleotides 47 and 74. From there, the number of accessible points up to a distance of 5 Å was calculated using a grid spacing of 0.5 Å. From comparison to the ligand-bound crystal structures, a binding site was considered intact, i.e., able to accommodate a ligand, if the RMSD of the binding nucleotides was < 3 Å and the number of accessible points in the ligand binding site was > 150. The initial structures for Gsw^{apt} and Gsw^{loop} yielded ~190 and ~230 accessible points in the ligand binding site, respectively. The RMSD threshold was chosen based on our previous observations on the RMSD in junction regions [22].

Calculation of reweighted histogram for χ torsion

In order to probe to what extent the use of the ff99 force field for our Gsw systems leads to a sampling of χ torsions in the high anti region [31-34], we calculated a weighted histogram of the χ dihedrals in the simulations. The reweighting was done for each sampled conformation in a bin (i, j) of the 2D PMF using a weight $w_{ij} = exp(-\frac{G_{ij}}{RT})$, where G_{ij} is the relative free energy in the 2D PMF that the conformation is assigned to.

Results

General REMD characteristics

We performed REMD simulations of the two unbound aptamer variants of Gsw in the absence and presence of Mg^{2+} ions in explicit solvent using 64 replica in the temperature range from 273 to 363 K. For this we applied force field parameters and system setups that had previously be shown to be very well suited to describe structural and dynamics changes for the Gsw systems [21, 22]. The upper temperature is ~20 K higher than the melting temperature of the secondary structure elements in Gsw^{apt} [18]. The REMD simulations showed exchange rates between neighboring replica of about 30% (Table S1), which is in the typically desired range[23,24] and related to well-overlapping potential energy distributions of neighboring replica (Figure S1, Table S2). To conclude, the number of replica and the temperature steps were appropriately chosen to successfully allow for sufficient replica exchanges.

The lack of two hydrogen bonds in the tertiary interactions destabilizes the tertiary fold of the aptamer domain

In order to investigate the deviation of a Gsw structure from the initial, X-ray crystal structure in the bound state, we computed the number of native contacts Q following the definition by Best et al. [25, 26] (eq. 1). The Q values were computed for the entire aptamer (Q_{all}) for each structural ensemble extracted from a respective replica temperature. From Q_{all}, the fraction of folded structures was determined considering a structure folded if $Q_{all} > 0.75$. Similar Q_{all} values to determine folded fractions have been used before [27].

In the absence of Mg^{2+} ions (Figure S3), we observe a decrease in the fraction of folded structures with increasing temperature for both Gsw^{apt} and Gsw^{loop}, with a particularly strong decline beyond temperatures of 320 and 310 K, respectively. The fraction of folded structures is generally higher by ~10-20% for a given temperature for Gsw^{apt} than for Gsw^{loop}. Our calculated melting temperature for Gsw^{apt} in the absence of Mg^{2+} ions of ~342 K is ~40 K higher than the experimentally determined melting temperature for the tertiary interactions of the aptamer [18]. For Gsw^{loop} in the absence of Mg^{2+} ions we observe a melting temperature of ~330 K, which is at variance with experiments where no melting temperature for the tertiary interactions could be determined [18]. These results show that, even though our simulations are likely not long enough to sample the states with broken tertiary interactions sufficiently when starting from Gsw structures derived from the bound crystal conformation, our simulation conditions still allow observing a difference in the overall stability of the aptamers caused by the destabilized tertiary interactions, as expected from experiment [18, 19].

The presence of Mg^{2+} has a stabilizing effect on both aptamers: The fraction of folded structures only decreases little now with increasing temperature (~10% between 273 K and 363 K) (data not shown). In contrast to our simulation results, experiments were able to determine a difference in the melting temperature for Gsw^{apt} and Gsw^{loop} in the presence of Mg^{2+} ions of ~30 K [18]. Still, in our simulations, Gsw^{apt} yields a larger fraction of folded structures (between 3 and 25% higher) than Gsw^{loop} in the presence of Mg^{2+} ions (data not shown). Hence, even though our results do not match the experimental values quantitatively, we observe qualitatively that Gsw^{apt} is more stable than Gsw^{loop}.

To conclude, the lack of two hydrogen bonds due to the exchange of the Watson-Crick base pair G37-C61 in the upper base quadruple of Gsw^{apt} by the Watson-Crick base pair A37-U61 in Gsw^{loop} results in an overall pronounced destabilization of the tertiary fold.

The lack of two hydrogen bonds in the tertiary interactions changes the number and distribution of free energy basins on the free energy landscape of the unbound aptamer

In order to investigate the influence of destabilized tertiary interactions on the free energy landscape of the aptamer domain in the unbound state, we calculated a two-dimensional potential of mean force (2D-PMF) by applying a reweighting scheme on the REMD results using the Pymbar software [28]. As we are primarily interested in how the stability of the tertiary interactions influences the ligand binding ability of the aptamer, we chose the number of native contacts within the nucleotides forming the tertiary interactions in the loop region ($Q_{loopinteractions}$) and the number of contacts within the ligand binding site ($Q_{bindingsite}$) as reaction coordinates.

For Gsw^{apt} in the absence of Mg^{2+} ions (Figure 2A), we observe one prominent free energy basin ($M1_0^{apt}$, indicated by 1 in Figure 2A) on the free energy landscape. This basin is located at high $Q_{loopinteractions}$ values of ~0.98, suggesting that the tertiary interactions in the loop region are mostly conserved. However, we observe a spread in the $Q_{bindingsite}$ values between ~0.44 and 0.90 in the basin. Interestingly, there are no notable energy barriers that separate conformations with $Q_{bindingsite}$ > ~0.4 within this free energy basin, and the overall landscape (also considering smaller $Q_{loopinteractions}$) has a rather smooth, funnel-like character towards this basin [29].

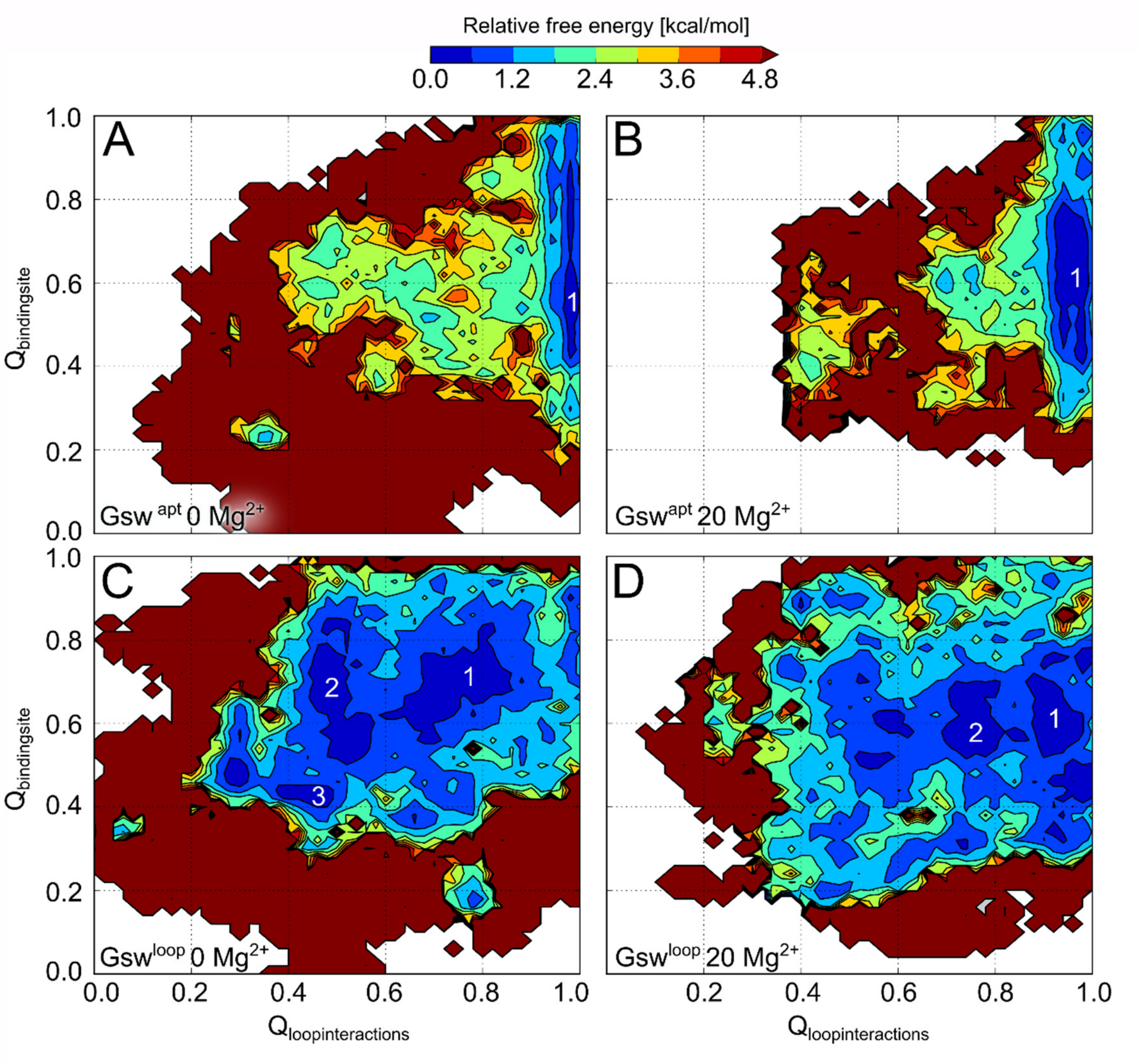


Figure 2: 2D-PMF computed for the reaction coordinates $Q_{\text{bindingsite}}$ and $Q_{\text{loopinteractions}}$ for Gsw$^{\text{apt}}$ in the absence and presence of 20 Mg^{2+} ions per RNA molecule (A and B, respectively) and for Gsw$^{\text{loop}}$ in the absence and presence of 20 Mg^{2+} ions per RNA molecule (C and D, respectively). The white numbers identify free energy basins. The color scale represents the relative free energies with respect to the global basin. White areas were not sampled in the simulations.

We observe a similar picture for Gsw$^{\text{apt}}$ in the presence of Mg^{2+} ions (Figure 2B). Here, there is one prominent free energy basin ($M1_{20}^{apt}$, indicated by 1 in Figure 2B), which is slightly wider with respect to $Q_{loopinteractions}$ (values between 0.92 and 0.99), which suggests that the presence of Mg^{2+} ions allows for the formation of other than or the loss of some of the native tertiary interactions in the loop region, while in the absence of Mg^{2+} ions this appears energetically more unfavorable. However, despite the broader free energy basin, the adjacent regions of higher free energy are less extended, revealing a steeper yet equally smooth funnel towards $M1_{20}^{apt}$.

In contrast, for Gsw$^{\text{loop}}$ in the absence of Mg^{2+} ions (Figure 2C), we observe a free energy landscape that is flatter, making about 1/3 of the free energy landscape energetically equal to within 3 kT at $T = 300$ K. There are at least three prominent free energy basins, which are almost equally low ($M1_0^{loop}$, $M2_0^{loop}$, and $M3_0^{loop}$, indicated by 1, 2, and 3 in Figure 2C). However, in contrast to Gsw^{apt}, none of these is associated with $Q_{loopinteractions}$ values considerably larger than 0.85, and all of them show $Q_{bindingsite}$ values almost always < 0.8. Furthermore, a local basin at ($Q_{loopinteractions}$, $Q_{bindingsite}$) ≈ (0.95, 0.9) is higher in energy by 1 kT and separated from $M1_0^{loop}$, $M2_0^{loop}$, and $M3_0^{loop}$ by a barrier of about 3 kT.

For Gsw^{loop} in the presence of Mg^{2+} ions (Figure 2D), we observe, similarly to the absence of Mg^{2+} ions, an extended region on the free energy landscape that is energetically favorable. However, the most prominent basins ($M1_{20}^{loop}$ and $M2_{20}^{loop}$ marked by 1 and 2 in Figure 2D) are located closer to that region where the free energy basins $M1_0^{apt}$ and $M1_{20}^{apt}$ for Gsw^{apt} were found. In particular, $M1_{20}^{loop}$ can be reached now from almost all regions of the extended low free energy region without the need to cross barriers.

Overall, for Gsw^{apt} the free energy landscape is dominated by one free energy basin close to the native state, which is restricted in its extension to high $Q_{loopinteractions}$, but at the same time exhibits variability in $Q_{bindingsite}$. In contrast, the lack of two hydrogen bonds in the tertiary interactions in Gsw^{loop} results in a flatter free energy landscape with multiple free energy basins further away from the native state. The presence of Mg^{2+} ions shifts the free energy basins for Gsw^{loop} towards the native state.

Structural analysis shows a conformationally homogeneous ensemble for Gsw^{apt} and a heterogeneous ensemble for Gsw^{loop}

In order to gain insight into the structural variability within the local basins, we performed a cluster analysis on the structural ensemble in the free energy basins using as a cluster criterion an RMSD of the phosphorus atoms of 2 Å, excluding the highly mobile P1 region. See section "Definition of free energy basins" in the Methods section for how the boundaries of a free energy basin were defined. For Gsw^{apt}, both in the absence and presence of Mg^{2+} ions, we observe one big cluster each for $M1_0^{apt}$ and $M1_{20}^{apt}$ (containing almost all ~106,400 and ~210,000 structures in that basin, respectively) (Table S3). Thus, all structures within one basin are structurally highly similar in this case. For Gsw^{loop} in the absence of Mg^{2+} ions, both $M1_0^{loop}$ and $M2_0^{loop}$ also show a rather homogeneous structural ensemble: For $M1_0^{loop}$, one large cluster (containing ~49,400 structures) and three smaller clusters (containing < 1,500 structures) were found; for $M2_0^{loop}$, a somewhat more heterogeneous distribution with one large cluster (~28,700 structures) and several smaller ones (one containing ~2,200 structures and two containing ~1,000 structures each) was found. An even more heterogeneous ensemble is found for the basin $M3_{0}^{loop}$ with two almost equally populated clusters (each containing ~2,500 structures) and several smaller clusters. Thus, this basin contains structures that largely differ from each other despite similar $Q_{loopinteractions}$ and $Q_{bindingsite}$ values as well as similar conformational free energies. For Gsw^{loop} in the absence of Mg^{2+} ions, the cluster analysis for $M1_{20}^{loop}$ and $M2_{20}^{loop}$ resulted in one large cluster each (containing ~40,600 and ~39,600 structures, respectively).

To investigate whether alternative base pairs are formed by structures located in the free energy basins, we calculated hydrogen bond occupancies within the respective structural ensembles (Figure S4). The hydrogen bonds were determined based on geometric criteria (see section "Trajectory analysis" in the Methods section for details). Generally, we observe that the Watson-Crick base pairs in the L2-L3 region (between nucleotides 37 and 61, and 38 and 60) are formed in most structures in all identified free energy basins. Furthermore, for Gsw^{apt}, both in the absence and presence of Mg^{2+} ions, the native hydrogen bonds (Figure S4A) are maintained over the structural ensembles (Figure S4B, C). Similarly, for Gsw^{loop} (Figure S4D) in the presence of Mg^{2+} ions, in $M1_{20}^{loop}$ most native hydrogen bonds are formed (Figure S4H). In contrast, in $M2_{20}^{loop}$, the base pair between nucleotides 61 and 65 is only present in about 16% of the structures (Figure S4I). For Gsw^{loop} in the absence of Mg^{2+} ions, the native hydrogen bond network is destabilized, as expected for this system [18, 19] (Figure S4E-G). In particular, the structures in $M3_{0}^{loop}$ show the lowest hydrogen bond occupancies (Figure S4G). Additionally, we observe alternative base pairs for $M3_{0}^{loop}$ including nucleotides from the native base quadruples and other nucleotides in the loops (base pairs 62-65 and 34-64) or between the two base quadruples (base pair 61-66).

To investigate whether the ligand binding site is intact in the structural ensembles of the basins (Table S3), we combined RMSD calculations with volume calculations for the binding site. Both measures provide an indication whether a structure would be ligand binding-competent: The RMSD threshold of 3 Å was chosen regarding the conformational variability in the junction region observed in our previous study [22]. Furthermore, in the bound conformation with the ligand removed, the ligand binding site contains ~190 and ~230 “accessible points” computed for Gsw^{apt} and Gsw^{loop}, respectively. We use these “accessible points” (with a spacing of 0.5 Å) to estimate whether a ligand binding site could enclose a guanine molecule. Hence, we consider volumes with more than 150 accessible points as an indicator for ligand binding capability. We observe that for $M1_{0}^{apt}$ more than half of the structures (~55,400 structures) in the free energy basin have a ligand binding site that is intact according to these criteria. For $M1_{20}^{apt}$, around 79,300 structures (~40% of the structures in the basin) have a ligand binding site that is intact. In contrast, for $\mathrm{Gsw^{loop}}$ in the absence of $\mathrm{Mg^{2+}}$ ions, in $M3_{0}^{loop}$ only ~150 structures (~1.5 % of the structures in the basin) have a ligand binding site that is intact. For $M1_{0}^{loop}$ and $M2_{0}^{loop}$, ~27,200 (~55% of the structures in the basin) and ~9,800 (~30% of the structures in the basin) structures have intact ligand binding sites, respectively. The presence of $\mathrm{Mg^{2+}}$ ions in the case of $\mathrm{Gsw^{loop}}$ results in a larger number of structures with intact ligand binding site (~21,000 (~50% of the structures in the basin) and ~16,700 (~42% of the structures in the basin) structures for $M1_{20}^{loop}$ and $M2_{20}^{loop}$, respectively).

Visual inspection of the representative structures of the clusters (Figure 3, Table S3) reveals that the structure for $\mathrm{Gsw^{apt}}$ in the absence of $\mathrm{Mg^{2+}}$ ions is slightly less compact than the crystal structure (Figure 3A,B). This is likely due to the lack of charge compensation, in particular between the negatively charged RNA backbones, as observed before by us [21] [22]. However, the overall structure of the representative is highly similar to the initial structure with an RMSD of the phosphorus atoms with respect to the crystal structure of 2.5 Å. The cluster representative of $\mathrm{Gsw^{apt}}$ in the presence of $\mathrm{Mg^{2+}}$ ions (Figure 3C) is even more similar, with an RMSD of only 1.6 Å. For $\mathrm{Gsw^{loop}}$ in the absence of $\mathrm{Mg^{2+}}$ ions, we obtained three prominent free energy basins. Within $M1_{0}^{loop}$, the cluster analysis resulted in one cluster, where the cluster representative has an RMSD value with respect to the crystal structure of around 2.5 Å (Figure 3E). The most notable differences are, as for $\mathrm{Gsw^{apt}}$, a less compact structure reflected in the larger distance between the backbones. In $M2_{0}^{loop}$, which is more distant to the native structure on the free energy landscape (Figure 2C), we obtained a cluster representative with a notably different fold in the tertiary interactions (RMSD ≈ 2.6 Å; Figure 3F). In contrast, for $M3_{0}^{loop}$, we obtained three almost equally populated clusters with cluster representatives that show larger structural differences to the crystal structure in many regions of the aptamer domain, resulting in RMSD values of 3.4 to 5 Å (Figure 3G-I). Here, neither the region of the loop interactions nor the ligand binding site are native-like anymore. Finally, for $\mathrm{Gsw^{loop}}$ in the presence of $\mathrm{Mg^{2+}}$ ions, the overall structures are more native-like than in the absence of $\mathrm{Mg^{2+}}$ ions (Figure 3J, K). In particular, the overall fold is similar again to the crystal structure (RMSD of 1.5 and 1.6 Å for $M1_{20}^{loop}$ and $M2_{20}^{loop}$).

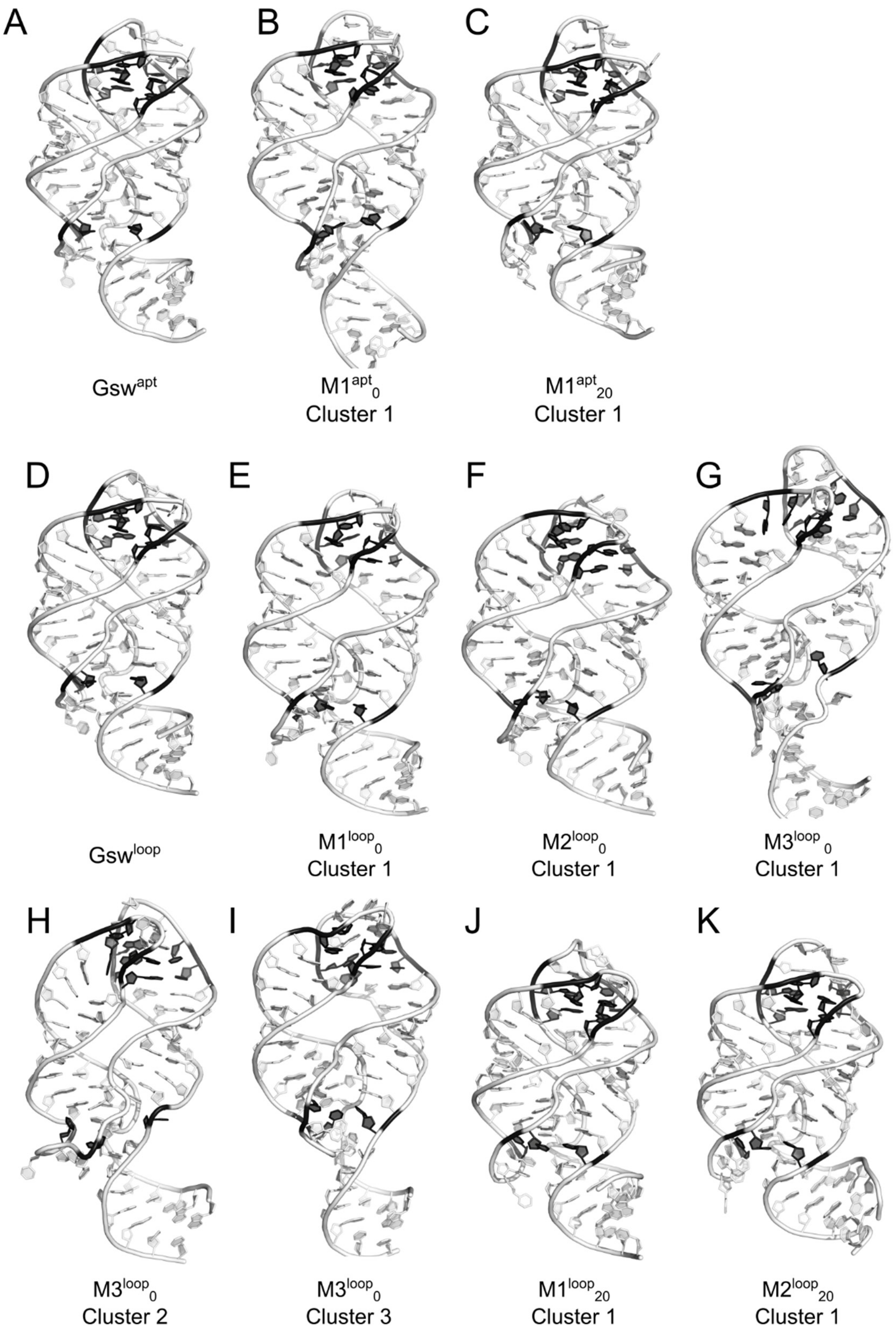


Figure 3: Structures of the cluster representatives from free energy basins. The nucleotides involved in the formation of the tertiary interactions in the loop region and the nucleotides binding the ligand are colored in black (top black nucleotides and bottom black nucleotides, respectively). A: Initial structure of Gsw^{apt}, which is highly similar to the crystal structure with

PDB ID 4FE5 [11, 61]; B: Cluster representative of the largest cluster from $\boldsymbol{M1_{0}^{apt}}$ for Gsw$^{\mathrm{apt}}$ in the absence of Mg^{2+} ions; C: Cluster representative of the largest cluster from $\boldsymbol{M1_{20}^{apt}}$ for Gsw$^{\mathrm{apt}}$ with 20 Mg^{2+} ions per RNA molecule; D: Initial structure of Gsw$^{\mathrm{loop}}$, corresponding to the crystal structure with PDB ID 3RKF [19]; E-I: Cluster representatives for Gsw$^{\mathrm{loop}}$ in the absence of Mg^{2+} ions. Cluster representatives for the largest clusters for $\boldsymbol{M1_{0}^{loop}}$ (E) and $\boldsymbol{M2_{0}^{loop}}$ (F), and for $\boldsymbol{M3_{0}^{loop}}$ for the largest (G), the second-largest (H) and the third-largest cluster (I) are shown. J, K: Cluster representatives for Gsw$^{\mathrm{loop}}$ in the presence of 20 Mg^{2+} ions for the largest cluster in $\boldsymbol{M1_{20}^{loop}}$ (J) and $\boldsymbol{M2_{20}^{loop}}$ (K).

In conclusion, we observe a homogeneous conformational ensemble for Gsw$^{\mathrm{apt}}$, in which most structures are similar to the ligand-bound crystal structure. In contrast, for Gsw$^{\mathrm{loop}}$ conformations within the same free energy basins show a larger conformational heterogeneity. Overall, the absence of the stable hydrogen bond network in the loop region results in conformations that strongly differ from the native state. In particular, this also strongly impedes the formation of an intact binding site, in agreement with the missing ligand binding ability of Gsw$^{\mathrm{loop}}$ [18, 19].

Discussion

In this study, we investigated the influence of the extent of tertiary interactions and Mg^{2+} ions on the conformational heterogeneity of the unbound aptamer domain of the guanine-sensing riboswitch. For this, we performed REMD simulations to enhance the sampling of the conformational space compared to standard MD simulations and subsequently calculated 2D free energy landscapes with respect to the closeness to the native state in the loop region and the ligand binding site. We found that the tertiary interactions focus functional conformational variability on the binding site region. I.e., while the tertiary interactions in the loop region foster Gsw conformations to be ligand binding-competent, they allow for variability in the ligand binding site, which is required for ligand binding. Furthermore, the tertiary interactions and Mg^{2+} ions influence the conformational heterogeneity of the aptamer domain and the funneledness of the free energy landscape and, hence, might impact the folding kinetics of the aptamer domain.

The simulation setup with respect to the choice of the force field and the treatment of Mg^{2+} ions was based on our previous findings [21, 22] that the Amber ff99 force field [45] and Mg^{2+}

parameters of Åqvist [30] yielded the best agreement with experimental observations compared to more recent force field / parameter variants. However, the ff99 force field was reported to yield "ladder-like" structures in long nucleic acid simulations caused by a shift of the χ dihedral to the *high anti* region [31-34]. Investigating the conformational Gsw ensembles from the REMD simulations with respect to sampled χ torsions revealed a low occurrence frequency of χ torsions in the *high anti* region (Figure S5). However, as noted earlier [21, 22], also the crystal structures of the Gsw systems contain nucleotides with χ torsions in that region, and often this torsion angle value is found in functionally relevant regions [11, 19]. Still, to reduce a possible influence on our analyses due to a force field artefact, we removed such conformations prior to the structural analyses.

As to the REMD simulations, the selected number of replica and their distribution in temperature space ensured a sufficient exchange between replica, which is a requirement for optimal sampling of the conformational space [35, 36]. Furthermore, comparing free energy landscapes generated with only the first or second half of sampled conformations revealed that the free energy landscapes are sufficiently converged in those regions that we further investigate (Figure S2), although in the remaining regions convergence was not achieved despite an aggregate MD simulation time of ~4.1 μs per system because of the slow and complex (un-)folding kinetics of RNA [37]. Thus, we focused our analyses of the free energy landscapes on the converged areas.

As detailed experimental information on the unbound state of Gsw and in particular for the mutant for comparison to simulation results is largely missing, we computed melting curves from our conformational ensembles to compare those to curves obtained from CD melting experiments [18]. Our curves correctly display at a qualitative level that $\mathrm{Gsw}^{\mathrm{apt}}$ is in general more stable than $\mathrm{Gsw}^{\mathrm{loop}}$, also at higher temperatures. Furthermore, we observed that the presence of Mg^{2+} ions stabilizes both aptamers, albeit $\mathrm{Gsw}^{\mathrm{apt}}$ still remains more stable than $\mathrm{Gsw}^{\mathrm{loop}}$. This result is in line with the known stabilizing role of Mg^{2+} ions for RNA molecules [9, 38] and, in particular, for Gsw [18]. However, our curves did not allow for a quantitative determination of melting points. Likely, our simulations are not long enough to sample sufficiently, e.g., states with completely broken tertiary interactions, even in the case of the least stable system, $\mathrm{Gsw}^{\mathrm{loop}}$ in the absence of Mg^{2+} ions.

The computed 2D free energy landscapes for unbound $\mathrm{Gsw}^{\mathrm{apt}}$ reveal one prominent basin near the conformation found in the bound state ($Q_{loopinteractions} \approx 0.95$) irrespective of the presence or absence of Mg^{2+}. However, Mg^{2+} influences the steepness of the free energy landscape, as the energy barrier of ~1 kcal mol^{-1} located between $Q_{loopinteractions} \approx 0.7$ and

0.8 in the absence of Mg^{2+} is shifted towards $Q_{loopinteractions}$ ≈ 0.8 in the presence of Mg^{2+}, or even vanishes, resulting in a steeper and/or smoother descent towards the basin. Furthermore, Mg^{2+} restricts sampling of the conformational space with lower $Q_{loopinteractions}$, fostering a steeper free energy landscape towards the main basin. Notably, with respect to $Q_{bindingsite}$, the basin is located at 0.42 – 0.88. First, this finding demonstrates that the binding site in the absence of the ligand does not retain its bound conformation, although the structural changes are small as only one cluster of conformations is found in that basin. Second, this finding demonstrates a certain degree of mobility in the binding site region in the unbound state. Both aspects are in line with previous reports that stressed the necessity of a flexible binding site because the ligand in the crystal structure is deeply buried [11] and ligand binding is assumed to happen via opening of the ligand binding site at the J2/3 region [8, 12, 16, 17]. Taken together, our results, in connection with a lack of (large) free energy barriers in the direction of $Q_{bindingsite}$ but also in the direction of $Q_{loopinteractions}$, suggest that Gsw^{apt} is designed to allow for a quick access to the ligand binding site without the need for large conformational rearrangements. In view of our previous findings that ligand binding information might be transferred via changes in the dynamics to the terminal P1 region and, subsequently, to the expression platform [22], our results thus stress again the importance of changes in the dynamics over conformational changes for the function of Gsw^{apt}. Furthermore, the steepness and smoothness of the free energy landscape suggests a downhill folding pathway without energetic traps or intermediates. A directed and fast folding would be in agreement with the short time available for regulation by the transcriptionally acting guanine-sensing riboswitch [39, 40]. Finally, the presence of Mg^{2+} ions allows for larger variations in $Q_{loopinteractions}$ within the energy basin. Apparently, interactions between Gsw^{apt} and Mg^{2+} can compensate for non-optimal formation of the tertiary interactions in the loop region.

In contrast, for Gsw^{loop} in the absence of Mg^{2+} ions, the free energy landscape is flatter, with several energetically equally favorable free energy basins. These basins are more widespread and located further away from the native bound state than the energy basin in Gsw^{apt}. A barrier of ≥ 1.2 kcal mol^{-1} at $Q_{loopinteractions}$ ≈ 0.9 separates basin $M1_0^{loop}$ from the native state. Structures within the basins are globally similar to the bound conformation but strongly differ from this conformation with respect to RMSD, native hydrogen bonds, and characteristics of the ligand binding site. The presence of several free energy basins results in a flatter and more rugged free energy landscape, which suggests that the folding of Gsw^{loop} is less straight forward compared to Gsw^{apt}. In particular, Gsw^{loop} might encounter several folding traps, intermediate states, or misfolded structures [41], which will make the structural ensemble of Gsw^{loop} more heterogeneous. This finding is in agreement with a more heterogeneous ensemble for Gsw^{loop} in the absence of Mg^{2+} ions described based on NMR experiments [19]. The presence of Mg^{2+} ions partially "rescues" the Gsw^{loop} aptamer in that it shifts the population more towards the native bound conformation. However, compared to Gsw^{apt}, still the presence of multiple basins persists, again resulting in a flatter and more rugged free energy landscape. This finding is in line with the experimental observation that Mg^{2+} partially compensates the destabilized tertiary interactions in the G37A/C61U mutant with respect to ligand binding and the formation of the tertiary interactions in the loop region [18, 19].

In summary, our data provide evidence that suggests that the unbound state of Gsw^{apt} is conformationally rather homogeneous from a global point of view, although the ligand binding site shows the (functionally necessary) mobility required for ligand binding. For Gsw^{loop}, which lacks the stable tertiary interactions in the loop region, our data suggest a more heterogeneous ensemble, in particular in the absence of Mg^{2+}. In that respect, Gsw^{loop} resembles the adenine-sensing riboswitch, which requires Mg^{2+} ions for the formation of stable tertiary interactions and for which a more heterogeneous unbound state is assumed [13-15]. Furthermore, our data allows to speculate that already the weakening of the tertiary interactions by two hydrogen bonds shifts the kinetics of folding from a quick downhill folding without traps or intermediate states for Gsw^{apt} to a slower folding including intermediates and misfolded structures for Gsw^{loop}. A slowed-down folding of the aptamer might be sufficient to favor a decision during transcriptional regulation for the off-path even if the ligand binds [39, 40]. It would be interesting to probe this prediction at the cellular level. Further complexity of the free energy landscapes may arise when considering the full-length guanine-sensing riboswitch because experimental studies showed the presence of additional (transient) states when investigating the aptamer domain and the expression platform [7, 40, 42]. The lack of detailed structural knowledge of the expression platform hampers computational studies of the full-length riboswitch at the atomistic level, however. Despite the discussed limitations in obtaining fully converged simulations, the above aspects demonstrate that enhanced sampling techniques can provide valuable insights at the atomistic level in the conformational heterogeneity of unbound aptamers. Such knowledge may be useful for the development of novel antibiotics targeting riboswitches, with alternative binding site conformations as potential drug targets.

Acknowledgements

We gratefully acknowledge the computing time granted by the John von Neumann Institute for Computing (NIC) and provided on the supercomputer JUROPA at Jülich Supercomputing Center (JSC) (NIC project 4722, HDD08). Additional computational support was provided by the "Center for Information and Media Technology" (ZIM) at the Heinrich Heine University Düsseldorf.

References

1. Mandal, M.; Breaker, R. R., Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **2004**, 5, 451-463.

2. Coppins, R. L.; Hall, K. B.; Groisman, E. A., The intricate world of riboswitches. *Curr. Opin. Microbiol.* **2007**, 10, 176-181.

3. Serganov, A., The long and the short of riboswitches. *Curr. Opin. Struct. Biol.* **2009**, 19, 251-259.

4. Stoddard, C. D.; Montange, R. K.; Hennelly, S. P.; Rambo, R. P.; Sanbonmatsu, K. Y.; Batey, R. T., Free state conformational sampling of the SAM-I riboswitch aptamer domain. *Structure* **2010**, 18, 787-797.

5. Liberman, J. A.; Wedekind, J. E., Riboswitch structure in the ligand-free state. *Wiley Interdiscip. Rev. RNA* **2012**, 3, 369-384.

6. Serganov, A.; Patel, D. J., Metabolite recognition principles and molecular mechanisms underlying riboswitch function. *Annu. Rev. Biophys.* **2012**, 41, 343-370.

7. Furtig, B.; Nozinovic, S.; Reining, A.; Schwalbe, H., Multiple conformational states of riboswitches fine-tune gene regulation. *Curr. Opin. Struct. Biol.* **2015**, 30, 112-124.

8. Batey, R. T., Structure and mechanism of purine-binding riboswitches. *Q. Rev. Biophys.* **2012**, 45, 345-381.

9. Stagno, J. R.; Liu, Y.; Bhandari, Y. R.; Conrad, C. E.; Panja, S.; Swain, M.; Fan, L.; Nelson, G.; Li, C.; Wendel, D. R.; White, T. A.; Coe, J. D.; Wiedorn, M. O.; Knoska, J.; Oberthuer, D.; Tuckey, R. A.; Yu, P.; Dyba, M.; Tarasov, S. G.; Weierstall, U.; Grant, T. D.; Schwieters, C. D.; Zhang, J.; Ferre-D'Amare, A. R.; Fromme, P.; Draper, D. E.; Liang, M.; Hunter, M. S.; Boutet, S.; Tan, K.; Zuo, X.; Ji, X.; Barty, A.; Zatsepin, N. A.; Chapman, H. N.; Spence, J. C.; Woodson, S. A.; Wang, Y. X., Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature* **2017**, 541, 242-246.

10. Porter, E. B.; Marcano-Velazquez, J. G.; Batey, R. T., The purine riboswitch as a model system for exploring RNA biology and chemistry. *Biochim. Biophys. Acta, Gene Regul. Mech.* **2014**, 1839, 919-930.

11. Batey, R. T.; Gilbert, S. D.; Montange, R. K., Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **2004**, 432, 411-415.

12. Stoddard, C. D.; Gilbert, S. D.; Batey, R. T., Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA* **2008**, 14, 675-684.

13. Prychyna, O.; Dahabieh, M. S.; Chao, J.; O'Neill, M. A., Sequence-dependent folding and unfolding of ligand-bound purine riboswitches. *Biopolymers* **2009**, 91, 953-965.

14. Noeske, J.; Schwalbe, H.; Wöhnert, J., Metal-ion binding and metal-ion induced folding of the adenine-sensing riboswitch aptamer domain. *Nucleic Acids Res.* **2007**, 35, 5262-5273.

15. Noeske, J.; Buck, J.; Fürtig, B.; Nasiri, H. R.; Schwalbe, H.; Wöhnert, J., Interplay of ‘induced fit’and preorganization in the ligand induced folding of the aptamer domain of the guanine binding riboswitch. *Nucleic Acids Res.* **2007**, 35, 572-583.

16. Ottink, O. M.; Rampersad, S. M.; Tessari, M.; Zaman, G. J. R.; Heus, H. A.; Wijmenga, S. S., Ligand-induced folding of the guanine-sensing riboswitch is controlled by a combined predetermined–induced fit mechanism. *RNA* **2007**, 13, 2202-2212.

17. Gilbert, S. D.; Stoddard, C. D.; Wise, S. J.; Batey, R. T., Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *J. Mol. Biol.* **2006**, 359, 754-768.

18. Buck, J.; Noeske, J.; Wöhnert, J.; Schwalbe, H., Dissecting the influence of Mg^{2+} on 3D architecture and ligand-binding of the guanine-sensing riboswitch aptamer domain. *Nucleic Acids Res.* **2010**, 38, 4143-4153.

19. Buck, J.; Wacker, A.; Warkentin, E.; Wöhnert, J.; Wirmer-Bartoschek, J.; Schwalbe, H., Influence of ground-state structure and Mg^{2+} binding on folding kinetics of the guanine-sensing riboswitch aptamer domain. *Nucleic Acids Res.* **2011**, 39, 9768-9778.

20. Sponer, J.; Banas, P.; Jurecka, P.; Zgarbova, M.; Kuhrova, P.; Havrila, M.; Krepl, M.; Stadlbauer, P.; Otyepka, M., Molecular Dynamics Simulations of Nucleic Acids. From Tetranucleotides to the Ribosome. *J. Phys. Chem. Lett.* **2014**, 5, 1771-1782.

21. Hanke, C. A.; Gohlke, H. Force Field Dependence of Riboswitch Dynamics. In *Computational methods for understanding riboswitches*, Chen, S.-J.; Burke-Aguero, D. H., Eds.; Academic Press: Burlington, 2015; Vol. 553, pp 163-191.

22. Hanke, C. A.; Gohlke, H., Ligand-mediated and tertiary interactions cooperatively stabilize the P1 region in the guanine-sensing riboswitch. *PLOS ONE* **2017**, 12, e0179271.

23. Rathore, N.; Chopra, M.; de Pablo, J. J., Optimal allocation of replicas in parallel tempering simulations. *J. Chem. Phys.* **2005**, 122, 024111.

24. Earl, D. J.; Deem, M. W., Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.* **2005**, 7, 3910-3916.

25. Best, R. B.; Mittal, J., Microscopic events in beta-hairpin folding from alternative unfolded ensembles. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, 108, 11087-11092.

26. Best, R. B.; Mittal, J., Balance between a and beta Structures in Ab Initio Protein Folding. *J. Phys. Chem. B* **2010**, 114, 8790-8798.

27. Cavalli, A.; Haberthur, U.; Paci, E.; Caflisch, A., Fast protein folding on downhill energy landscape. *Protein Sci.* **2003**, 12, 1801-1803.

28. Shirts, M. R.; Chodera, J. D., Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, 129, 124105.

29. Dill, K. A.; Chan, H. S., From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **1997**, 4, 10-9.

30. Aqvist, J., Modeling of Ion Ligand Interactions in Solutions and Biomolecules. *Theochem-J. Mol. Struc.* **1992**, 88, 135-152.

31. Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham III, T. E.; Sponer, J.; Otyepka, M., Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* **2010**, 6, 3836-3849.

32. Mlynsky, V.; Banas, P.; Hollas, D.; Reblova, K.; Walter, N. G.; Sponer, J.; Otyepka, M., Extensive molecular dynamics simulations showing that canonical G8 and protonated $A38H^+$ forms are most consistent with crystal structures of hairpin ribozyme. *J. Phys. Chem. B* **2010**, 114, 6642-6652.

33. Sklenovsky, P.; Florova, P.; Banas, P.; Reblova, K.; Lankas, F.; Otyepka, M.; Sponer, J., Understanding RNA Flexibility Using Explicit Solvent Simulations: The Ribosomal and Group I Intron Reverse Kink-Turn Motifs. *J. Chem. Theory Comput.* **2011**, 7, 2963-2980.

34. Banas, P.; Sklenovsky, P.; Wedekind, J. E.; Sponer, J.; Otyepka, M., Molecular mechanism of $preQ_1$ riboswitch action: A molecular dynamics study. *J. Phys. Chem. B* **2012**, 116, 12721-12734.

35. Sindhikara, D.; Meng, Y.; Roitberg, A. E., Exchange frequency in replica exchange molecular dynamics. *J. Chem. Phys.* **2008**, 128, 024103.

36. Sindhikara, D. J.; Emerson, D. J.; Roitberg, A. E., Exchange Often and Properly in Replica Exchange Molecular Dynamics. *J. Chem. Theory Comput.* **2010**, 6, 2804-2808.

37. Chen, S. J., RNA folding: Conformational statistics, folding kinetics, and ion electrostatics. *Annu. Rev. Biophys.* **2008**, 37, 197-214.

38. Draper, D. E.; Grilley, D.; Soto, A. M., Ions and RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, 34, 221-243.

39. Quarta, G.; Sin, K.; Schlick, T., Dynamic energy landscapes of riboswitches help interpret conformational rearrangements and function. *PLoS Comput. Biol.* **2012**, 8, e1002368.

40. Steinert, H.; Sochor, F.; Wacker, A.; Buck, J.; Helmling, C.; Hiller, F.; Keyhani, S.; Noeske, J.; Grimm, S.; Rudolph, M. M.; Keller, H.; Mooney, R. A.; Landick, R.; Suess, B.; Furtig, B.; Wohnert, J.; Schwalbe, H., Pausing guides RNA folding to populate transiently stable RNA structures for riboswitch-based transcription regulation. *eLife* **2017**, 6, e21297.

41. Thirumalai, D.; Hyeon, C., RNA and protein folding: common themes and variations. *Biochemistry* **2005**, 44, 4957-4970.

42. Reining, A.; Nozinovic, S.; Schlepckow, K.; Buhr, F.; Furtig, B.; Schwalbe, H., Three-state mechanism couples ligand and temperature sensing in riboswitches. *Nature* **2013**, 499, 355-359.

43. Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, 26, 1668-1688.

44. Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvai, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A., In; University of California, San Francisco, 2010.

45. Wang, J. M.; Cieplak, P.; Kollman, P. A., How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, 21, 1049-1074.

46. Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **1999**, 314, 141-151.

47. Patriksson, A.; van der Spoel, D., A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.* **2008**, 10, 2073-2077.

48. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, 79, 926-935.

49. Gohlke, H.; Kiel, C.; Case, D. A., Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RaIGDS complexes. *J. Mol. Biol.* **2003**, 330, 891-913.

50. Roe, D. R.; Cheatham III, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, 9, 3084-3095.

51. Case, D. A.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A., In; University of California, San Francisco, 2016.

52. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M., The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J. Comput. Chem.* **1992**, 13, 1011-1021.

53. Galindo-Murillo, R.; Roe, D. R.; Cheatham, T. E., Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAACGAACGAACGC). *Biochim. Biophys. Acta* **2015**, 1850, 1041-1058.

54. Fajer, M.; Swift, R. V.; McCammon, J. A., Using Multistate Free Energy Techniques to Improve the Efficiency of Replica Exchange Accelerated Molecular Dynamics. *J. Comput. Chem.* **2009**, 30, 1719-1725.

55. Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Dill, K. A., Long-time protein folding dynamics from short-time molecular dynamics simulations. *Multiscale Model Sim* **2006**, 5, 1214-1226.

56. Chu, B. K.; Fu, I. W.; Markegard, C. B.; Choi, S. E.; Nguyen, H. D., A Tail of Two Peptide Amphiphiles: Effect of Conjugation with Hydrophobic Polymer on Folding of Peptide Sequences. *Biomacromolecules* **2014**, 15, 3313-3320.

57. Moradi, M.; Tajkhorshid, E., Computational Recipe for Efficient Description of Large-Scale Conformational Changes in Biomolecular Systems. *J. Chem. Theory Comput.* **2014**, 10, 2866-2880.

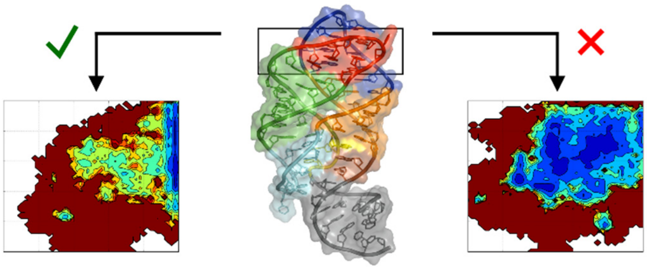
58. Li, P. C.; Miyashita, N.; Im, W.; Ishido, S.; Sugita, Y., Multidimensional Umbrella Sampling and Replica-Exchange Molecular Dynamics Simulations for Structure Prediction of Transmembrane Helix Dimers. *J. Comput. Chem.* **2014**, 35, 300-308.

59. Kalinin, S.; Peulen, T.; Sindbert, S.; Rothwell, P. J.; Berger, S.; Restle, T.; Goody, R. S.; Gohlke, H.; Seidel, C. A. M., A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Meth.* **2012**, 9, 1218-1129.

60. Serganov, A.; Yuan, Y. R.; Pikovskaya, O.; Polonskaia, A.; Malinina, L.; Phan, A. T.; Hobartner, C.; Micura, R.; Breaker, R. R.; Patel, D. J., Structural basis for discriminative regulation of gene expression by adenine-and guanine-sensing mRNAs. *Chem. Biol.* **2004**, 11, 1729-1741.

61. Stoddard, C. D.; Widmann, J.; Trausch, J. J.; Marcano-Velazquez, J. G.; Knight, R.; Batey, R. T., Nucleotides adjacent to the ligand-binding pocket are linked to activity tuning in the purine riboswitch. *J. Mol. Biol.* **2013**, 425, 1596-1611.

TOC Figure



Supporting Information

Supplemental Figures

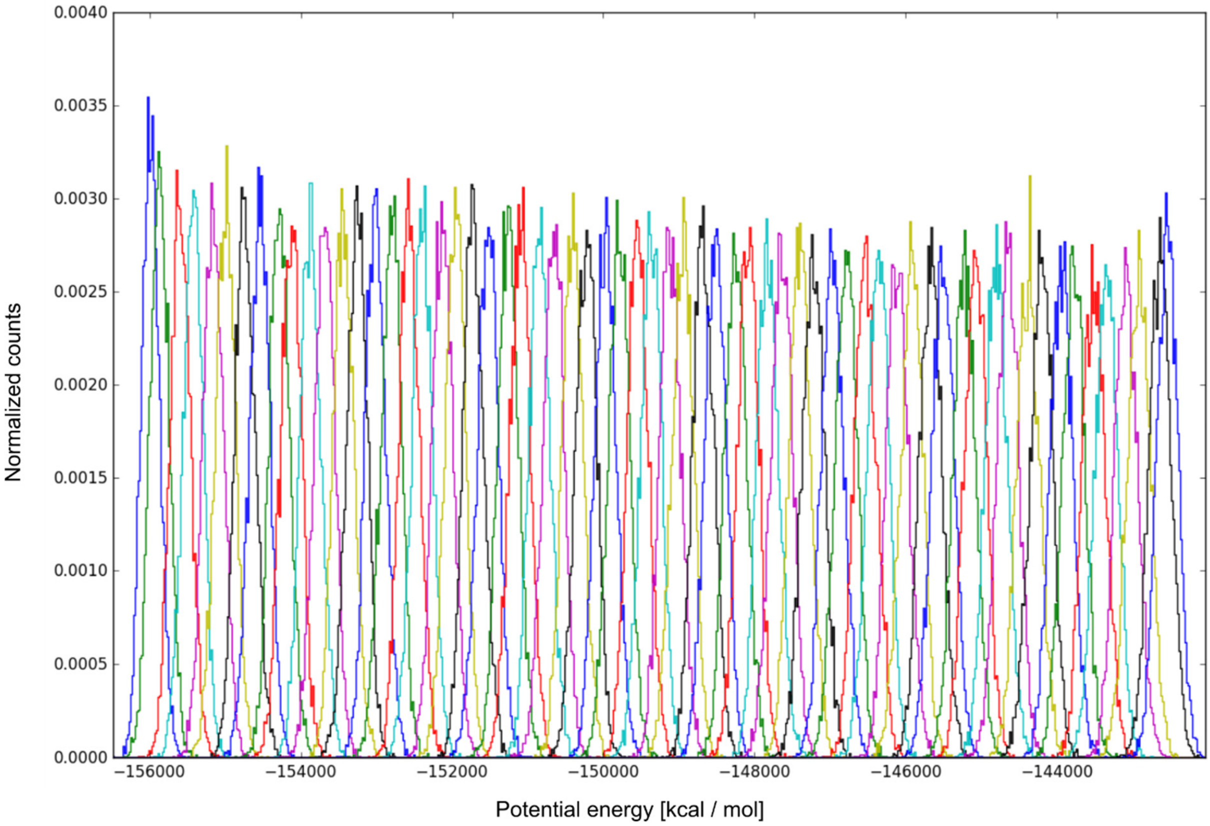


Figure S1. Normalized potential energy distributions for all 64 replica (differently colored) for the REMD simulation of Gsw^{apt} in the absence of Mg^{2+} ions.

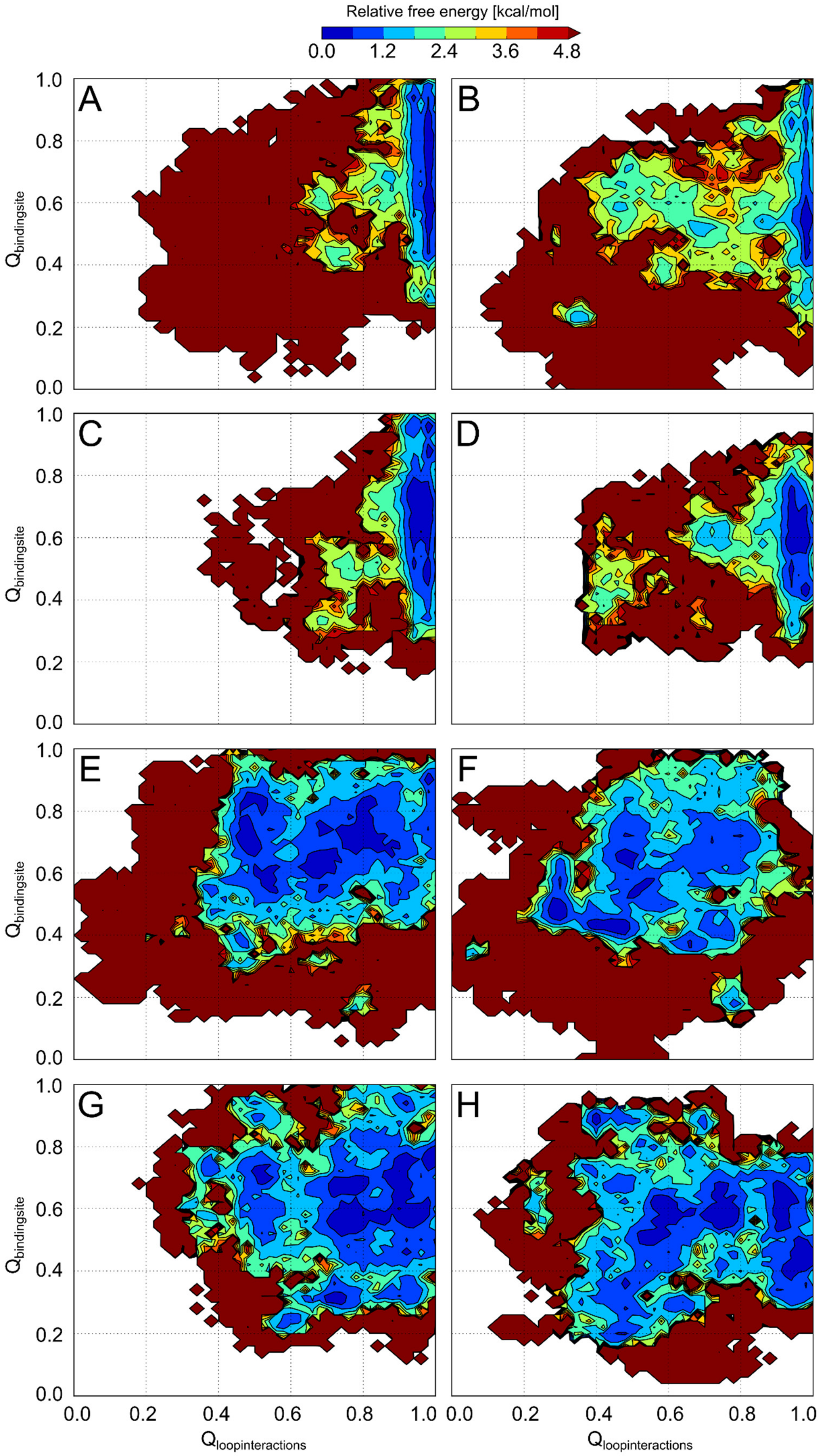


Figure S2. Comparison of 2D PMF calculated based on the first (left column) and second (right column) half of the trajectory for Gsw^{apt} at 0 Mg^{2+} (A, B), Gsw^{apt} at 20 Mg^{2+} (C, D), Gsw^{loop} at 0 Mg^{2+} (E, F), and Gsw^{loop} at 20 Mg^{2+} (G, H).

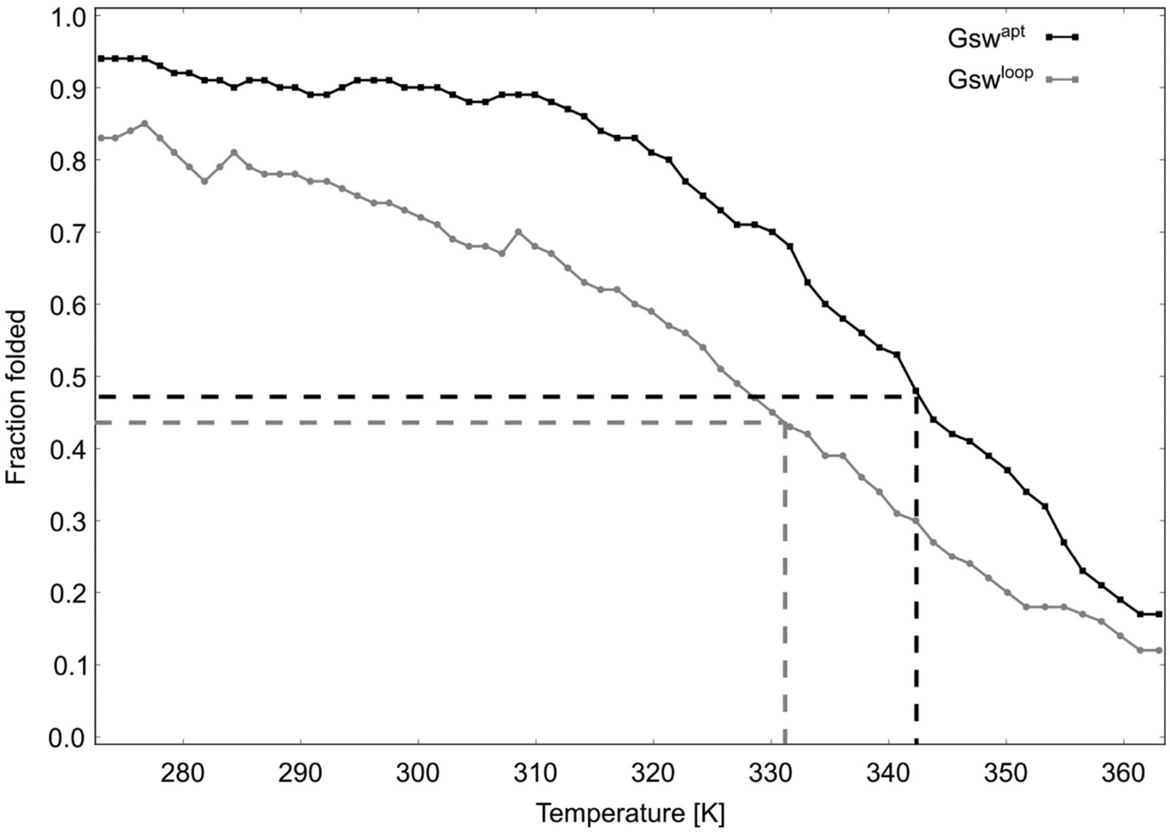


Figure S3. Melting curve for Gsw^{apt} (black) and Gsw^{loop} (grey) in the absence of Mg^{2+} ions. A structure is considered folded if $\boldsymbol{Q_{all}} > 0.75$ (eq. 1). The dashed lines indicate the melting temperatures determined as 50% of the initial folded fraction.

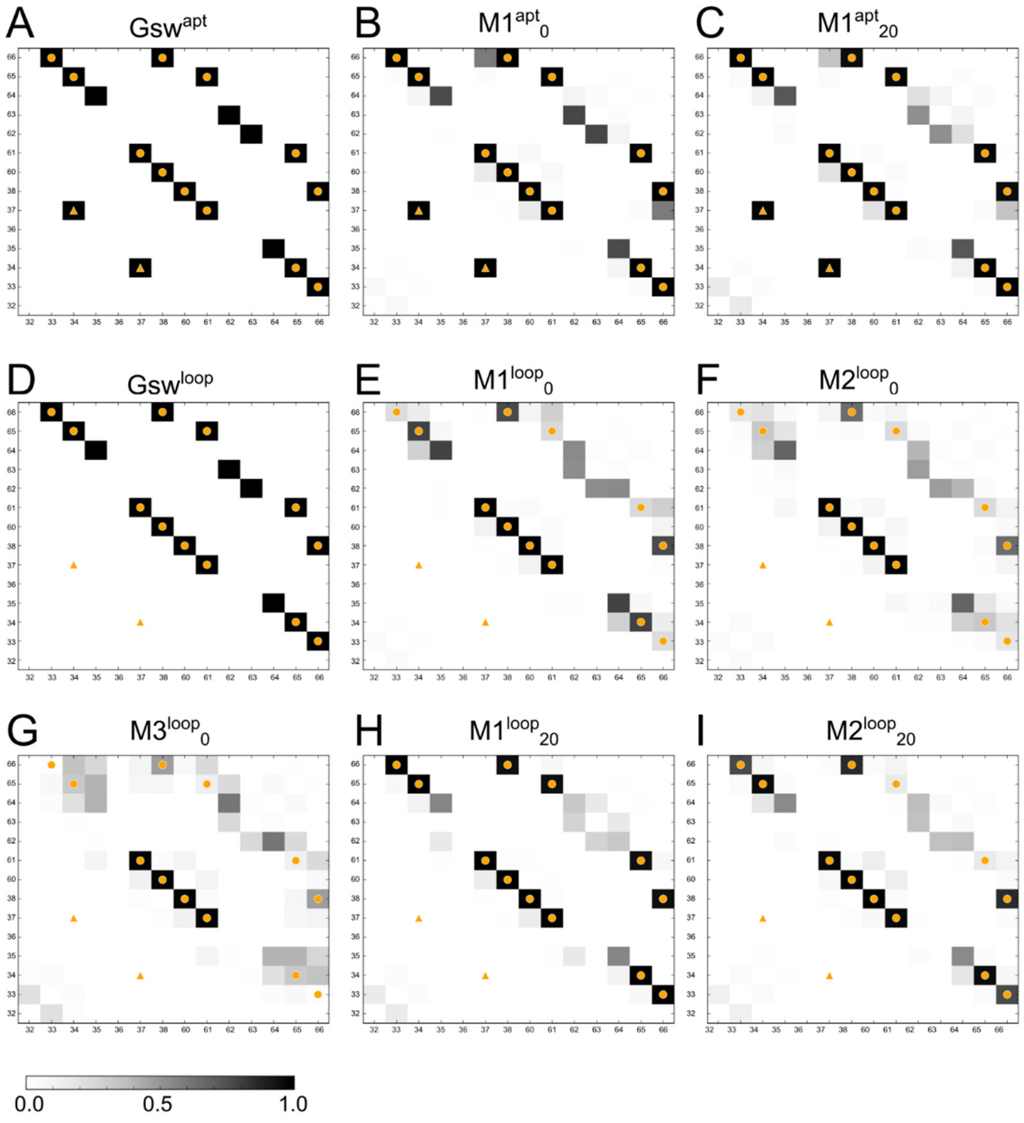


Figure S4. Hydrogen bond occupancy for the L2 and L3 region calculated for the structural ensemble in the free energy basins (M1 to M3) indicated in Figure 2. Orange dots indicate native hydrogen bonds formed in both Gsw^{apt} and Gsw^{loop}, the orange triangles indicate the base pair formation between nucleotides 34 and 37, which is lost in Gsw^{loop} due to the G37A/C61U mutation. Gsw^{apt} (A) and Gsw^{loop} (D) indicate the hydrogen bond networks in the loop regions of the crystal structures [1,2].

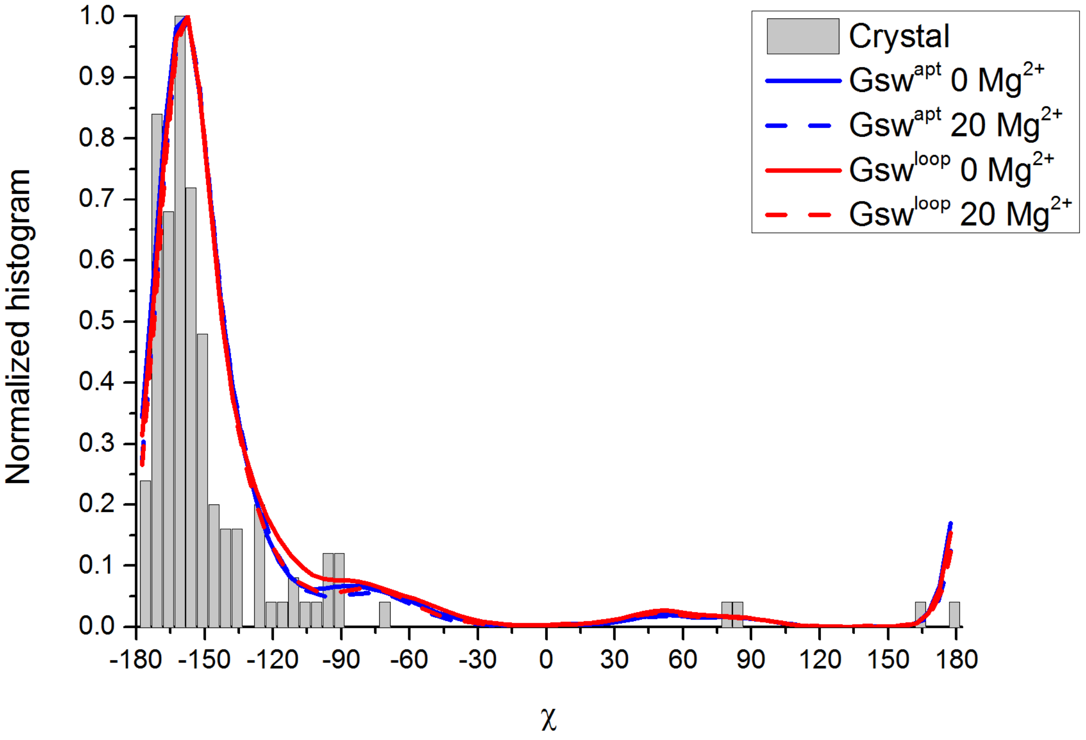


Figure S5. Normalized histogram of χ torsions reweighted based on the 2D PMF. See the legend for which simulated systems and simulation conditions are displayed by the blue and red curves. The grey bars represent the histogram for the torsion angles in the crystal structures of Gsw^{apt} and Gsw^{loop} (PDB IDs 4FE5 1 and 3RKF 2, respectively).

Supplemental Tables

Table S1. Percentage of successful exchanges per replica index[a].

Replica	Success rate[b]	Replica	Success rate[b]
1	32.5	33	33.6
2	31.9	34	31.8
3	31.2	35	32.2
4	32.5	36	31.6
5	31.4	37	32.8
6	32.3	38	30.7
7	30.9	39	33.7
8	33.8	40	31.6
9	32.4	41	32.6
10	32.4	42	33.5
11	32.1	43	32.5
12	34.2	44	32.1
13	32.6	45	31.9
14	31.9	46	33.5
15	30.5	47	34.2
16	32.0	48	31.7
17	31.5	49	31.7
18	31.6	50	32.1
19	32.8	51	33.2
20	32.8	52	32.9
21	32.3	53	31.6
22	31.3	54	32.0
23	32.5	55	31.0
24	32.8	56	32.2
25	30.7	57	33.4
26	32.9	58	32.4
27	33.5	59	33.9
28	32.0	60	32.8
29	32.2	61	34.1
30	32.7	62	31.1
31	33.1	63	34.4
32	33.6	64	32.3

[a] For the REMD simulation of Gsw^{apt} in the absence of Mg^{2+} ions.

[b] In %.

Table S2: Overlap of potential energy distributions between neighboring replica[a].

T 1[b]	T 2[b]	Overlap of potential energy distribution[c]	T 1[b]	T 2[b]	Overlap of potential energy distribution[c]
273.0	274.2	63.9	315.5	316.9	57.3
274.2	275.5	29.0	316.9	318.4	33.5
275.5	276.7	54.2	318.4	319.8	57.2
276.7	278.0	31.5	319.8	321.3	34.8
278.0	279.2	56.0	321.3	322.7	58.0
279.2	280.5	32.8	322.7	324.2	35.9
280.5	281.8	49.3	324.2	325.7	54.4
281.8	283.1	31.7	325.7	327.1	39.6
283.1	284.3	58.4	327.1	328.6	55.0
284.3	285.6	34.9	328.6	330.1	38.0
285.6	286.9	52.2	330.1	331.6	55.3
286.9	288.2	35.1	331.6	333.1	38.2
288.2	289.5	53.0	333.1	334.6	57.8
289.5	290.8	35.9	334.6	336.1	38.2
290.8	292.2	46.2	336.1	337.7	54.0
292.2	293.5	37.3	337.7	339.2	39.3
293.5	294.8	54.7	339.2	340.7	59.8
294.8	296.2	32.0	340.7	342.3	34.0
296.2	297.5	54.9	342.3	343.8	62.3
297.5	298.8	37.4	343.8	345.4	35.1
298.8	300.2	50.5	345.4	346.9	61.2
300.2	301.6	33.6	346.9	348.5	36.8
301.6	302.9	57.8	348.5	350.1	57.5
302.9	304.3	34.6	350.1	351.7	37.1
304.3	305.7	52.5	351.7	353.3	59.4
305.7	307.1	36.1	353.3	354.9	37.8
307.1	308.5	53.3	354.9	356.5	60.2
308.5	309.9	35.9	356.5	358.1	38.7
309.9	311.3	54.0	358.1	359.7	59.4
311.3	312.7	37.3	359.7	361.4	34.9
312.7	314.1	54.6	361.4	363.0	73.2

[a] For the REMD simulation of Gsw^{apt} in the absence of Mg^{2+} ions.

[b] Temperature of the replica, in K.

[c] In %. The overlap was analytically calculated for two Gaussians fitted to the respective distribution.

Table S3. Characteristics of the free energy basins shown in Figure 2.

System	Basin	Number of structures in basin	Cluster		Number of structures with intact ligand binding site[b]
			Number of structures	RMSD of representative structure[a]	
$\mathrm{Gsw}^{\mathrm{apt}}$ 0 Mg^{2+}	$M1_{0}^{apt}$	106,449	106,347	2.5	55,408
$\mathrm{Gsw}^{\mathrm{apt}}$ 20 Mg^{2+}	$M1_{20}^{apt}$	209,803	209,799	1.6	79,335
$\mathrm{Gsw}^{\mathrm{loop}}$ 0 Mg^{2+}	$M1_{0}^{loop}$	51,834	49,370	2.5	27,220
	$M2_{0}^{loop}$	34,999	28,663	2.6	9,794
	$M3_{0}^{loop}$	9,624	2,615 2,332 1,114	5.0 4.3 3.4	154[c]
$\mathrm{Gsw}^{\mathrm{loop}}$ 20 Mg^{2+}	$M1_{20}^{loop}$	40,648	40,646	1.5	21,031
	$M2_{20}^{loop}$	39,588	39,578	1.6	16,713

[a] In Å. The RMSD was calculated with respect to the initial structure using the phosphorus atoms but omitting the P1 region.

[b] An intact ligand binding site was defined as having an RMSD of nucleotides 47, 51, and 74 < 3 Å and a volume of the ligand binding site > 150 accessible points.

[c] Total number in all three clusters.

Supplemental References

1. Batey, R. T.; Gilbert, S. D.; Montange, R. K., Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **2004**, 432, 411-415.
2. Buck, J.; Wacker, A.; Warkentin, E.; Wöhnert, J.; Wirmer-Bartoschek, J.; Schwalbe, H., Influence of ground-state structure and Mg^{2+} binding on folding kinetics of the guanine-sensing riboswitch aptamer domain. *Nucleic Acids Res.* **2011**, 39, 9768-9778.

References

1. Berg, J.M., J.L. Tymoczko, and L. Stryer, *Biochemistry*. 7th ed. 2010, New York: W. H. Freeman.
2. Tinoco, I. and C. Bustamante, *How RNA folds.* J. Mol. Biol., 1999. **293**(2): p. 271-281.
3. Mandal, M., et al., *Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria.* Cell, 2003. **113**(5): p. 577-586.
4. Mandal, M. and R.R. Breaker, *Gene regulation by riboswitches.* Nat. Rev. Mol. Cell Biol., 2004. **5**(6): p. 451-463.
5. Blount, K.F. and R.R. Breaker, *Riboswitches as antibacterial drug targets.* Nat. Biotechnol., 2006. **24**(12): p. 1558-1564.
6. Coppins, R.L., K.B. Hall, and E.A. Groisman, *The intricate world of riboswitches.* Curr. Opin. Microbiol., 2007. **10**(2): p. 176-181.
7. Breaker, R.R., *Riboswitches: from ancient gene-control systems to modern drug targets.* Future Microbiol., 2009. **4**(7): p. 771-773.
8. Deigan, K.E. and A.R. Ferré-D'Amaré, *Riboswitches: Discovery of drugs that target bacterial gene-regulatory RNAs.* Acc. Chem. Res., 2011: p. 1329-1338.
9. Lünse, C.E., A. Schuller, and G. Mayer, *The promise of riboswitches as potential antibacterial drug targets.* Int. J. Med. Microbiol., 2014. **304**(1): p. 79-92.
10. Garst, A.D. and R.T. Batey, *A switch in time: detailing the life of a riboswitch.* Biochim. Biophys. Acta, Gene Regul. Mech., 2009. **1789**(9-10): p. 584-591.
11. Porter, E.B., J.G. Marcano-Velazquez, and R.T. Batey, *The purine riboswitch as a model system for exploring RNA biology and chemistry.* Biochim. Biophys. Acta, Gene Regul. Mech., 2014. **1839**(10): p. 919-930.
12. Wickiser, J.K., et al., *The kinetics of ligand binding by an adenine-sensing riboswitch.* Biochemistry, 2005. **44**(40): p. 13404-13414.
13. Stoddard, C.D., et al., *Free state conformational sampling of the SAM-I riboswitch aptamer domain.* Structure, 2010. **18**(7): p. 787-797.
14. Batey, R.T., *Structure and mechanism of purine-binding riboswitches.* Q. Rev. Biophys., 2012. **45**(3): p. 345-381.
15. Stoddard, C.D. and R.T. Batey, *Beyond crystallography: Investigating the conformational dynamics of the purine riboswitch*, in *Non-Protein Coding RNAs*, N.G. Walter, Editor. 2009, Springer-Verlag: Berlin Heidelberg.
16. Watson, J.D. and F.H.C. Crick, *Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid.* Nature, 1953. **171**(4356): p. 737-738.
17. Batey, R.T., R.P. Rambo, and J.A. Doudna, *Tertiary motifs in RNA structure and folding.* Angew. Chem. Int. Edit., 1999. **38**(16): p. 2327-2343.
18. Moore, P.B., *Structural motifs in RNA.* Annu. Rev. Biochem., 1999. **68**: p. 287-300.
19. Hendrix, D.K., S.E. Brenner, and S.R. Holbrook, *RNA structural motifs: building blocks of a modular biomolecule.* Q. Rev. Biophys., 2005. **38**(3): p. 221-243.
20. Leontis, N.B. and E. Westhof, *Analysis of RNA motifs.* Curr. Opin. Struct. Biol., 2003. **13**(3): p. 300-308.
21. Leontis, N.B., A. Lescoute, and E. Westhof, *The building blocks and motifs of RNA architecture.* Curr. Opin. Struct. Biol., 2006. **16**(3): p. 279-287.
22. Leontis, N.B., J. Stombaugh, and E. Westhof, *The non-Watson-Crick base pairs and their associated isostericity matrices.* Nucleic Acids Res., 2002. **30**(16): p. 3497-3531.
23. Reblova, K., et al., *Non-Watson-Crick basepairing and hydration in RNA motifs: Molecular dynamics of 5S rRNA loop E.* Biophys. J., 2003. **84**(6): p. 3564-3582.
24. Drew, H.R., et al., *Structure of a B-DNA dodecamer - conformation and dynamics.* Proc. Natl. Acad. Sci. U. S. A., 1981. **78**(4): p. 2179-2183.

25. Hingerty, B., R.S. Brown, and A. Jack, *Further refinement of structure of yeast transfer-RNA Phe.* J. Mol. Biol., 1978. **124**(3): p. 523-534.
26. Batey, R.T., S.D. Gilbert, and R.K. Montange, *Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine.* Nature, 2004. **432**(7015): p. 411-415.
27. Howe, J.A., et al., *Selective small-molecule inhibition of an RNA structural element.* Nature, 2015. **526**(7575): p. 672-677.
28. Kapral, G.J., et al., *New tools provide a second look at HDV ribozyme structure, dynamics and cleavage.* Nucleic Acids Res., 2014. **42**(20): p. 12833-12846.
29. Mattick, J.S. and I.V. Makunin, *Non-coding RNA.* Hum. Mol. Genet., 2006. **15**: p. R17-R29.
30. Crick, F.H., *On protein synthesis.* Symp Soc Exp Biol, 1958. **12**: p. 138-163.
31. Morris, K.V. and J.S. Mattick, *The rise of regulatory RNA.* Nat. Rev. Genet., 2014. **15**(6): p. 423-437.
32. Noller, H.F., *Structure of ribosomal RNA.* Annu. Rev. Biochem., 1984. **53**: p. 119-162.
33. Nissen, P., et al., *The structural basis of ribosome activity in peptide bond synthesis.* Science, 2000. **289**(5481): p. 920-930.
34. Doherty, E.A. and J.A. Doudna, *Ribozyme structures and mechanisms.* Annu. Rev. Biophys. Biomol. Struct., 2001. **30**: p. 457-475.
35. John, B., et al., *Human microRNA targets.* PLoS Biol., 2004. **2**(11): p. 1862-1879.
36. Montange, R.K. and R.T. Batey, *Riboswitches: emerging themes in RNA structure and function.* Annu. Rev. Biophys., 2008. **37**: p. 117-133.
37. Roth, A. and R.R. Breaker, *The structural and functional diversity of metabolite-binding riboswitches.* Annu. Rev. Biochem., 2009. **78**: p. 305-334.
38. Schwalbe, H., et al., *Structures of RNA switches: insight into molecular recognition and tertiary structure.* Angew. Chem. Int. Edit., 2007. **46**(8): p. 1212-1219.
39. Breaker, R.R., *Riboswitches and the RNA World.* Cold Spring Harb. Perspect. Biol., 2012. **4**(2).
40. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-1712.
41. Gilbert, W., *Origin of Life - the RNA World.* Nature, 1986. **319**(6055): p. 618-618.
42. Cech, T.R., *The RNA Worlds in Context.* Cold Spring Harb. Perspect. Biol., 2012. **4**(7).
43. Robertson, M.P. and G.F. Joyce, *The Origins of the RNA World.* Cold Spring Harb. Perspect. Biol., 2012. **4**(5).
44. Vitreschak, A.G., et al., *Riboswitches: the oldest mechanism for the regulation of gene expression?* Trends Genet., 2004. **20**(1): p. 44-50.
45. Winkler, W.C. and R.R. Breaker, *Regulation of bacterial gene expression by riboswitches.* Annu. Rev. Microbiol., 2005. **59**: p. 487-517.
46. Batey, R.T., *Structures of regulatory elements in mRNAs.* Curr. Opin. Struct. Biol., 2006. **16**(3): p. 299-306.
47. Winkler, W.C., *Riboswitches and the role of noncoding RNAs in bacterial metabolic control.* Curr. Opin. Chem. Biol., 2005. **9**(6): p. 594-602.
48. Tucker, B.J. and R.R. Breaker, *Riboswitches as versatile gene control elements.* Curr. Opin. Struct. Biol., 2005. **15**(3): p. 342-348.
49. Serganov, A. and D.J. Patel, *Metabolite recognition principles and molecular mechanisms underlying riboswitch function.* Annu. Rev. Biophys., 2012. **41**: p. 343-370.
50. Breaker, R.R., *Prospects for riboswitch discovery and analysis.* Mol. Cell, 2011. **43**(6): p. 867-879.
51. Mandal, M. and R.R. Breaker, *Adenine riboswitches and gene activation by disruption of a transcription terminator.* Nat. Struct. Mol. Biol., 2004. **11**(1): p. 29-35.

52. Serganov, A., et al., *Structural basis for discriminative regulation of gene expression by adenine-and guanine-sensing mRNAs.* Chem. Biol., 2004. **11**(12): p. 1729-1741.
53. Wang, J.X. and R.R. Breaker, *Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine.* Biochem. Cell Biol., 2008. **86**(2): p. 157-168.
54. Montange, R.K. and R.T. Batey, *Structure of the S-adenosylmethionine riboswitch regulatory mRNA element.* Nature, 2006. **441**(7097): p. 1172-1175.
55. Lu, C., et al., *Crystal structures of the SAM-III/S-MK riboswitch reveal the SAM-dependent translation inhibition mechanism.* Nat. Struct. Mol. Biol., 2008. **15**(10): p. 1076-1083.
56. Matzner, D. and G. Mayer, *(Dis)similar analogues of riboswitch metabolites as antibacterial lead compounds.* J. Med. Chem., 2015. **58**(8): p. 3275-3286.
57. Nahvi, A., et al., *Genetic control by a metabolite binding mRNA.* Chem. Biol., 2002. **9**(9): p. 1043-1049.
58. Nahvi, A., J.E. Barrick, and R.R. Breaker, *Coenzyme B-12 riboswitches are widespread genetic control elements in prokaryotes.* Nucleic Acids Res., 2004. **32**(1): p. 143-150.
59. Mironov, A.S., et al., *Sensing small molecules by nascent RNA: A mechanism to control transcription in bacteria.* Cell, 2002. **111**(5): p. 747-756.
60. Winkler, W., A. Nahvi, and R.R. Breaker, *Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression.* Nature, 2002. **419**(6910): p. 952-956.
61. Winkler, W.C., S. Cohen-Chalamish, and R.R. Breaker, *An mRNA structure that controls gene expression by binding FMN.* Proc. Natl. Acad. Sci. U. S. A., 2002. **99**(25): p. 15908-15913.
62. Corbino, K.A., et al., *Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria.* Genome Biol., 2005. **6**(8): p. R70.1-R70.10.
63. Epshtein, V., A.S. Mironov, and E. Nudler, *The riboswitch-mediated control of sulfur metabolism in bacteria.* Proc. Natl. Acad. Sci. U. S. A., 2003. **100**(9): p. 5052-5056.
64. Fuchs, R.T., F.J. Grundy, and T.M. Henkin, *The S-MK box is a new SAM-binding RNA for translational regulation of SAM synthetase.* Nat. Struct. Mol. Biol., 2006. **13**(3): p. 226-233.
65. McDaniel, B.A.M., et al., *Transcription termination control of the S box system: Direct measurement of S-adenosylmethionine by the leader RNA.* Proc. Natl. Acad. Sci. U. S. A., 2003. **100**(6): p. 3083-3088.
66. Poiata, E., et al., *A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria.* RNA, 2009. **15**(11): p. 2046-2056.
67. Weinberg, Z., et al., *The aptamer core of SAM-IV riboswitches mimics the ligand-binding site of SAM-I riboswitches.* RNA, 2008. **14**(5): p. 822-828.
68. Weinberg, Z., et al., *Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes.* Genome Biol., 2010. **11**(3): p. R31.1-R31.17.
69. Grundy, F.J. and T.M. Henkin, *The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria.* Mol. Microbiol., 1998. **30**(4): p. 737-749.
70. Winkler, W.C., et al., *An mRNA structure that controls gene expression by binding S-adenosylmethionine.* Nat. Struct. Biol., 2003. **10**(9): p. 701-707.
71. Wang, J.X., et al., *Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling.* Mol. Cell, 2008. **29**(6): p. 691-702.
72. Ames, T.D., et al., *A eubacterial riboswitch class that senses the coenzyme tetrahydrofolate.* Chem. Biol., 2010. **17**(7): p. 681-685.

73. Regulski, E.E., et al., *A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism.* Mol. Microbiol., 2008. **68**(4): p. 918-932.
74. Kim, J. and R. Breaker, *Purine sensing by riboswitches.* Biol. Cell, 2008. **100**: p. 1-11.
75. Meyer, M.M., et al., *Confirmation of a second natural preQ$_1$ aptamer class in Streptococcaceae bacteria.* RNA, 2008. **14**(4): p. 685-695.
76. Roth, A., et al., *A riboswitch selective for the queuosine precursor preQ$_1$ contains an unusually small aptamer domain.* Nat. Struct. Mol. Biol., 2007. **14**(4): p. 308-317.
77. Kim, J.N., A. Roth, and R.R. Breaker, *Guanine riboswitch variants from Mesoplasma florum selectively recognize 2 '-deoxyguanosine.* Proc. Natl. Acad. Sci. U. S. A., 2007. **104**(41): p. 16092-16097.
78. Lee, E.R., et al., *An allosteric self-splicing ribozyme triggered by a bacterial second messenger.* Science, 2010. **329**(5993): p. 845-848.
79. Sudarsan, N., et al., *Riboswitches in eubacteria sense the second messenger cyclic di-GMP.* Science, 2008. **321**(5887): p. 411-413.
80. Nelson, J.W., et al., *Riboswitches in eubacteria sense the second messenger c-di-AMP.* Nat. Chem. Biol., 2013. **9**(12): p. 834-841.
81. Sudarsan, N., et al., *An mRNA structure in bacteria that controls gene expression by binding lysine.* Genes Dev., 2003. **17**(21): p. 2688-2697.
82. Grundy, F.J., S.C. Lehman, and T.M. Henkin, *The L box regulon: Lysine sensing by leader RNAs of bacterial lysine biosynthesis genes.* Proc. Natl. Acad. Sci. U. S. A., 2003. **100**(21): p. 12057-12062.
83. Rodionov, D.A., et al., *Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch?* Nucleic Acids Res., 2003. **31**(23): p. 6748-6757.
84. Mandal, M., et al., *A glycine-dependent riboswitch that uses cooperative binding to control gene expression.* Science, 2004. **306**(5694): p. 275-279.
85. Ames, T.D. and R.R. Breaker, *Bacterial aptamers that selectively bind glutamine.* RNA Biol., 2011. **8**(1): p. 82-89.
86. Winkler, W.C., et al., *Control of gene expression by a natural metabolite-responsive ribozyme.* Nature, 2004. **428**(6980): p. 281-286.
87. Klein, D.J. and A.R. Ferre-D'Amare, *Structural basis of glmS ribozyme activation by glucosamine-6-phosphate.* Science, 2006. **313**(5794): p. 1752-1756.
88. Cromie, M.J., et al., *An RNA sensor for intracellular Mg^{2+}.* Cell, 2006. **125**(1): p. 71-84.
89. Dann, C.E., et al., *Structure and mechanism of a metal-sensing regulatory RNA.* Cell, 2007. **130**(5): p. 878-892.
90. Li, S.S. and R.R. Breaker, *Fluoride enhances the activity of fungicides that destabilize cell membranes.* Bioorg. Med. Chem. Lett., 2012. **22**(9): p. 3317-3322.
91. Peselis, A. and A. Serganov, *Themes and variations in riboswitch structure and function.* Biochim. Biophys. Acta, Gene Regul. Mech., 2014. **1839**(10): p. 908-918.
92. Ramesh, A., *Second messenger - sensing riboswitches in bacteria.* Semin. Cell Dev. Biol., 2015. **47-48**: p. 3-8.
93. Serganov, A. and E. Nudler, *A decade of riboswitches.* Cell, 2013. **152**(1-2): p. 17-24.
94. Barrick, J.E. and R.R. Breaker, *The distributions, mechanisms, and structures of metabolite-binding riboswitches.* Genome Biol., 2007. **8**(11): p. R239.1-R239.19.
95. Cheah, M.T., et al., *Control of alternative RNA splicing and gene expression by eukaryotic riboswitches.* Nature, 2007. **447**(7143): p. 497-501.
96. Li, S.S. and R.R. Breaker, *Eukaryotic TPP riboswitch regulation of alternative splicing involving long-distance base pairing.* Nucleic Acids Res., 2013. **41**(5): p. 3022-3031.
97. Croft, M.T., et al., *Thiamine biosynthesis in algae is regulated by riboswitches.* Proc. Natl. Acad. Sci. U. S. A., 2007. **104**(52): p. 20770-20775.

98. Wachter, A., et al., *Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs.* Plant Cell, 2007. **19**(11): p. 3437-3450.
99. Bocobza, S., et al., *Riboswitch-dependent gene regulation and its evolution in the plant kingdom.* Genes Dev., 2007. **21**(22): p. 2874-2879.
100. Gusarov, I. and E. Nudler, *The mechanism of intrinsic transcription termination.* Mol. Cell, 1999. **3**(4): p. 495-504.
101. Yarnell, W.S. and J.W. Roberts, *Mechanism of intrinsic transcription termination and antitermination.* Science, 1999. **284**(5414): p. 611-615.
102. Lee, D.N., et al., *Transcription pausing by Escherichia coli RNA-polymerase is modulated by downstream DNA-sequences.* J. Biol. Chem., 1990. **265**(25): p. 15145-15153.
103. Wickiser, J.K., et al., *The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch.* Mol. Cell, 2005. **18**(1): p. 49-60.
104. Stoddard, C.D., et al., *Nucleotides adjacent to the ligand-binding pocket are linked to activity tuning in the purine riboswitch.* J. Mol. Biol., 2013. **425**(10): p. 1596-1611.
105. Vogel, U. and K.F. Jensen, *The RNA chain elongation rate in Escherichia coli depends on the growth rate.* J. Bacteriol., 1994. **176**(10): p. 2807-2813.
106. Proshkin, S., et al., *Cooperation between translating ribosomes and RNA polymerase in transcription elongation.* Science, 2010. **328**(5977): p. 504-508.
107. Hollands, K., et al., *Riboswitch control of Rho-dependent transcription termination.* Proc. Natl. Acad. Sci. U. S. A., 2012. **109**(14): p. 5376-5381.
108. Nudler, E. and A.S. Mironov, *The riboswitch control of bacterial metabolism.* Trends Biochem. Sci., 2004. **29**(1): p. 11-17.
109. Chen, A.G.Y., N. Sudarsan, and R.R. Breaker, *Mechanism for gene control by a natural allosteric group I ribozyme.* RNA, 2011. **17**(11): p. 1967-1972.
110. Loh, E., et al., *A trans-acting riboswitch controls expression of the virulence regulator PrfA in Listeria monocytogenes.* Cell, 2009. **139**(4): p. 770-779.
111. Peselis, A., A. Gao, and A. Serganov, *Cooperativity, allostery and synergism in ligand binding to riboswitches.* Biochimie, 2015. **117**: p. 100-109.
112. Butler, E.B., et al., *Structural basis of cooperative ligand binding by the glycine riboswitch.* Chem. Biol., 2011. **18**(3): p. 293-298.
113. Sherman, E.M., et al., *An energetically beneficial leader-linker interaction abolishes ligand-binding cooperativity in glycine riboswitches.* RNA, 2012. **18**(3): p. 496-507.
114. Baird, N.J. and A.R. Ferre-D'Amare, *Modulation of quaternary structure and enhancement of ligand binding by the K-turn of tandem glycine riboswitches.* RNA, 2013. **19**(2): p. 167-176.
115. Ruff, K.M. and S.A. Strobel, *Ligand binding by the tandem glycine riboswitch depends on aptamer dimerization but not double ligand occupancy.* RNA, 2014. **20**(11): p. 1775-1788.
116. Sudarsan, N., et al., *Tandem riboswitch architectures exhibit complex gene control functions.* Science, 2006. **314**(5797): p. 300-304.
117. Quarta, G., K. Sin, and T. Schlick, *Dynamic energy landscapes of riboswitches help interpret conformational rearrangements and function.* PLoS Comput. Biol., 2012. **8**(2): p. e1002368.
118. Ferre-D'Amare, A.R. and W.C. Winkler, *The roles of metal ions in regulation by riboswitches.* Met. Ions Life Sci., 2011. **9**: p. 141-73.
119. Serganov, A., *The long and the short of riboswitches.* Curr. Opin. Struct. Biol., 2009. **19**(3): p. 251-259.
120. Liberman, J.A. and J.E. Wedekind, *Riboswitch structure in the ligand-free state.* Wiley Interdiscip. Rev. RNA, 2012. **3**(3): p. 369-384.

121. Jenkins, J.L., et al., *Comparison of a preQ$_1$ riboswitch aptamer in metabolite-bound and free states with implications for gene regulation.* J. Biol. Chem., 2011. **286**(28): p. 24626-24637.
122. Garst, A.D., et al., *Crystal structure of the lysine riboswitch regulatory mRNA element.* J. Biol. Chem., 2008. **283**(33): p. 22347-22351.
123. Serganov, A., L.L. Huang, and D.J. Patel, *Structural insights into amino acid binding and gene control by a lysine riboswitch.* Nature, 2008. **455**(7217): p. 1263-U76.
124. Wilson, R.C., et al., *Tuning riboswitch regulation through conformational selection.* J. Mol. Biol., 2011. **405**(4): p. 926-938.
125. Reining, A., et al., *Three-state mechanism couples ligand and temperature sensing in riboswitches.* Nature, 2013. **499**(7458): p. 355-359.
126. Furtig, B., et al., *Multiple conformational states of riboswitches fine-tune gene regulation.* Curr. Opin. Struct. Biol., 2015. **30**: p. 112-124.
127. Davies, J. and D. Davies, *Origins and Evolution of Antibiotic Resistance.* Microbiol. Mol. Biol. Rev., 2010. **74**(3): p. 417-433.
128. Mulhbacher, J., P. St-Pierre, and D.A. Lafontaine, *Therapeutic applications of ribozymes and riboswitches.* Curr. Opin. Pharmacol., 2010. **10**(5): p. 551-556.
129. Sudarsan, N., et al., *Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine.* Chem. Biol., 2005. **12**(12): p. 1325-1335.
130. Kim, J.N., et al., *Design and antimicrobial action of purine analogues that bind guanine riboswitches.* ACS Chem. Biol., 2009. **4**(11): p. 915-927.
131. Mulhbacher, J., et al., *Novel riboswitch ligand analogs as selective inhibitors of guanine-related metabolic pathways.* PLoS Pathog., 2010. **6**(4): p. e1000865.
132. Blount, K.F., et al., *Novel riboswitch-binding flavin analog that protects mice against Clostridium difficile infection without inhibiting cecal flora.* Antimicrob. Agents Chemother., 2015. **59**(9): p. 5736-5746.
133. Ott, E., et al., *The RFN riboswitch of Bacillus subtilis is a target for the antibiotic roseoflavin produced by Streptomyces davawensis.* RNA Biol., 2009. **6**(3): p. 276-280.
134. Lee, E.R., K.F. Blount, and R.R. Breaker, *Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression.* RNA Biol., 2009. **6**(2): p. 187-194.
135. Ster, C., et al., *Experimental treatment of Staphylococcus aureus bovine intramammary infection using a guanine riboswitch ligand analog.* J. Dairy Sci., 2013. **96**(2): p. 1000-1008.
136. Hanke, C.A. and H. Gohlke, *Force field dependence of riboswitch dynamics*, in *Computational methods for understanding riboswitches*, S.-J. Chen and D.H. Burke-Aguero, Editors. 2015, Academic Press: Burlington. p. 163-191.
137. Gilbert, S.D., et al., *Adaptive ligand binding by the purine riboswitch in the recognition of guanine and adenine analogs.* Structure, 2009. **17**(6): p. 857-868.
138. Noeske, J., et al., *Ligand binding and conformational changes in the purine-binding riboswitch aptamer domains*, in *Non-Protein Coding RNAs*, N.G. Walter, Editor. 2009.
139. Wacker, A., et al., *Mechanisms for differentiation between cognate and near-cognate ligands by purine riboswitches.* RNA Biol., 2012. **9**(5): p. 672-680.
140. Jain, N., et al., *Heterogeneity and dynamics of the ligand recognition mode in purine-sensing riboswitches.* Biochemistry, 2010. **49**(17): p. 3703-3714.
141. Noeske, J., et al., *An intermolecular base triple as the basis of ligand specificity and affinity in the guanine- and adenine-sensing riboswitch RNAs.* Proc. Natl. Acad. Sci. U. S. A., 2005. **102**(5): p. 1372-1377.
142. Gilbert, S.D., et al., *Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain.* J. Mol. Biol., 2006. **359**(3): p. 754-768.

143. Stoddard, C.D., S.D. Gilbert, and R.T. Batey, *Ligand-dependent folding of the three-way junction in the purine riboswitch.* RNA, 2008. **14**(4): p. 675-684.

144. Ottink, O.M., et al., *Ligand-induced folding of the guanine-sensing riboswitch is controlled by a combined predetermined–induced fit mechanism.* RNA, 2007. **13**(12): p. 2202-2212.

145. Rieder, R., et al., *Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control.* ChemBioChem, 2007. **8**(8): p. 896-902.

146. Noeske, J., et al., *Interplay of 'induced fit' and preorganization in the ligand induced folding of the aptamer domain of the guanine binding riboswitch.* Nucleic Acids Res., 2007. **35**(2): p. 572-583.

147. Gilbert, S.D., et al., *Mutational analysis of the purine riboswitch aptamer domain.* Biochemistry, 2007. **46**(46): p. 13297-13309.

148. Noeske, J., H. Schwalbe, and J. Wöhnert, *Metal-ion binding and metal-ion induced folding of the adenine-sensing riboswitch aptamer domain.* Nucleic Acids Res., 2007. **35**(15): p. 5262-5273.

149. Lemay, J.F., et al., *Folding of the adenine riboswitch.* Chem. Biol., 2006. **13**(8): p. 857-868.

150. de la Pena, M., D. Dufour, and J. Gallego, *Three-way RNA junctions with remote tertiary contacts: A recurrent and highly versatile fold.* RNA, 2009. **15**(11): p. 1949-1964.

151. Buck, J., et al., *Dissecting the influence of Mg^{2+} on 3D architecture and ligand-binding of the guanine-sensing riboswitch aptamer domain.* Nucleic Acids Res., 2010. **38**(12): p. 4143-4153.

152. Buck, J., et al., *Influence of ground-state structure and Mg^{2+} binding on folding kinetics of the guanine-sensing riboswitch aptamer domain.* Nucleic Acids Res., 2011. **39**(22): p. 9768-9778.

153. Prychyna, O., et al., *Sequence-dependent folding and unfolding of ligand-bound purine riboswitches.* Biopolymers, 2009. **91**(11): p. 953-965.

154. Buck, J., et al., *Time-resolved NMR methods resolving ligand-induced RNA folding at atomic resolution.* Proc. Natl. Acad. Sci. U. S. A., 2007. **104**(40): p. 15699-15704.

155. Brenner, M.D., et al., *Multivector fluorescence analysis of the xpt guanine riboswitch aptamer domain and the conformational role of guanine.* Biochemistry, 2010. **49**(8): p. 1596-1605.

156. Sun, E.I. and D.A. Rodionov, *Computational analysis of riboswitch-based regulation.* Biochim. Biophys. Acta, Gene Regul. Mech., 2014. **1839**(10): p. 900-907.

157. Clote, P., *Computational Prediction of Riboswitches*, in *Computational methods for understanding riboswitches*, S.-J. Chen and D.H. Burke-Aguero, Editors. 2015, Academic Press: Burlington. p. 287-312.

158. Sanbonmatsu, K.Y., *Dynamics of riboswitches: Molecular simulations.* Biochim. Biophys. Acta, Gene Regul. Mech., 2014. **1839**(10): p. 1046-1050.

159. Villa, A., J. Wöhnert, and G. Stock, *Molecular dynamics simulation study of the binding of purine bases to the aptamer domain of the guanine sensing riboswitch.* Nucleic Acids Res., 2009. **37**(14): p. 4774-4786.

160. Ling, B., et al., *Theoretical studies on the interaction of modified pyrimidines and purines with purine riboswitch.* J. Mol. Graph. Model., 2009. **28**(1): p. 37-45.

161. Sharma, M., 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, 2009. **15**(9): p. 1673-1692.

162. Priyakumar, U.D., *Atomistic details of the ligand discrimination mechanism of SMK/SAM-III riboswitch.* J. Phys. Chem. B, 2010. **114**(30): p. 9920-9925.

163. Gong, Z., et al., *Insights into ligand binding to preQ₁ riboswitch aptamer from molecular dynamics simulations.* PLoS One, 2014. **9**(3): p. e92247.
164. Banas, P., et al., *Molecular mechanism of preQ₁ riboswitch action: A molecular dynamics study.* J. Phys. Chem. B, 2012. **116**(42): p. 12721-12734.
165. Kelley, J.M. and D. Hamelberg, *Atomistic basis for the on–off signaling mechanism in SAM-II riboswitch.* Nucleic Acids Res., 2010. **38**(4): p. 1392-1400.
166. Sharma, P., et al., *Modeling the noncovalent interactions at the metabolite binding site in purine riboswitches.* J. Mol. Model., 2009. **15**(6): p. 633-649.
167. Sund, J., C. Lind, and J. Aqvist, *Binding site preorganization and ligand discrimination in the purine riboswitch.* J. Phys. Chem. B, 2015. **119**(3): p. 773-782.
168. Huang, W., et al., *A mechanism for S-adenosyl methionine assisted formation of a riboswitch conformation: a small molecule with a strong arm.* Nucleic Acids Res., 2009. **37**(19): p. 6528-6539.
169. Priyakumar, U. and A.D. MacKerell Jr., *Role of the adenine ligand on the stabilization of the secondary and tertiary interactions in the adenine riboswitch.* J. Mol. Biol., 2010. **396**(5): p. 1422-1438.
170. Nguyen, P.H., P. Derreumaux, and G. Stock, *Energy flow and long-range correlations in guanine-binding riboswitch: A nonequilibrium molecular dynamics study.* J. Phys. Chem. B, 2009. **113**(27): p. 9340-9347.
171. Allner, O., L. Nilsson, and A. Villa, *Loop-loop interaction in an adenine-sensing riboswitch: A molecular dynamics study.* RNA, 2013. **19**(7): p. 916-926.
172. Lutz, B., et al., *Differences between cotranscriptional and free riboswitch folding.* Nucleic Acids Res., 2014. **42**(4): p. 2687-2696.
173. Leach, A.R., *Molecular modelling : principles and applications*. 2nd ed. 2001, Harlow, England ; New York: Prentice Hall.
174. Lindorff-Larsen, K., et al., *Systematic validation of protein force fields against experimental data.* PLoS One, 2012. **7**(2): p. e32131.
175. Martin-Garcia, F., et al., *Comparing molecular dynamics force fields in the essential subspace.* PLoS One, 2015. **10**(3).
176. Hashem, Y. and P. Auffinger, *A short guide for molecular dynamics simulations of RNA systems.* Methods, 2009. **47**(3): p. 187-197.
177. Cino, E.A., W.Y. Choy, and M. Karttunen, *Comparison of secondary structure formation using 10 different force fields in microsecond molecular dynamics simulations.* J. Chem. Theory Comput., 2012. **8**(8): p. 2725-2740.
178. Cornell, W.D., et al., *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules.* J. Am. Chem. Soc., 1995. **117**(117): p. 5179-5197.
179. Atkins, P.W. and J. De Paula, *Physical Chemistry*. 2006: Macmillan Higher Education.
180. Cheatham III, T.E. and D.A. Case, *Twenty-five years of nucleic acid simulations.* Biopolymers, 2013. **99**(12): p. 969-977.
181. Wang, J.M., 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., 2000. **21**(12): p. 1049-1074.
182. Case, D.A., et al., *The Amber biomolecular simulation programs.* J. Comput. Chem., 2005. **26**(16): p. 1668-1688.
183. Sponer, J., et al., *Molecular dynamics simulations of nucleic acids. From tetranucleotides to the ribosome.* J. Phys. Chem. Lett., 2014. **5**(10): p. 1771-1782.
184. Sponer, J., X.H. Cang, and T.E. Cheatham III, *Molecular dynamics simulations of G-DNA and perspectives on the simulation of nucleic acid structures.* Methods, 2012. **57**(1): p. 25-39.
185. *Biochemical Nomenclature and Related Documents*. 2nd ed. 1992: Portland Press.

186. Perez, A., et al., *Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers.* Biophys. J., 2007. **92**(11): p. 3817-3829.
187. Besseova, I., et al., *Dependence of A-RNA simulations on the choice of the force field and salt strength.* Phys. Chem. Chem. Phys., 2009. **11**(45): p. 10701-10711.
188. Reblova, K., et al., *Structure, dynamics, and elasticity of free 16S rRNA helix 44 studied by molecular dynamics simulations.* Biopolymers, 2006. **82**(5): p. 504-520.
189. Besseova, I., et al., *Simulations of A-RNA duplexes. The effect of sequence, solute force field, water model, and salt concentration.* J. Phys. Chem. B, 2012. **116**(33): p. 9899-9916.
190. Mlynsky, V., et al., *Extensive molecular dynamics simulations showing that canonical G8 and protonated A38H$^+$ forms are most consistent with crystal structures of hairpin ribozyme.* J. Phys. Chem. B, 2010. **114**(19): p. 6642-6652.
191. Zgarbova, M., et al., *Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles.* J. Chem. Theory Comput., 2011. **7**(9): p. 2886-2902.
192. Yildirim, I., et al., *Reparameterization of RNA chi torsion parameters for the AMBER force field and comparison to NMR spectra for cytidine and uridine.* J. Chem. Theory Comput., 2010. **6**(5): p. 1520-1531.
193. Yildirim, I., et al., *Revision of AMBER torsional parameters for RNA improves free energy predictions for tetramer duplexes with GC and iGiC base pairs.* J. Chem. Theory Comput., 2012. **8**(1): p. 172-181.
194. Banas, P., et al., *Performance of molecular mechanics force fields for RNA simulations: Stability of UUCG and GNRA hairpins.* J. Chem. Theory Comput., 2010. **6**(12): p. 3836-3849.
195. Bowman, J.C., et al., *Cations in charge: magnesium ions in RNA folding and catalysis.* Curr. Opin. Struct. Biol., 2012. **22**(3): p. 262-272.
196. Pyle, A.M., *Metal ions in the structure and function of RNA.* J. Biol. Inorg. Chem., 2002. **7**(7-8): p. 679-690.
197. Draper, D.E., *A guide to ions and RNA structure.* RNA, 2004. **10**(3): p. 335-343.
198. Draper, D.E., D. Grilley, and A.M. Soto, *Ions and RNA folding.* Annu. Rev. Biophys. Biomol. Struct., 2005. **34**: p. 221-243.
199. Ohtaki, H. and T. Radnai, *Structure and dynamics of hydrated ions.* Chem. Rev., 1993. **93**(3): p. 1157-1204.
200. Mills, R. and V.M.M. Lobo, *Self-diffusion in electrolyte solutions: a critical examination of data compiled from the literature.* 1989, Amsterdam: Elsevier.
201. Auffinger, P., *Ions in molecular dynamics simulations of RNA systems*, in *RNA 3D Structure Analysis and Prediction*, N. Leontis and E. Westhof, Editors. 2012, Springer: Berlin, Heidelberg. p. 299-318.
202. Li, P.F., et al., *Rational design of Particle Mesh Ewald compatible Lennard-Jones parameters for +2 metal cations in explicit solvent.* J. Chem. Theory Comput., 2013. **9**(6): p. 2733-2748.
203. Aqvist, J., *Modeling of ion ligand interactions in solutions and biomolecules.* Theochem-J. Mol. Struc., 1992. **88**: p. 135-152.
204. Allner, O., L. Nilsson, and A. Villa, *Magnesium ion-water coordination and exchange in biomolecular simulations.* J. Chem. Theory Comput., 2012. **8**(4): p. 1493-1502.
205. Li, P.F. and K.M. Merz, *Taking into account the ion-induced dipole interaction in the nonbonded model of ions.* J. Chem. Theory Comput., 2014. **10**(1): p. 289-297.
206. Li, P.F., L.F. Song, and K.M. Merz, *Parameterization of highly charged metal ions using the 12-6-4 LJ-type nonbonded model in explicit water.* J. Phys. Chem. B, 2015. **119**(3): p. 883-895.

207. Hansmann, U.H.E., *Parallel tempering algorithm for conformational studies of biological molecules.* Chem. Phys. Lett., 1997. **281**(1-3): p. 140-150.
208. Sugita, Y. and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding.* Chem. Phys. Lett., 1999. **314**(1-2): p. 141-151.
209. Earl, D.J. and M.W. Deem, *Parallel tempering: Theory, applications, and new perspectives.* Phys. Chem. Chem. Phys., 2005. **7**(23): p. 3910-3916.
210. Kone, A. and D.A. Kofke, *Selection of temperature intervals for parallel-tempering simulations.* J. Chem. Phys., 2005. **122**(20).
211. Rathore, N., M. Chopra, and J.J. de Pablo, *Optimal allocation of replicas in parallel tempering simulations.* J. Chem. Phys., 2005. **122**(2).
212. Denschlag, R., M. Lingenheil, and P. Tavan, *Optimal temperature ladders in replica exchange simulations.* Chem. Phys. Lett., 2009. **473**(1-3): p. 193-195.
213. Nadler, W. and U.H.E. Hansmann, *Optimized explicit-solvent replica exchange molecular dynamics from scratch.* J. Phys. Chem. B, 2008. **112**(34): p. 10386-10387.
214. Sindhikara, D.J., Y. Meng, and A.E. Roitberg, *Exchange frequency in replica exchange molecular dynamics.* J. Chem. Phys., 2008. **128**: p. 024103.
215. Sindhikara, D.J., D.J. Emerson, and A.E. Roitberg, *Exchange often and properly in replica exchange molecular dynamics.* J. Chem. Theory Comput., 2010. **6**(9): p. 2804-2808.
216. Trebst, S., M. Troyer, and U.H.E. Hansmann, *Optimized parallel tempering simulations of proteins.* J. Chem. Phys., 2006. **124**(17).
217. Garcia, A.E., H. Herce, and D. Paschek, *Simulations of temperature and pressure unfolding of peptides and proteins with replica exchange molecular dynamics.* Annu. Rep. Comput. Chem., 2006. **2**: p. 83-95.
218. Garcia, A.E. and D. Paschek, *Simulation of the pressure and temperature folding/unfolding equilibrium of a small RNA hairpin.* J. Am. Chem. Soc., 2008. **130**(3): p. 815-817.
219. Kar, P., W. Nadler, and U.H.E. Hansmann, *Microcanonical replica exchange molecular dynamics simulation of proteins.* Phys. Rev. E., 2009. **80**(5): p. 056703.
220. Li, Y. and T. Schlick, *Modeling DNA polymerase μ motions: subtle transitions before chemistry.* Biophys. J., 2010. **99**(10): p. 3463-3472.
221. Zheng, W.H., et al., *Kinetic network study of the diversity and temperature dependence of Trp-cage folding pathways: Combining transition path theory with stochastic simulations.* J. Phys. Chem. B, 2011. **115**(6): p. 1512-1523.
222. Jiang, W., et al., *Calculation of free energy landscape in multi-dimensions with Hamiltonian-exchange umbrella sampling on petascale supercomputer.* J. Chem. Theory Comput., 2012. **8**(11): p. 4672-4680.
223. Kumar, S., et al., *The weighted histogram analysis method for free-energy calculations on biomolecules .1. The method.* J. Comput. Chem., 1992. **13**(8): p. 1011-1021.
224. Chodera, J.D., et al., *Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations.* J. Chem. Theory Comput., 2007. **3**(1): p. 26-41.
225. Shirts, M.R. and J.D. Chodera, *Statistically optimal analysis of samples from multiple equilibrium states.* J. Chem. Phys., 2008. **129**(12): p. 124105.
226. Fajer, M., R.V. Swift, and J.A. McCammon, *Using multistate free energy techniques to improve the efficiency of replica exchange accelerated molecular dynamics.* J. Comput. Chem., 2009. **30**(11): p. 1719-1725.
227. Chodera, J.D., et al., *Long-time protein folding dynamics from short-time molecular dynamics simulations.* Multiscale Model. Sim., 2006. **5**(4): p. 1214-1226.

228. Chu, B.K., et al., *A tail of two peptide amphiphiles: Effect of conjugation with hydrophobic polymer on folding of peptide sequences.* Biomacromolecules, 2014. **15**(9): p. 3313-3320.

229. Moradi, M. and E. Tajkhorshid, *Computational recipe for efficient description of large-scale conformational changes in biomolecular systems.* J. Chem. Theory Comput., 2014. **10**(7): p. 2866-2880.

230. Li, P.C., et al., *Multidimensional umbrella sampling and replica-exchange molecular dynamics simulations for structure prediction of transmembrane helix dimers.* J. Comput. Chem., 2014. **35**(4): p. 300-308.

231. Gohlke, H., L.A. Kuhn, and D.A. Case, *Change in protein flexibility upon complex formation: Analysis of Ras-Raf using molecular dynamics and a molecular framework approach.* Proteins, 2004. **56**(2): p. 322-337.

232. Jacobs, D.J., et al., *Protein flexibility predictions using graph theory.* Proteins: Struct., Funct., Genet., 2001. **44**(2): p. 150-165.

233. Jacobs, D.J., L.A. Kuhn, and M.F. Thorpe, *Flexible and rigid regions in proteins*, in *Rigidity theory and applications*. 2002, Springer. p. 357-384.

234. Thorpe, M.F., et al., *Protein flexibility and dynamics using constraint theory.* J. Mol. Graph. Model., 2001. **19**(1): p. 60-69.

235. Hespenheide, B.M., D.J. Jacobs, and M.F. Thorpe, *Structural rigidity in the capsid assembly of cowpea chlorotic mottle virus.* J. Phys.: Condens. Matter, 2004. **16**(44): p. S5055-S5064.

236. Wang, Y.M., et al., *Global ribosome motions revealed with elastic network model.* J. Struct. Biol., 2004. **147**(3): p. 302-314.

237. Fulle, S. and H. Gohlke, *Constraint counting on RNA structures: Linking flexibility and function.* Methods, 2009. **49**(2): p. 181-188.

238. Fulle, S. and H. Gohlke, *Statics of the ribosomal exit tunnel: Implications for cotranslational peptide folding, elongation regulation, and antibiotics binding.* J. Mol. Biol., 2009. **387**(2): p. 502-517.

239. Rathi, P.C., et al., *Statics of biomacromolecules*, in *Modeling of Molecular Properties*. 2011. p. 281-299.

240. Pfleger, C., et al., *Constraint Network Analysis (CNA): a Python software package for efficiently linking biomacromolecular structure, flexibility, (thermo-)stability, and function.* J. Chem. Inf. Model., 2013. **53**(4): p. 1007-1015.

241. Jacobs, D.J. and M.F. Thorpe, *Generic rigidity percolation: The pebble game.* Phys. Rev. Lett., 1995. **75**(22): p. 4051-4054.

242. Jacobs, D.J. and B. Hendrickson, *An algorithm for two-dimensional rigidity percolation: The pebble game.* J. Comp. Phys., 1997. **137**(2): p. 346-365.

243. Fulle, S. and H. Gohlke, *Analyzing the flexibility of RNA structures by constraint counting.* Biophys. J., 2008. **94**(11): p. 4202-4219.

244. Schafmeister, C.E.A.F., W.S. Ross, and V. Romanovski, *LEaP*. 1995: University of California, San Francisco

245. Saini, J.S., et al., *Determinants of the species selectivity of oxazolidinone antibiotics targeting the large ribosomal subunit.* Biol. Chem., 2013. **394**(11): p. 1529-1541.

246. Robinson, H., et al., *Hexahydrated magnesium ions bind in the deep major groove and at the outer mouth of A-form nucleic acid duplexes.* Nucleic Acids Res., 2000. **28**(8): p. 1760-1766.

247. Dixon, N., et al., *Reengineering orthogonally selective riboswitches.* Proc. Natl. Acad. Sci. U. S. A., 2010. **107**(7): p. 2830-2835.

248. Woodson, S.A., *Metal ions and RNA folding: a highly charged topic with a dynamic future.* Curr. Opin. Chem. Biol., 2005. **9**(2): p. 104-109.

249. Hanke, C.A. and H. Gohlke, *How tertiary interactions between the L2 and L3 loops affect the dynamics of the distant ligand binding site in the guanine sensing riboswitch*, in *Proceedings of the NIC Symposium 2014*, K. Binder, G. Münster, and M. Kremer, Editors. 2014: Jülich. p. 99-106.
250. Hanke, C.A. and H. Gohlke, *Ligand-mediated and tertiary interactions cooperatively stabilize the P1 region in the guanine-sensing riboswitch.* PLOS ONE, 2017. **12**(6): p. e0179271.
251. Cooper, A. and D.T.F. Dryden, *Allostery without Conformational Change - a Plausible Model.* Eur. Biophys. J., 1984. **11**(2): p. 103-109.
252. Best, R.B. and J. Mittal, *Balance between α and β structures in ab initio protein folding.* J. Phys. Chem. B, 2010. **114**(26): p. 8790-8798.
253. Best, R.B. and J. Mittal, *Microscopic events in β-hairpin folding from alternative unfolded ensembles.* Proc. Natl. Acad. Sci. U. S. A., 2011. **108**(27): p. 11087-11092.
254. Weigand, J.E. and B. Suess, *Aptamers and riboswitches: perspectives in biotechnology.* Appl. Microbiol. Biotechnol., 2009. **85**(2): p. 229-236.
255. Groher, F. and B. Suess, *Synthetic riboswitches - A tool comes of age.* Biochim. Biophys. Acta, Gene Regul. Mech., 2014. **1839**(10): p. 964-973.