U n i v e r z i t a   P a l a c k é h o   v   O l o m o u c i

Přírodovědecká fakulta

Katedra fyzikální chemie

Molekulární dynamika RNA přepínačů

Diplomová práce

Autor:	Bc. Michal Janeček
Školitel:	doc. Mgr. Pavel Banáš, Ph.D.
Studijní program:	N 1407 Chemie
Studijní obor:	Fyzikální chemie
Forma studia:	Prezenční

Olomouc 2016

Palacky University Olomouc

Faculty of Science

Department of Physical Chemistry

Molecular dynamics of riboswitches

Diploma Thesis

Author:	Bc. Michal Janeček
Supervisor:	doc. Mgr. Pavel Banáš, Ph.D.
Study programme:	N 1407 Chemistry
Major:	Physical Chemistry
Study form:	Daily

Olomouc 2016

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně a veškeré literární prameny použité v této práci jsem uvedl v seznamu použité literatury.

V Olomouci dne:

                                                            _____

                                                                    Michal Janeček

Bibliografická identifikace

Autor: Bc. Michal Janeček

Název práce: Molekulární dynamika RNA přepínačů

Typ práce: Diplomová

Vedoucí práce: doc. Mgr. Pavel Banáš, Ph.D.

Rok obhajoby: 2016

Pracoviště: Katedra fyzikální chemie, UP v Olomouci

Abstrakt:

RNA přepínače jsou strukturní prvky lokalizované v nekódující části mRNA, které regulují expresi příslušného genu, a to tak, že v závislosti na koncentraci určité malé molekuly mění svou konformaci. Protože se nacházejí především v bakteriích, včetně patogenních druhů, mohly by RNA přepínače být slibnými cíli pro nově navrhovaná léčiva. Proto je pochopení mechanismu účinku RNA přepínače důležitým předmětem zájmu. Jsou navrhovány dva obecné typy mechanismu účinku a navázání ligandu – „conformational capture" a „induced fit". Ke zjištění, o jaký mechanismus jde, může významně přispět rentgenové difrakce a určení trojrozměrné struktury RNA přepínače ve stavu s navázaným ligandem a bez něj. Nicméně, i když má tato metoda klíčový význam pro studium struktury RNA přepínačů, její prediktibilita může být do jisté míry omezena, protože uspořádání v monokrystalu může být ovlivněno krystalovými kontakty. Tento artefakt však může být odhalen pomocí výpočetních metod.

V této práci jsme použili současné výpočetní metody (konkrétně molekulovou dynamiku, metadynamiku a „replica exchange" molekulovou dynamiku) pro studium strukturní dynamiky dvou RNA přepínačů, a také abychom objasnili, jakou úlohu v mechanismu účinku těchto RNA přepínačů hrají struktury určené rentgenovou difrakcí. Konkrétně jsme studovali strukturní dynamiku SAM-III RNA přepínače z bakterie *Enterococcus faecalis*, a preQ1-I RNA přepínače z bakterie *Thermoanaerobacter tengcongensis*, a to jak ve stavu s navázaným ligandem, tak bez ligandu.

Klíčová slova:

RNA přepínače, riboswitche, molekulární dynamika, metadynamika, molekulární dynamika s výměnou replik, REST2, REFT, SAM-III riboswitch, preQ1-I riboswitch

Počet stran: 70

Jazyk: Anglický

Bibliographic identification

Author: Bc. Michal Janeček

Title: Molecular dynamics of riboswitches

Type of Thesis: Diploma

Supervisor: Assoc. Prof. Mgr. Pavel Banáš, Ph.D.

Department: Department of Physical Chemistry, UP Olomouc

The Year of Presentation: 2016

Abstract:

Riboswitches are structural elements of non-coding mRNA regulating expression of their associated genes by changing their conformation in response to concentration of small molecules. As they are present mainly in bacteria, including the pathogenic bacteria species, they could be promising targets of newly designed therapeutics. Thus understanding of mechanism of riboswitch action is important object of interest. Two general types of action and ligand binding are suggested – conformational selection vs. induced fit. The mechanism can be revealed by determining the 3D structure of the riboswitch in ligand-bound and ligand-free states using X-ray diffraction, however, despite crucial role of the X-ray diffraction, its structural predictability might be to some extent limited by crystal packing. Nonetheless, the effect of crystal packing on the structure can be revealed using computational methods.

In this work, we applied contemporary computational methods (in particular molecular dynamics, metadynamics and replica exchange molecular dynamics) to study structural dynamics of two riboswitches and to reveal the roles of the structures obtained by X-ray crystallography in mechanisms of their action. In particular, we studied structural dynamics of SAM-III riboswitch from *Enterococcus faecalis*, and preQ1-I riboswitch from *Thermoanaerobacter tengcongensis*, both in their ligand-bound and ligand-free states, respectively.

Keywords:

riboswitches, molecular dynamics, metadynamics, replica exchange molecular dynamics, REST2, REFT, SAM-III riboswitch, preQ1-I riboswitch

Number of Pages: 70

Language: English

Contents

1 Introduction

Riboswitches provides important strategy of living organisms in regulation of the expression of their genes in response to the stimuli from inner and outer environment. In particular, they are structural elements of mRNA regulating gene expression by changing their conformation in response to concentration of small molecules.[1] Interestingly, they are composed exclusively from the mRNA molecule (or more precisely its non-coding part). The riboswitch is both able to specifically bind the small molecule and regulate the expression of its associated gene. As they are present mainly in bacteria, including the pathogenic bacteria species (e. g., SAM-III riboswitch occurs in *Enterococci* or *Streptococci*[2] and preQ1-I riboswitch in *Listeriae*, *Staphylococci* or *Streptococci*),[3] they could be promising targets of newly designed therapeutics, which aims to obstruct the metabolism of the bacteria.

In order to design such therapeutics, the mechanism of riboswitch action and ligand binding must be well understood. Two general types of action and ligand binding are suggested – conformational selection vs. induced fit.[1] The three-dimensional structures in the ligand-bound and ligand-free states obtained by X-ray diffraction might significantly help in revealing the mechanism of action in particular riboswitch. The method provides atomic-level resolution, however, the conformation of the structure can be affected by crystal packing. The artefact can be revealed by contemporary computational methods that can simulate the dynamics of the biomacromolecules in their natural environment, water solution, and thus provide useful complementary information.

This work aims to apply contemporary computational methods to crystal structures of two riboswitches, SAM-III riboswitch from *Enterococcus faecalis* and preQ1-I riboswitch from *Thermoanaerobacter tengcongensis*, in order to reveal the role of the structures in mechanism of the riboswitch action. In particular, we used both classical MD simulations as well as several enhanced sampling methods to study structural dynamics of the above mentioned riboswitches.

2 Theoretical part

2.1 Riboswitches

### 2.1.1 Definition and general properties

Riboswitches are the mRNA structural elements that have the ability to bind small molecules or ions as ligands and to regulate gene expression as the response to the ligand binding.[1] Whereas besides the mRNA molecule the ligand binding does not require any other macromolecular factor. They are very common in bacteria, some were also found in fungi[4,5] and plants[6] (e. g. the TPP riboswitch) and some are predicted in archaea based on bioinformatical studies.[7] Riboswitches usually regulate the production of proteins involved in synthesis, degradation or transport of the ligand. They are usually located in 5' untranslated region (UTR) of mRNA but some riboswitches were also found in introns of some species of fungi and bacteria (e. g. TPP riboswitch in *Neurospora crassa*[5] and c-di-GMP-II riboswitch in *Clostridium difficile*)[8] and in 3' UTR of some plants species (e. g. TPP riboswitch in *Arabidopsis thaliana*).[6] As they usually act on the same molecule they are encoded in, they belong to so-called cis acting elements.[1] The existence of riboswitches was proven in 2002[9–11] after earlier speculations.[12]

Two basic parts of the riboswitch can be usually distinguished, aptamer domain and expression platform (Figures 1 and 2). The aptamer domain specifically binds the ligand whereas the expression platform regulates gene expression.[1] The conformational changes of the aptamer caused by ligand binding are transferred to the expression platform by so-called switching sequence.
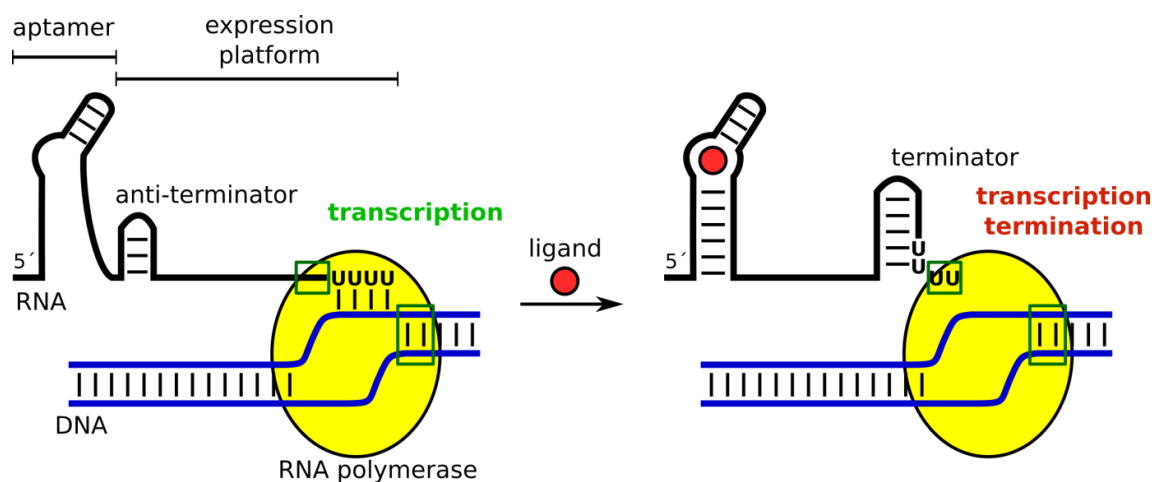
Figure 1 - Scheme of transcription termination by transcription acting "OFF" riboswitch. Within the **RNA polymerase** the growing **RNA** is bound to the template **DNA** by base pairing, and this interaction is also supported by binding to the protein (green square). The DNA is bound to protein as well (second green square). When **anti-terminator hairpin** in **expression platform** is formed, the transcription can carry on despite the relatively weakly bound **polyU sequence** as the RNA-DNA interaction is supported also by base-pairing with adjacent sequence. If the **ligand** is bound to the **aptamer**, its conformational changes lead to the formation of **terminator hairpin** involving sequence adjacent to polyU, which weaken RNA-DNA interaction even more and thus extracts the RNA from the polymerase.

(Note: proportions, number of uracils, neither of paired bases do not correspond to the reality in order to efficiently illustrate the mechanism within one picture)
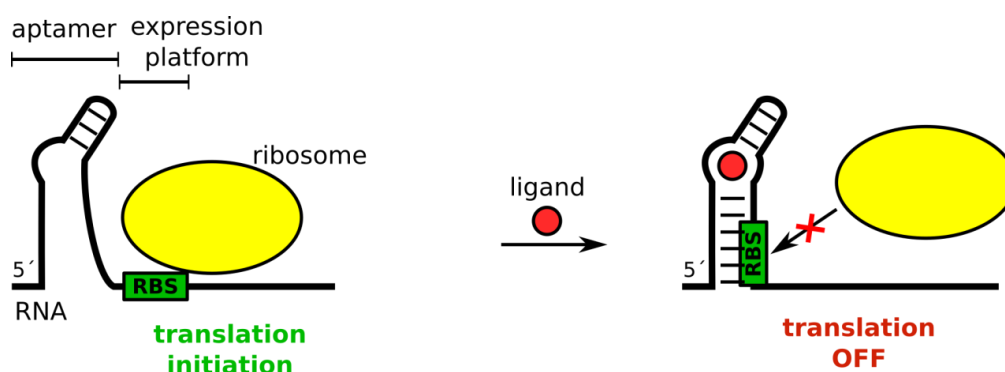
Figure 2 - Scheme of translation initiation regulation by "OFF" riboswitch. If the **ligand** is not bound to the **aptamer** the ribosome binding site (**RBS**) is accessible for the ribosome. If the ligand binds to the aptamer, the RBS pairs with complementary sequence from the RNA, which sequesters the RBS, so that the ribosome cannot bind it.

2.1.2 Strategy of gene expression regulation

The riboswitches usually use two basic distinct strategies for the regulation of the gene expression. They regulate either transcription termination or translation initiation.[13]

A transcription-acting riboswitch regulates transcription of DNA into RNA by switching between two distinct conformations of the expression platform involving either terminator or anti-terminator hairpin. The conformation of the expression platform is determined by the conformation of the aptamer domain, which is in turn sensitive to the presence or absence of the ligand. When so-called terminator hairpin is formed, its presence disrupts binding of RNA polymerase to the mRNA and terminates the transcription. Alternatively, formation of the so-called anti-terminator hairpin prevents formation of the terminator hairpin and thus allows transcription to carry on. The detail scheme of transcription termination is depicted on Figure 1.[14]

The transcription takes place in so-called transcription elongation complex, which is stabilized by 8-10 bp long pairing between the growing RNA and the template DNA and by 4-5 nt of RNA which bind to the enzyme. When the terminator hairpin sequence is transcribed and the anti-terminator hairpin is not formed, the terminator hairpin folds on a nascent mRNA outside the transcription elongation complex. The formation of the terminator hairpin and its binding to NusA protein then leads to stalling of the RNA polymerase. In addition, at this moment, the transcribed RNA is relatively weakly bound to DNA by the polyuracil sequence (as the U-A base pair has only two hydrogen bonds), and thus the formation of the terminator loop leads to increased likelihood of DNA/RNA hybrid destabilization, extraction of nascent mRNA, and transcription termination.

The formation of the terminator hairpin is allowed only when the anti-terminator hairpin is not formed. Formation of anti-terminator or terminator hairpins is determined by conformation of the aptamer domain and thus presence or absence of the cognate ligand in the binding site of aptamer. The ligand binding to the riboswitch aptamer either destabilizes anti-terminator (then the riboswitch acts as so-called genetic "OFF" switch, which is the most common in natural riboswitches)[13] or contrary the ligand binding stabilizes the anti-terminator (genetic "ON" switch, which is found rarely in natural riboswitches up to date, e. g., adenine riboswitch regulating *ydhL* gene[15] or glycine riboswitch regulating *gcvT* gene[16] both in *Bacillus subtilis*).

As the terminator must form nearby the ternary elongation complex in order to extract the growing RNA from the complex, it must form fast enough before the transcription elongates the RNA too much. The terminator formation at physiological ligand concentrations occurs at time scale of seconds. The same time-scale is observed also for the transcription slowed down by the polyU sequence. Thus transcription-regulating riboswitches are under kinetic control as the ligand binding, RNA structural rearrangement and the transcription are of the same time-scale and their speeds play crucial roles.[1,17]

The other group of riboswitches are translation-regulating riboswitches. They utilize also two distinct conformations of their expression platform for regulation of the translation initiation (Figure 2). Again, the switch between these two conformations of the expression platform is driven by the presence or absence of the ligand in the aptamer ligand-binding site. In the first conformation, the ribosome-binding site (RBS) of the mRNA is accessible for the ribosome binding and allows thus translation to be initiated. In the second conformation the nucleobases of the RBS form a base-pairing contacts with complementary RNA sequence, so the RBS cannot be accessed by the ribosome and the translation initiation is suppressed. In prokaryota, so-called Shine-Dalgarno (SD) sequence serves as the RBS. Only if the complementary sequence of the 16S rRNA pairs with the SD sequence of mRNA, the 30S rRNA subunit can be bound to the 16S rRNA, and the translation process can begin.

Similarly as in case of transcription-acting riboswitches, the ligand binding can either cause the sequestering of the RBS (genetic "OFF" switch) or contrary stabilize the structure where the RBS is accessible (genetic "ON" switch).

In contrast to the transcription-regulating riboswitches, the translation-regulating riboswitches have enough time to reach the equilibrium of the ligand binding process so the riboswitches are under thermodynamic control.[17]

Beside the described regulation pathways, others mechanisms are worth noting: In some riboswitches the dual transcription and translation regulation was found, where the translational riboswitch is located adjacent and downstream of the transcriptional riboswitch (e. g. SAM-II/SAM-V tandem arrangement in *Pelagibacter ubique*)[18] so the associated gene can be regulated at both levels. Another interesting mechanism is the

antisense regulation (found e. g. in *Listeria monocytogenes*).[19] This mechanism involves riboswitch, which regulates transcription of non-coding RNA which is antisense to the adjacent downstream gene whereas the antisense RNA inhibits expression of the complementary gene. And at last some riboswitches carry ribozyme as expression platform and thus perform self-cleavage (e. g. *glmS* riboswitch).[20]

2.1.3   Strategy of ligand binding and recognition

Till now the following types of molecules or ions have been found to serve as the ligands for riboswitch aptamers: nucleic acids precursors,[15,21] aminoacids,[16,22] enzyme cofactors,[23,24] second messengers,[25] metal ions[26,27] and inorganic anions.[28] As the chemical structures of the ligands are very diverse, each riboswitch utilizes its particular and unique mechanism for ligand recognition and binding. Moreover the same ligand can be recognized by more structurally distinct riboswitch aptamers. The unique ligand binding mechanism is formed using various secondary or tertiary structural elements such as internal bulges, terminal loops, junctions or pseudoknots. Usually every structural element is crucial and necessary for the riboswitch function.[1]

The mechanisms of the ligand binding can be classified into two groups: conformational capture and induced fit.[29] In case of conformational capture, first the aptamer must reach the appropriate conformation and then the ligand binds to it thereby fixing the aptamer conformation (Figure 3A). Contrary in case of induced fit mechanism, initially the ligand binds to the unfolded aptamer, which induces its conformation changes (Figure 3B). These two strategies can be combined. Usually the bound ligand is solvent inaccessible[1] so the induced fit mechanism take part in ligand binding in most of the cases.

The most important attribute of ligand recognition is its high specificity.[1] The riboswitch aptamer is usually able to distinguish between structurally very similar molecules, e. g., the guanine riboswitch from *Bacillus subtilis* discriminates between adenine and guanine by at least 100,000 fold.[15]
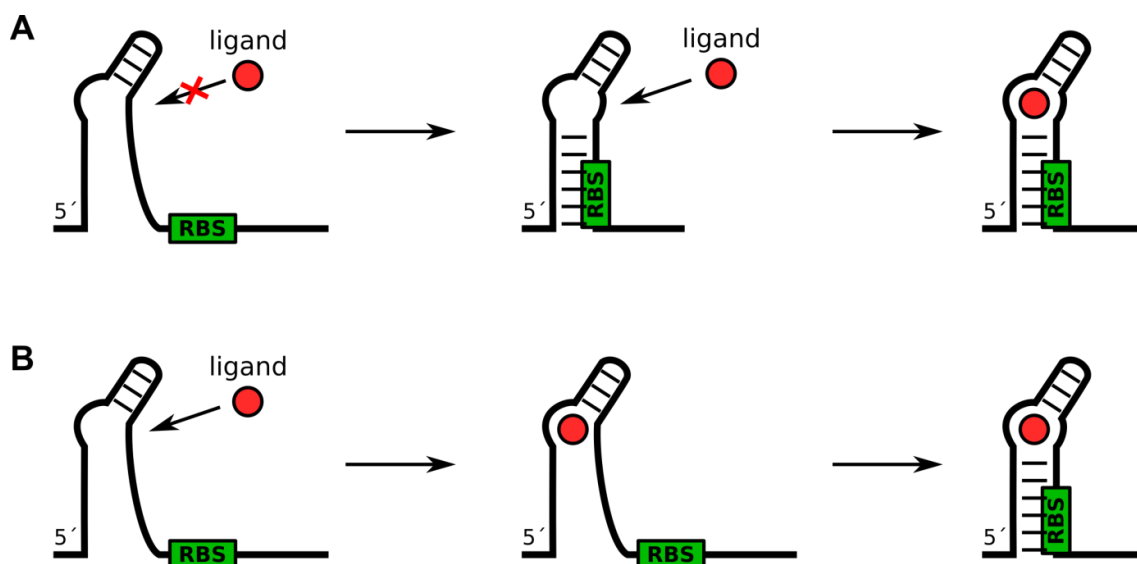
Figure 3 - Two basic mechanisms of the ligand binding by riboswitches demonstrated on translation regulating riboswitch **A** - conformational capture **B** - induced fit

2.1.4 Classification

The riboswitches are usually classified into families according to their cognate ligands and each family is further divided into classes according to typical secondary structure of aptamer and expression platform.[30] For instance, all the riboswitches that bind pre-queosine1 (preQ1) belong to preQ1 family and the family currently includes at least three classes (Figure 10) denoted preQ1-I, preQ1-II and preQ1-III, each having distinct secondary structure.[31] Moreover, riboswitch classes might be further divided into different types, each type containing a distinct conserved sequence, e. g., the preQ1-I riboswitch class involves three types denoted type 1, type 2 and type 3 (each carrying its own characteristic pattern within L1 and L2 loop and P2 stem).[3,31]

The exception in this classification system is purine riboswitch family. All its members have similar secondary structure but they are sensitive to multiple distinct ligands. The family is than divided into classes according to the ligand type. In particular, the purine family involves, e. g., adenine, guanine or 2'-deoxyguanosine classes.

The terminology of the classification, however, is not strict. Some authors[32–34] use the terms "family" and "subfamily" instead of "class" and "type", respectively, adopting these terms from the protein classification. Note that in protein classification, the term family involves proteins with clear sequence homology while the subfamily contains

subset of proteins of corresponding family having even higher sequence homology. Moreover the "superfamily" term (also denoted as "clan")[32,33] is used for proteins that share the same structural and functional features but the sequence relationship is not relevant. Applying the terminology to above mentioned example, the preQ1 binding riboswitches are distinguished into three families whereas preQ1-I family contains subfamilies 1, 2 and 3. The superfamily term is useful, e. g., for SAM binding riboswitches where SAM-I and SAM-IV families have almost the same ligand binding core but other structural elements differ, so these families are grouped to SAM-I superfamily.[33] As the classification into superfamilies, families and subfamilies is more universal, this classification will be used in following text. It is worth noting that currently at least 25 distinct riboswitch families have been found.[35]

2.1.5 SAM sensing riboswitches

The S-adenosyl-L-methionine (SAM, Figure 4A) is an important cofactor in every living organism. It serves as methyl source in most of methylation reactions as its methyl group is relatively weakly bound to the sulphur. The SAM is synthesized from methionine and ATP with assistance of SAM synthetases.[33]

In most bacteria the production of SAM is regulated by SAM sensing riboswitches. The SAM sensing riboswitches were shown to specifically bind the SAM while discriminating against the S-adenosyl-L-homocysteine (SAH, Figure 4B), which rests after the SAM is used as methyl donor in methylation reactions. Currently, three distinct groups of riboswitches binding SAM have been discovered, each representing unique way how to achieve the SAM binding and discriminate against the SAH: SAM-I and SAM-II superfamilies and SAM-III family.[33] Each of them is widespread among different bacteria species.

The SAM-I superfamily consists of SAM-I, SAM-IV and SAM-I/IV families sharing similar arrangement of the binding site, but the other part of aptamer differs and their genetic relationship was not found.[33]
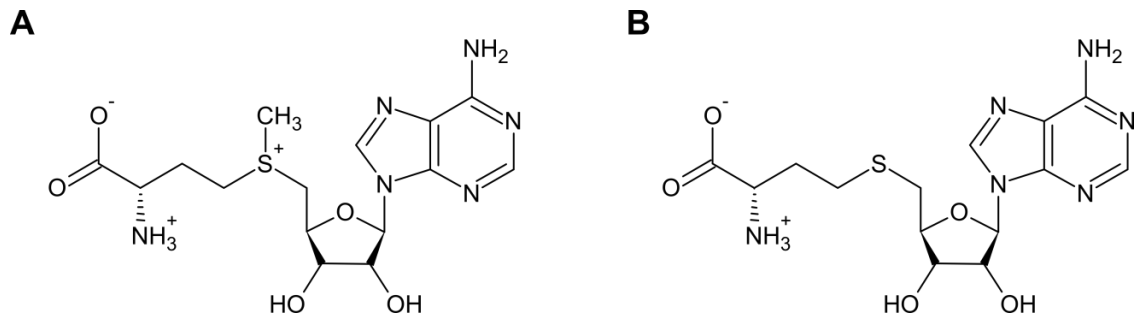
Figure 4 - **A** S-adenosyl-L-methionine **B** S-adenosyl-L-homocysteine
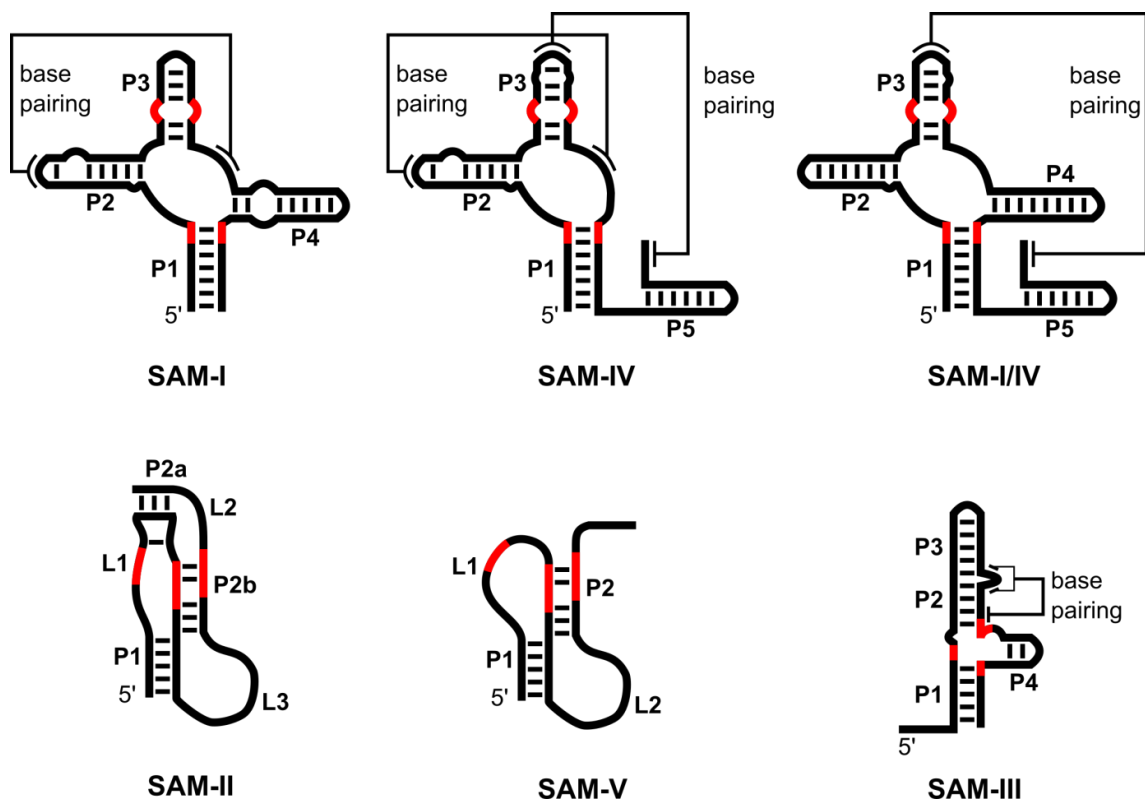
Figure 5 - The secondary structure schemes of all currently discovered SAM riboswitch families in ligand bound states. Sequences involved in SAM binding are highlighted in red. **SAM-I**[36] - Four stems P1-P4 are connected by four-way junction. The SAM is bound by sequences from P1 and P3 stems. **SAM-IV**[34] - Compared to SAM-I, the P4 helix is missing, another helix P5 is present and additional base pair interactions takes place between P3 terminal loop and bases located adjacent downstream the P5 stem. **SAM-I/IV**[32] - Both P4 and P5 helix are present, but in contrast to SAM-I and SAM-IV, base pairs between four-way junction and P2 terminal loop are missing. **SAM-II**[37]- By coaxial stacking between the P1 and P2b helices the H-type pseudoknot is formed. The SAM is bound using nucleotides from P2b stem and L1 loop. **SAM-V**[18] - Compared to SAM-II the P2 stem has different structure but nucleotides involved in SAM binding remains. **SAM-III**[38] - The P1, P2 and P4 helices are connected by three-way junction, which provides nucleotides for the SAM binding.

17

The SAM-I family (also called "S-box") has been found abundant in *Firmicutes* phylum namely *Bacilli* and *Clostridia* classes with exception of the *Lactobacillales* order, where it is rather rare.[24] The family members regulates various genes involved in SAM synthesis and very often the multiple riboswitches are found in one genome each regulating different gene.[39] The regulation mostly occurs at level of transcription termination, although some cases of SD sequence sequestration have been also found.[2] The aptamer comprises four helices connected by a four-way junction whereas the SAM is bound between two parallelly oriented helices (Figure 5).[36]

The very similar binding core is present in SAM-IV family though it is formed in context of different outside structural elements. In contrast to SAM-I, SAM-IV family lacks one helix at junction region, so its structure contains only three helices that are connected by a three-way junction while the fourth helix is located elsewhere (Figure 5). The family has been discovered in *Actinomycetales* order bacteria where it controls expression of various genes involved in sulphur metabolism. It is presumed to act at the level of translation as many SAM-IV aptamers were found nearby the translation starting site of the downstream gene but further evidences are needed to prove this hypothesis.[34]

SAM-I/IV family aptamer carries four helices connected by four-way junction and one additional helix (Figure 5) thus combining structural features of both SAM-I and SAM-IV riboswitches. Currently, only few representatives of this family have been found spread among multiple unrelated bacterial phyla and also some small marine organisms (e. g. *Tetrahymena thermophila*).[7,32]

The SAM-II superfamily consists of SAM-II and SAM-V families, which likewise the previous superfamily share almost the same binding core although the other aptamer elements differ.

The SAM-II family has been found abundant in *Proteobacteria* phylum. In most cases it regulates the translation initiation of genes involved in L-methionine biosynthesis but transcription termination regulating members have been also identified. The aptamer forms typical H-type pseudoknot whereas nucleotides from three strands are involved in the SAM binding (Figure 5).[33,37]

The SAM-V family was found exclusively in marine species of *Proteobacteria* and *Bacteroidetes* phyla. According to contemporary knowledge it appears to control translation initiation of genes involved in L-methionine synthesis. A crystal structure was not determined yet but the sequence analysis and its comparison with SAM-II family implies that the nucleotides crucial for SAM binding are present and H-type pseudoknot is formed (Figure 5).[18,33]

The SAM-III family (also called "S_{MK}-box") was identified almost exclusively in *Lactobacillales* order. It was found to always regulate the *metK* gene (which leaded to the name "S_{MK}-box") at the level of translation initiation.[24,38] The *metK* gene encodes SAM synthetase enzyme which catalyses the SAM synthesis from ATP and L-methionine. The aptamer in its ligand-bound state is formed by four stems. Three of them are connected by three-way junction involving ligand-binding site (Figures 5, 6A). The lengths of the sequences of P3 and P4 stems strongly vary in various species (see Figure 6A).[24] This is quite unique among known riboswitches, which might also explain the fact that this riboswitch family has not been identified yet in other bacteria orders and phyla.

So the genetic regulation using SAM binding riboswitches occurs in most bacteria and more distinct riboswitch families have developed independently in different bacterial lineages. However, there are bacteria species where the SAM sensing riboswitches have not been found up to date although the SAM is commonly used there in methylation reactions. This may be explained by the fact that the other riboswitches might be awaiting the discovery[33] or another mechanism of regulation might take place (e. g., protein based mechanism).[2]
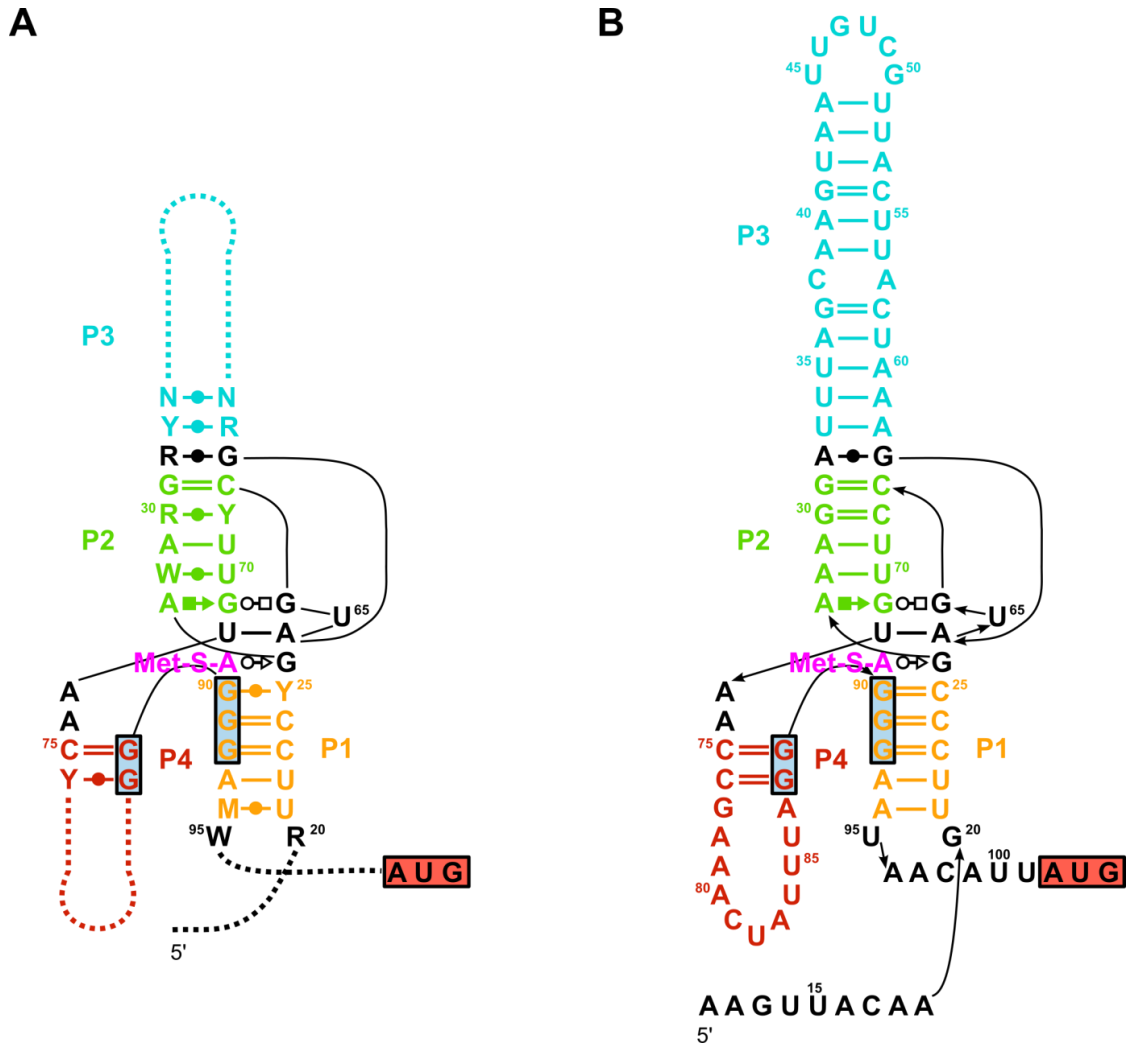
Figure 6 - Secondary structures of ligand bound (i. e. OFF state) SAM-III riboswitch. The secondary and tertiary interactions according to *E. faecalis* crystal structure (PDB ID: 3E5C).[38] In this state the riboswitch forms four stems P1-P4 (in colours) and three-way junction which connects P1, P2 and P4 stems. The SAM (in magenta) is bound within the three-way junction. The SD sequence (blue border) is paired and thus ribosome inaccessible. **A** - The consensus structure of whole family[24] using IUB nucleotide codes,[40] Leontis–Westhof symbols[41] and dashed lines representing sequences with variable length, composition and pairing **B** - The full length sequence of the riboswitch from *E. faecalis*[24]
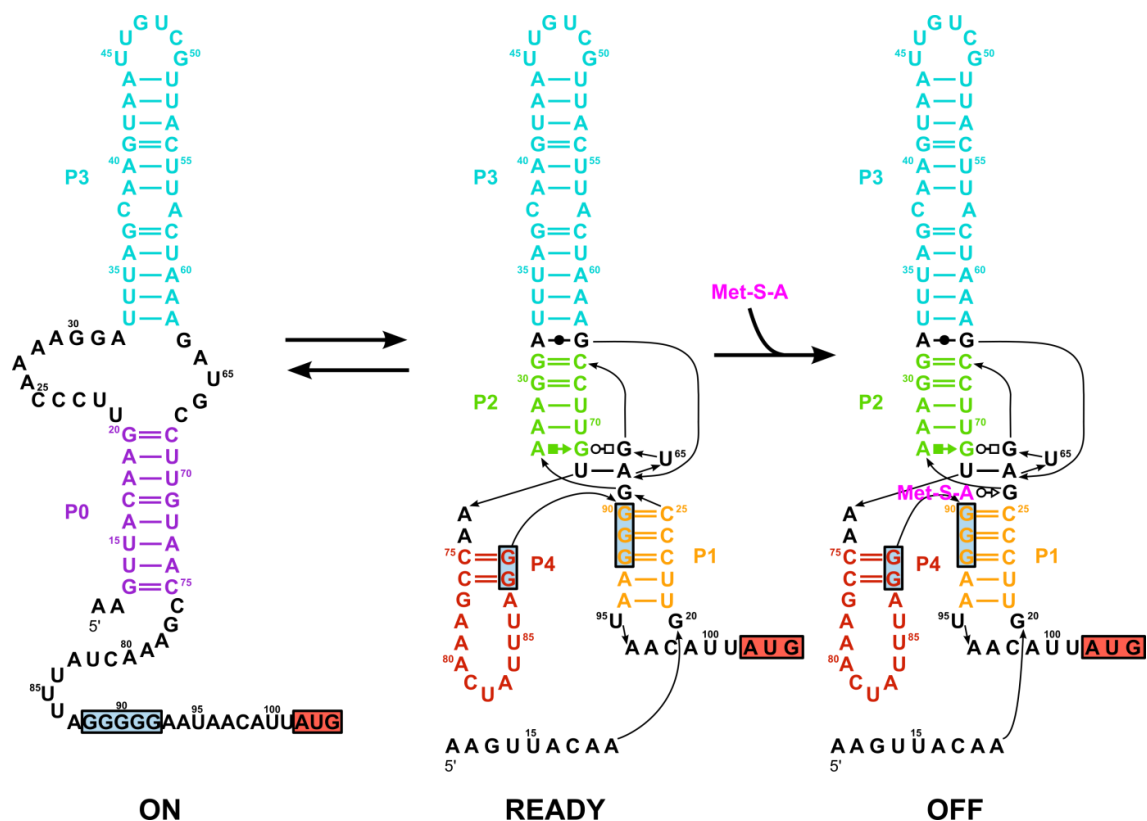
Figure 7 - The most probable mechanism of the SAM-III riboswitch action. In the **ON** state, the SD sequence is fully accessible and P0 helix is formed.[42–44] This state is in equilibrium with so-called **READY** state (at least 0.4% of the riboswitch molecules should populate READY state in the solution according to mutagenesis and prediction software analysis of full length riboswitch) which corresponds to the conformation very close to the **OFF** state[42–44] with P1-P4 stems, three-way junction and ligand binding pocket formed. If the SAM is present, it binds into the binding pocket of READY state and stabilizes the riboswitch in the OFF conformation[38,42,43] (the apparent dissociation constant of the SAM was estimated[38] to 0.5 μM for full length riboswitch at 298K so standard free energy change between ON and OFF state should by ~8.6 kcal/mol).

21
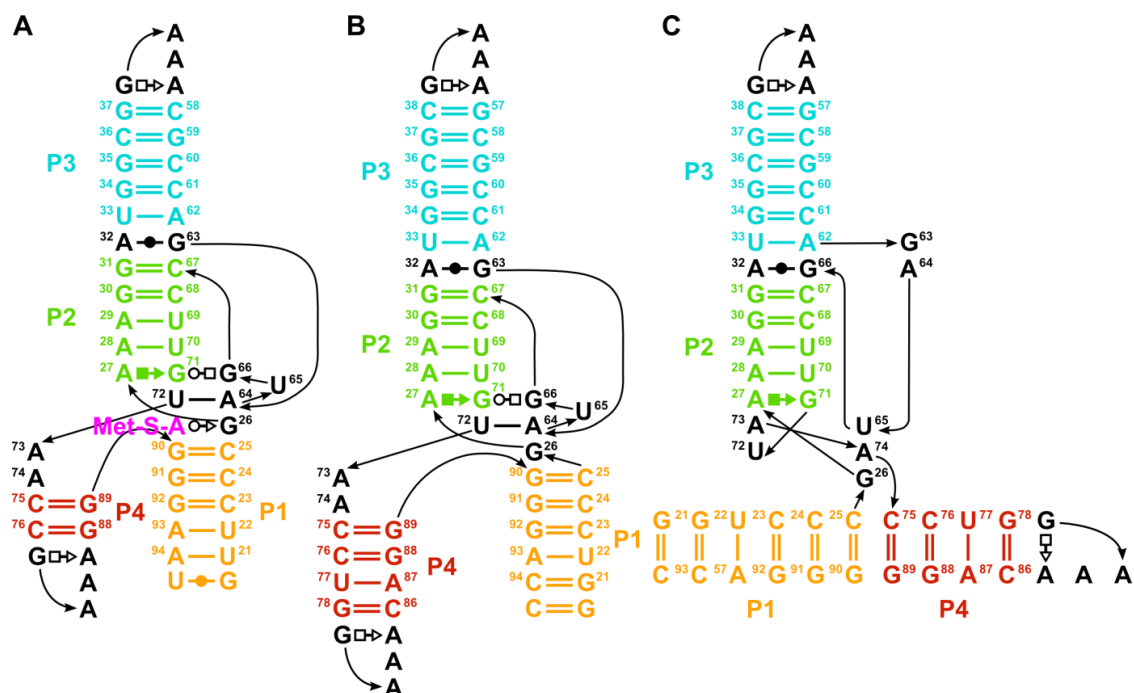
Figure 8 – Secondary structures as observed in the crystal structures of the SAM-III riboswitch from *E. faecalis* in SAM-bound OFF state (**A**, PDB ID: 3E5C) and SAM free ready1 and ready2 states (**B** and **C**, respectively, unpublished data),[44] The stems and the S-adenosylmethionine (Met-S-A) are highlighted in colours and the bases numbering is consistent with previous studies.[24] Compared to wild-type riboswitch from *E. faecalis* (Figure 6B) the 5' and 3' ends were cut and P3 and P4 stems were modified in order to obtain high resolution crystal structure.

2.1.6 SAM-III riboswitch from *Enterococcus faecalis*

SAM-III riboswitch has been widely studied and characterized in *Enterococcus faecalis* species although some studies involved also other species such as *Lactobacillus plantarum*[24,45] and *Streptococcus gordonii*.[24] The studies suggest that the riboswitch behaviour should be very similar through the whole family although such comparison might be limited by the observation that the single nucleobases substitution might cause large changes in the thermodynamics and kinetics of riboswitch action.

The riboswitch acts at the level of translation initiation as proved by nitrocellulose binding assay.[45] This experiment demonstrated that SAM-mediated inhibition of 30S ribosomal subunit binding to the *metK* gene RNA transcript from *E. faecalis* occurred in vitro. During the transcript lifetime the riboswitch can bind and release the SAM several times as the half-time of riboswitch-SAM complex is several times lower than the half-life of the transcript in the cell, which was determined by 2-aminopurinefluorescence assay and qRT-PCR,[46] respectively. Thus the riboswitch can make multiple decisions reacting at given cell conditions.

The ribosome binding is inhibited by SD sequence sequestration which was proved by crystal structure of the ligand-bound riboswitch (PDB ID: 3E5C,[38] Figure 8A). The SD sequence is GGGGG representing unusual sequence compared to other known riboswitches. Another specific feature of the SAM-III riboswitch family is that the aptamer and expression platform are not separated but on the contrary integrated within single structure responsible for both ligand binding and SD sequestration.

If the SAM is bound, the SD sequence is sequestered and riboswitch is in so-called OFF state (Figure 7). In such a state the riboswitch adopts a conformation, in which four A-RNA stems P1-P4 are formed and the P1,P2 and P4 stems are connected by a three-way junction as was shown by the crystal structure (PDB ID: 3E5C,[38] Figure 8A). The junction region and nearby nucleobases form so-called ligand binding pocket. Some SD bases are part of this pocket.

In ligand-free state, riboswitch is in so-called ON state and the SD sequence is fully accessible because P1, P2 and P4 helices are disrupted and instead P0 helix is formed (Figure 7) as was observed by NMR based base pairs probing,[43] SHAPE analysis,[42] and also unpublished crystal structure.[44] As the structure do not poses the ligand binding

pocket, which formation appears to be inseparably connected with SD sequence sequestration, the open question was how the ligand binding influences the structure. This issue was solved by discovery of so-called READY state (Figure 7) which was observed by both NMR based base pairs probing[43] and SHAPE analysis[42] for the structure with destabilized ON structure by mutagenesis. The state adopts structure very close to the OFF state, so that the ligand binding pocket is formed.

As was demonstrated by mutagenesis,[42] the equilibrium between the ON and the READY states can be regulated by various bases substitutions. The most crucial is that even single base substitution, G77 to U, shifts the equilibrium towards the READY state as the substitution allows U77-A87 base pair formation and thus adjacent U-A pairs formation in P4 stem (Figure 6B). As was calculated by structure prediction programs such as RNAfold,[47] the G77 to U substitution should stabilize the P4 stem by ~3.2 kcal/mol. Thus the standard free energy difference between ON and READY states of wild-type riboswitch should not be higher and at least 0.4% of riboswitch molecules should populate the READY state at 298K in solution. So according to contemporary knowledge, the suggested mechanism of riboswitch acting is illustrated on Figure 7.

New question, however, arose when the crystal structure of the READY state was determined.[44] The structure revealed two crystalographically independent conformations coexisting in one crystal. As expected, one was corresponding to previously described READY state (henceforth labelled ready1, Figure 8B). But the second one possesses a different arrangement of the three-way junction and lacks the ligand binding pocket (henceforth labelled ready2, Figure 8C). Furthermore, after soaking the SAM into the crystal, the SAM was successfully bound only by the ready1 structure (the resulting ligand-bound structure henceforth labelled soaked1) but not by the ready2 structure (the resulting structure henceforth labelled soaked2 although it is ligand-free). The question is whether the ready2 structure might play some role in the riboswitch mechanism. This could be revealed using contemporary computational methods.

Another interesting feature of this riboswitch is the specificity of SAM binding and discrimination against SAH. The SAM is bound in the binding pocket by adenosyl moiety and the positively charged sulfonium ion. The methionine tail seems to be ignored as there were not observed any interactions with the riboswitch in the crystal

structure.[38] In vitro assays demonstrated that SAM-III riboswitch prefers the SAM against SAH by at least 100 fold. The crucial differences used to discriminate between SAM and SAH are the most likely the charge on the sulphur atom and distinct conformation of the ligand.[38]

2.1.7 preQ1 sensing riboswitches

PreQ1 is a precursor of modified nucleoside queuosine (Q, Figure 9). The Q usually occurs at wobble position of GUN anticodon (N means any nucleotide)[40] of certain tRNAs which increases the efficiency of reading degenerated codons as the Q pairs well with both C and U and also enhances the maintenance of reading frame among others biological roles.[48,49] The prokaryotes synthesize Q de novo from GTP (Figure 9) whereas the eukaryotes acquire the Q from their diet or intestinal bacteria in form of the free nucleobase, queuine. The preQ1 and associated riboswitches are thus found only in the prokaryotes. The riboswitches bind selectively preQ1 and discriminate against its natural analogue preQ0 although not very tightly (e. g., the apparent dissociation constant of preQ0 is in preQ1-I riboswitch from *Thermoanaerobacter tengcongensis* only 17-fold lower than that of preQ1).[50] Three distinct families of preQ1 sensing riboswitches were identified.[3,31]

The preQ1-I family was found widespread among members of *Firmicutes*, *Proteobacteria* and *Fusobacteria* phyla where it usually resides in 5' UTR of the operon which encodes QueC, QueD, QueE and QueF enzymes (Figure 9) or sometimes in 5' UTR of a gene of putative Q transporter.[51] The family carries the smallest aptamer of all currently known riboswitches with minimal length of 34 nt.[31,52] In the ligand-bound state, the aptamer forms H-type pseudoknot and ligand resides between two stems P1 and P2 (Figure 10).[50,53] The aptamer was found to be associated with expression platforms regulating translation initiation[50,51] as well as those regulating transcription termination[51,53] which demonstrates the modularity of the preQ1-I aptamer domain. Three subfamilies were identified so far, each carrying specific sequence within L1, L2 and/or P2 area (see Figure 10).[31,52] The subfamily 1 members contain 95% conserved 5'-UUCR sequence at the beginning of the L1 loop and 5'-UC at the end of the L2. The subfamily 2 members carry 5'-YUAR sequence in the P2 stem and 5'-AC at the end of the L2. Finaly, the subfamily 3 members contain 5'-UUC at the beginning of the L1 and 5'-CCU within the P2 stem. However, some species also carry patterns of multiple
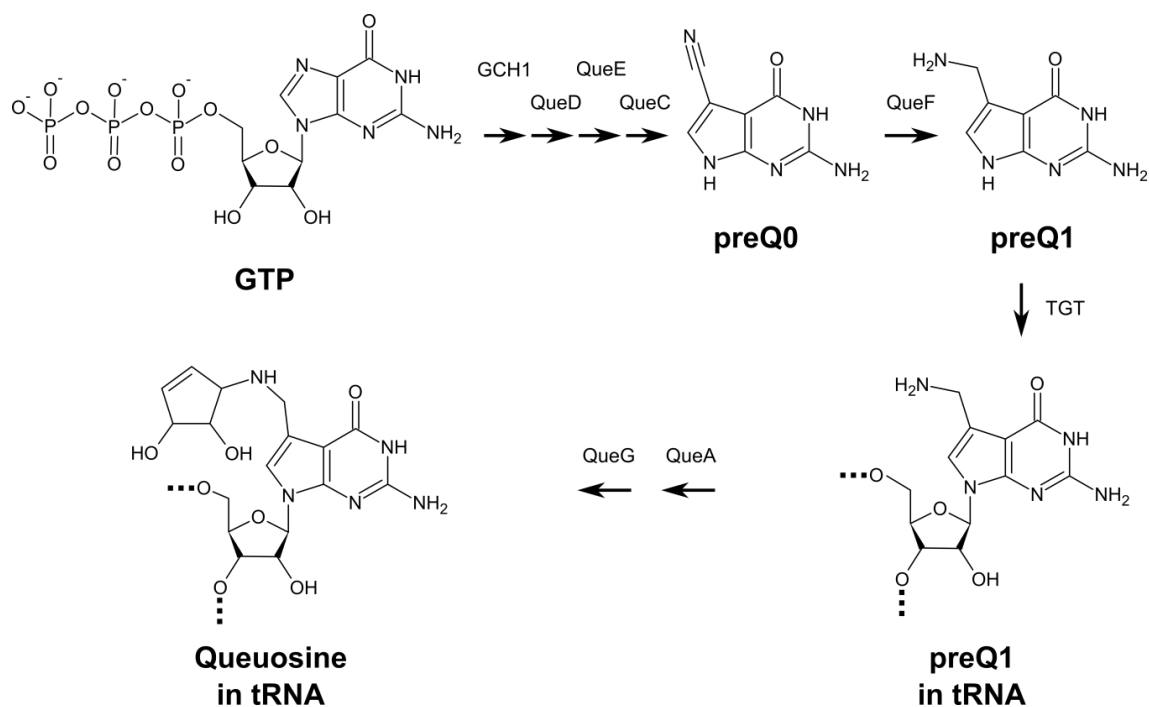
Figure 9 - Scheme of biosynthetic pathway of queuosine from GTP highlighting the role of preQ0 and preQ1. First, the **GTP** is conversed into **preQ0** by cascade of multiple enzymes, GTP cyclohydrolase I (**GCH1**), 6-carboxy-5,6,7,8-tetrahydropterin synthase (**QueD** or also called ToyB),7-carboxy-7-deazaguanine synthase (**QueE** or ToyC) and preQ0 synthetase (**QueC** or ToyM). The preQ0 reductase (**Que F**) transforms preQ0 into **preQ1** which than substitutes guanine in his position in tRNA using tRNA:guanine transglycosylase (**TGT**). preQ1 in the tRNA is finally transformed into **queuosine** in two steps involving S-adenosylmethionine:tRNA ribosyltransferase-isomerase (**QueA**) and oQ reductase (**QueG**).[49]

subfamilies, e. g., the subfamily 1 riboswitch from *Thermoanaerobacter tengcongensis* carries 5'-UCG at the beginning of the L1 loop, which classifies it to the subfamily 1,[3] but at the end of the L2 loop it carries 5'-AC, a typical feature of the subfamily 2.

The preQ1-II family was found exclusively in *Lactobacillales* order, where their members regulate translation of genes of membrane proteins likely transporting Q.[3,31] Its aptamer carries relatively large (average is 58 nt) and complex structure (Figure 10). This may be the reason why it was not identified in other orders. The affinities for preQ1 and discrimination against its analogues are comparable to I family.

preQ1-I **preQ1-II** **preQ1-III**
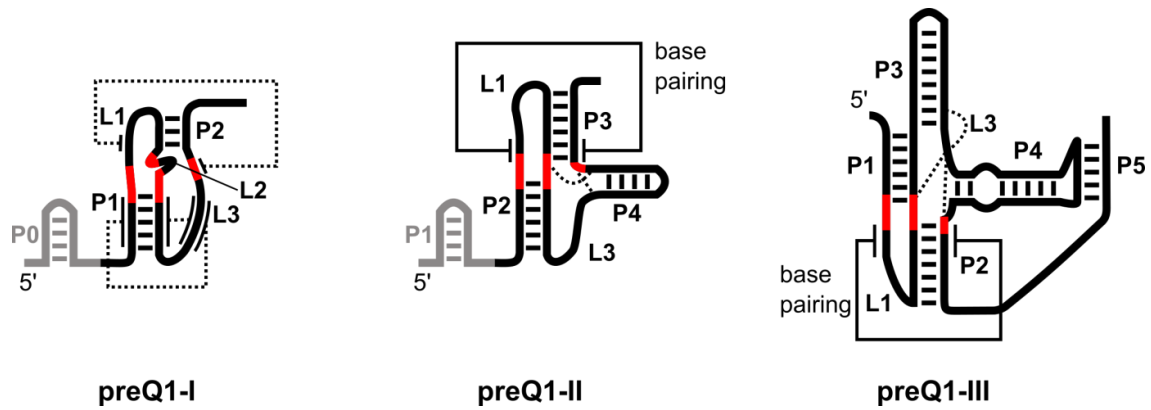
Figure 10 - The scheme of secondary structures of all preQ1 binding riboswitch families in ligand bound states. Sequences involved in preQ1 binding are highlighted in red, dashed lines represent base pairing or other hydrogen bonds. **preQ1-I**[50,53]- In the presence of the preQ1 the riboswitch forms H-type pseudoknot involving two stems (P1, P2) and three loops (L1, L2, L3). The additional P0 hairpin loop was found only in some species (such as *Bacillus subtilis*)[52] and is not necessary for the riboswitch function (only slightly increases affinity for ligand). **preQ-II**[54,55] - The riboswitch forms atypical H-type pseudoknot, which contains additional P4 hairpin within L3 loop. P1 hairpin preceding P2 stem is not necessary for the preQ1 binding and is not conserved. Stems numbering is according to crystal structure.[55] **preQ1-III**[56] - The P1 and P2 helices form HL_{out}-type pseudoknot, whereas the L3 loop contains the P3 and P4 hairpins. The P5 stem serves to RBS sequestration.

The preQ1-III family was discovered recently in *Ruminococcaceae*, the family of *Clostridiales* order bacteria.[31] It was found to regulate the same genes as the family II, also at the level of translation. Its aptamer is very complex and large (about 100 nt) and forms HL_{out}-type pseudoknot (Figure 10). Unlike the case of families I and II, the expression platform is distal to the binding core and the RBS is sequestered within the P5 stem.
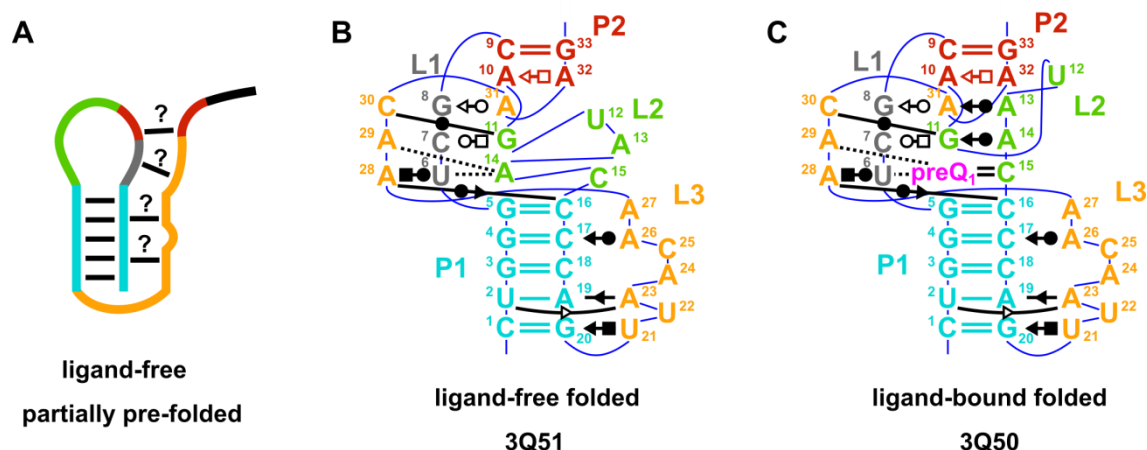
2.1.8   PreQ1-I riboswitch from *Thermoanaerobacter tengcongensis*

The preQ1 sensing family I subfamily 1 riboswitch from *Thermoanaerobacter tengcongensis* regulates expression of operon containing two genes most likely involved in the Q transfer. It acts at the level of the translation initiation involving SD sequestration as was proved at level of native mRNA sequence.[57] The SD sequence AGGGAGGU only slightly differs from the consensus (AGGAGG).[57]

In the ligand bound state, the riboswitch is folded in typical H-type pseudoknot and the first two nucleotides of SD sequence are sequestered within P2 stem (Figure 11C) as revealed by crystal structures (PDB ID: 3Q50).[50] The stability of this structure in water solution was confirmed by SAXS,[50] smFRET[51] and classical MD simulations.[21] The aptamer and expression platform are thus integrated within one domain which allows to study them together.

The ligand-free state was studied by smFRET which revealed that in solution the distance between U12 and G35 fluctuates dominantly around 45 Å, however, a diminished distance below ~30 Å was also observed in ~11% population.[51] According to coarse grain simulations, the former state can be assigned to ensemble of partially folded conformations where the P1 stem is always folded but only some of the interactions within P2 stem and between the L3 loop and P1 stem are present (Figure 11A). The latter state can be very probably assigned to folded-like structure observed in crystal (Figure 11B).[50]

The folded-like ligand-free structure adopts majority of features of ligand-bound state (Figure 11B) comprising P1 and P2 stems, L3 loops interactions with P1, L1 and L2 as well as all L1 interactions. The only changes occur at L2 region where the A14 adopts the ligand position in the ligand binding site. Thus in contrast to ligand-bound structure, the G11 ◀●A14 as well as A13 ●▶A31 base pairs and A13..A32 stacking interaction are missing. As was demonstrated by MD simulations[21] of the ligand-free folded structure adopted from the crystal structure (PDB ID: 3Q51), the lack of these interactions namely the stacking between A13 and A32 causes P2 stem instability. This may be the reason why SD sequence is accessible in the ligand-free state and why two signals corresponding to partially folded and folded-like structures (Figure 11A and B, respectively) were observed in the smFRET study.

Figure 11 - Secondary structure schemes of conformations of preQ-I riboswitch aptamer from *Thermoanaerobacter tengcongensis* in ligand-free and ligand-bound states. As predicted by coarse grain simulations combined with smFRET data,[51] the aptamer is in a solution adopts ensemble of partially folded conformations where the P1 stem is always folded but only some of the interactions within P2 stem and between the L3 loop and P1 stem are present (**A**). Crystal structures revealed folded-like conformation of ligand-free aptamer (**B**, PDB ID: 3Q51) and folded ligand-bound aptamer (**C**, PDB ID: 3Q50).

Two studies were performed exploring riboswitch behaviour and dynamics in the solution and its response to the ligand binding. The above mentioned smFRET study[51] explored also the response to preQ1 concentrations. The addition of the preQ1 increased the low-distance population showing that the ligand binds to riboswitch and stabilizes the folded conformation. The apparent dissociation constant of riboswitch-ligand complex was estimated to 69±22 nM from the course of the low-distance population at various ligand concentrations ranging from 0 to 10 µM. Within the same study, the isothermal titration calorimetry estimated the dissociation constant to 7.3±2.3 nM.

The SiM-KARTS study[57] investigated the accessibility of SD sequence and its dependence on the preQ1 concentration. The study found that the probe binding of the anti-SD to the SD sequence occurs in the "bursts". The period of "bursts", in which probe binds very to the SD of riboswitch, often alternates with period of "non-bursts", in which such binding event occurs rarely. These periods are present both without ligand and with 16 µM preQ1 concentration. The presence of preQ1 only decreases burst numerosity and burst duration. This suggests that in both absence and presence of the ligand the riboswitch switches between two distinct states: one carrying SD accessible

for the anti-SD probe binding, while the second with sequestered SD. The study also determined the binding rate constant of the probe to the SD k_{on} to $2.4 \pm 0.3 \cdot 10^6$ M^{-1}s^{-1} in absence of the ligand and $3.7 \pm 0.2 \cdot 10^6$ M^{-1}s^{-1} at 16 µM preQ1 concentration, and unbinding rate constant k_{off} to 3.7 ± 0.2 s^{-1} in absence of the ligand and 2.1 ± 0.4 s^{-1} at 16 µM preQ1 concentration.

The Go model simulations support in consistence with smFRET data[51] the induced fit model where ligand forms the pair with C15 within the partially pre-folded aptamer before the P2 stem and ligand binding pocket are formed. However this does not necessarily excludes participation of the folded ligand-free state in the riboswitch mechanism.

2.2 Computational chemistry

Full understanding of various chemical phenomena ranging from electronic transitions over chemical reactions to biomicromolecules conformational changes requires often the atomic or even electronic level description of the chemical system. However, not all phenomena can be studied by experimental techniques at such resolution. E. g. the only experimental technique capable of determining structure of macromolecules at atomic-level is X-ray diffraction which cannot be used in natural environment of biomolecules, i.e., water solutions.

Such issues might be partially overcome by so-called computational chemistry, which is ensemble of methods aiming to predict the behaviour of chemical systems usually at atomic or electronic level. The most common computational methods involve quantum mechanics (QM) based and molecular mechanics (MM) based methods.[58]

The QM based methods[59] (also called *ab initio* methods) often utilize numeric solution of time-independent electronic Schrödinger equation at various levels of approximation. The methods provide high level of accuracy but are suitable only for systems consisting of maximally tens of atoms, thus biomacromolecular systems are out of their applicability.

The MM methods[60,61] describe interactions in the chemical system by empirical potentials based on classical Newtonian mechanics, so they allow efficient description of relatively large chemical systems with reasonable computational cost. Thus the MM

based methods can be applied to systems containing around thousands of atoms, including biomacromoleculecules. As these methods are empirically based, they should be used carefully and always validated by experiments. On the other hand, when properly used, MM methods might provide useful complementary information. One of the MM based methods, which is very often used to study the biomacromolecules is classical molecular dynamics.

2.2.1 Classical molecular dynamics

The classical molecular dynamics[60,61] (MD) is a computational method, which uses MM potential to simulate development of the chemical system in time. In such MM description, all the interactions in the system are described by the pairwise interactional potential, which is function of the positions of all the particles in the system and the development in time is than simulated by numerical integrations of Newtonian equations of motion.

The all-atom MM potential describes all the interactions in the system as follows. The atoms are considered as mass points and all their bonded interactions are often described using three contributions, the energy of bonds $\sum E_{bond}$, the energy of angles $\sum E_{angle}$ and the energy of dihedral angles $\sum E_{dihedral}$ whereas the non-bonded interactions are described by energy of electrostatic interactions $\sum E_{electrostatics}$ and energy of Van der Waals interactions $\sum E_{VdW}$. The sum of those five contributions than gives total potential energy of the whole system E_{pot} (Equation 1).

$$E_{\text{pot}} = \sum E_{\text{bond}} + \sum E_{\text{angle}} + \sum E_{\text{dihedral}} + \sum E_{\text{electrostatics}} + \sum E_{\text{VdW}} \qquad (1)$$

The energy of a bond between two atoms is described as the energy of harmonic oscillator (Equation 2) where $k_{r(ij)}$ is the force constant of the "spring" between atoms i and j, the r_{ij} is the actual distance between the atoms and $r_{ij(0)}$ is the equilibrium distance.

$$E_{\text{bond}} = \frac{1}{2} k_{\text{r(ij)}} (r_{ij} - r_{ij\,(0)})^2 \qquad (2)$$

The interaction between atoms i and k, which are connected by two bonds via one atom (i. e., atom i- bond - atom j- bond - atom k) is described by the energy of the angle α_{ijk} in a similar way as a bond (Equation 3) where $k_{\alpha(ijk)}$ is the force constant of the "spring", the α_{ijk} is the actual angle and $\alpha_{ijk(0)}$ is the equilibrium angle.

$$E_{\text{angle}} = \frac{1}{2} k_{\alpha(\text{ijk})}(\alpha_{ijk} - \alpha_{ijk\,(0)})^2 \tag{3}$$

The interaction between atoms i and l which are connected by three bonds via two atoms (i. e., atom i- bond - atom j- bond - atom k- bond - atom l) is described by the energy of the dihedral angle φ_{ijkl} (Equation 4) which has periodical nature. The n is number of periods within 360°, $E_{\varphi(ijkl)}$ is the maximal energetic barrier, the φ_{ijkl} is the actual angle and $\varphi_{ijkl(0)}$ is the phase shift.

$$E_{\text{dihedral}} = \frac{1}{2} E_{\varphi(\text{ijkl})}(1 + \cos(n\varphi_{ijkl} - \varphi_{ijkl\,(0)})) \tag{4}$$

The MM uses an approximation where every atom carries a partial point charge. The interactions of those partially charged atoms as well as of ions are described by Coulomb's law (Equation 5) where r_{ij} is the actual distance between the atoms i and j, q_i and q_j their charges, ε_0 permittivity of the vacuum and ε_r relative permittivity.

$$E_{\text{electrostatics}} = \frac{1}{4\pi\varepsilon_0\varepsilon_r} \frac{q_i q_j}{r_{ij}} \tag{5}$$

The Van der Waals interaction is approximated by so-called Lennard-Jones potential (Equation 6) where r_{ij} is the actual distance between the atoms, ε_{ij} is the minimum of the energy of the potential and σ_{ij} is the distance, where attractive and repulsive contributions are just compensated and the resulting interaction energy equals zero.

32

$$E_{VdW} = 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \qquad (6)$$

From the interaction energy of the whole system, the force affecting every particle can be expressed as negative derivative of the potential energy with respect to particle position (Equation 7, for simplicity this calculation and following ones are described for the contribution in x-axes direction in Cartesian coordinate system).

$$F_x = -\frac{\partial E}{\partial x} \qquad (7)$$

Subsequently, the obtained forces F_x (being a function of coordinates and time) are combined with Newton's second law, which leads to second order differential equation (Eq 8).

$$F_x(x, y, z, t) = m \frac{\partial^2 x}{\partial t^2} \qquad (8)$$

This equation can be obtained for every coordinate of every particle so totally system of 3N equations is obtained for chemical system comprising N particles. Such a system of equations cannot be solved analytically, therefore the equations should be solved numerically. According to Newton's second law of motion the acceleration a_x can be calculated from the force F_x and the mass of particle m (Equation 9).

$$a_x = -\frac{F_x}{m} \qquad (9)$$

The known acceleration, actual velocities and usage of finite time step Δt give rise to the numerical system of equations of motion offering updated velocity v_x(Equation 10) and updated position of the particle x after timestep Δt (Equation 11). The timestep Δt must be small enough, usually 0.5 to 4 fs.

$$v_x\left(t + \frac{1}{2}\Delta t\right) = v_x\left(t - \frac{1}{2}\Delta t\right) + a_x(t)\Delta t \qquad (10)$$

$$x(t + \Delta t) = x(t) + v_x\left(t + \frac{1}{2}\Delta t\right)\Delta t \qquad (11)$$

The constants from all terms of the potential (Equations 2-6) are not determined for every single interaction and atom from the chemical system, but instead so-called atom types are defined. The atom type is an atom of one element from the same or similar chemical arrangement, which is supposed to behave in a similar way in different systems (e. g., the carbon and hydrogens from methyl group in various chemical systems). The ensemble of those parameters is called force field. Currently, variety of force fields exists, each of them designed for specific chemical systems (e. g., for nucleic acids or for proteins) or phenomena.

In case of biomacromolecules comprising thousands of atoms, the contemporary classical MD simulations are able to calculate the simulations up to microseconds in a reasonable computational time. This scale is often not enough to observe most phenomena, which are not probable on such time-scale, and in context of simulations are so-called rare events. E. g., the folding of riboswitch aptamers were observed to take around 1 second.[1] That is why so-called enhanced sampling methods were developed involving metadynamics and replica exchange molecular dynamics.

2.2.2 Metadynamics

The metadynamics[62] (MTD) is an enhanced sampling method, which accelerates rare events using so-called collective variables (CVs) and an history dependent external potential added to the system. The CV is a function of chosen system coordinates designed to describe desired rare event as efficiently as possible. During the simulation a small repulsive Gaussian potentials of width δs and height ω are added every timestep τ_G to at the actual position of CV $S(x)$. The value of external potential V_G at time t and at position $S(x)$ is as follows:

$$V_G(S(x),t) = \omega \sum_{\substack{t'=\tau_G, 2\tau_G, \dots \\ t'<t}} \exp\left(-\frac{(S(x)-s(t'))^2}{2\delta s^2}\right) \tag{12}$$

where $s(t) = S(x(t))$ which is the value of the CV at time t.

If two or more CVs are used at same time, the V_G at time t and at position $S_\alpha(x)$ (where α ranges from 1 to total number of CVs d) equals:

$$V_G(S_\alpha(x),t) = \omega \sum_{\substack{t'=\tau_G, 2\tau_G, \dots \\ t'<t}} \exp\left(-\sum_{\alpha=1}^{d} \frac{(S_\alpha(x)-s_\alpha(t'))^2}{2\delta s_\alpha^2}\right) \tag{13}$$

where the width δs_α is set for every CV and $s_\alpha(t) = S_\alpha(x(t))$.

Owing to this algorithm, the frequently sampled values of the CV are biased by numerous Gaussian potentials. Thus the system can escape local energy minima and sample other parts of the conformational space in context of the CV. Moreover the free energy profile along the CV can be reconstructed using the sum of Gaussian potentials after the MTD has converged.

The identification of one or more CVs must precede the MTD simulation and their choice is crucial for the method efficiency and reliability. The CV can be any function of coordinates, e. g., a distance, an angle or so-called coordination number. The coordination number can be efficiently used to detect the number of contacts within the r_0 distance between two groups of atoms. It is calculated as sum of switching functions (Eq 14) whereas the switching function (Eq 15) is calculated for every pair of atoms between groups A and B.

$$N = \sum_{i \in A} \sum_{j \in B} s_{ij} \tag{14}$$

$$s_{ij} = \frac{1-\left(\frac{r_{ij}}{r_0}\right)^n}{1-\left(\frac{r_{ij}}{r_0}\right)^m} \tag{15}$$

If the present distance r_{ij} is smaller than limit distance r_0 the s_{ij} is approximately 1, if $r_{ij} > r_0$ it is 0. The coefficients n and m are usually 6 and 12, respectively, and can be used to fine-tune the switching function.

The other crucial issue of the method is the convergence of the MTD simulation. The convergence was shown to be improved by introducing of so-called well-tempered metadynamics[63] (WT-MTD). The method uses an approach where the height w of each Gaussian potential depends on the external potential V_G which has been already deposited at the given position:

$$w = \omega\,\tau_G e^{-\left(\frac{V_G(S(x),t)}{\Delta T}\right)}$$
(16)

where ω is the initial height and ΔT temperature bias factor. Thus the higher is the already deposited biasing potential at the position the lower is the added Gaussian potential height at the position. Note that temperature bias factor ΔT might be used for fine-tuning of the WT-MTD, so that asymptotic behaviour of WT-MTD with ΔT of zero and infinity corresponds to classical unbiased simulations and classical MTD simulation, respectively.

2.2.3 Replica exchange molecular dynamics

In the replica exchange molecular dynamics[61,64] (REMD), the simulations of multiple replicas of a system are running in parallel. The replicas differ only in temperature and in given time periods the Monte Carlo calculation is used to attempt the exchanges of their temperatures. The probability of the exchange success between replicas i and j is related to the factor ΔE_{ij} that is a function of potential energies of the replicas:[64]

$$\Delta E_{ij} = \left(\frac{U(x_i)}{RT_j} + \frac{U(x_j)}{RT_i} - \frac{U(x_j)}{RT_j} - \frac{U(x_i)}{RT_i}\right)$$
(17)

where $U(x_n)$ is the potential energy of the system carrying coordinates of the n-th replica, T_n thermodynamic temperature of the n-th replica and R the gas constant. The factor ΔE_{ij} can be used to calculate the probability P_{ij} of exchange success between the replicas i and j as follows:

$$P_{ij} = \min\{1, \exp(-\Delta E_{ij})\}$$
(18)

The closer are potential energies of the replicas, the more likely is the exchange. The probability is then treated by the Monte Carlo calculation, which decides whether the exchange occurs. The decision is performed at regular given time periods and usually between all pairs of adjacent (in sense of temperature) replicas. The high temperatures

allow the replicas to bypass the enthalpic barriers. On the other hand, the sampling ensemble relevant to chosen temperature can be found at the replica simulation following such reference temperature.

The temperature is not the only parameter, which can be used for REMD that is why above described method is more precisely called temperature-REMD (T-REMD). The others commonly used parameters are generalized Hamiltonian (H-REMD), solution pH (pH-REMD), etc.

Large systems comprise also large number of solvent molecules and number of degrees of freedom of the system is high. Thus such systems require high number of replicas to guarantee sampling on sufficiently wide temperature range and ensuring reasonable exchange rates between adjacent replicas. To overcome this issue, the methods effectively changing the solute temperature only were developed.[65,66] It was shown that among these methods, one of the best performing is REST2 method (replica exchange with solute tempering version 2).[65]

The REST2 method is based on H-REMD, in which all the replicas are running at the same temperature T_0 and the potential energy of all solute-solute and solute-solvent interactions are scaled by β_i/β_0 and $\sqrt{\beta_i/\beta_0}$, respectively (where $\beta_i = 1/(RT_i)$). This is equivalent to changing temperature of the solute from T_0 to T_i. The potential energy $U_i^{REST2}(x)$ of replica i at given coordinates x at given parameter β_i than equals:

$$U_i^{REST2}(x) = \frac{\beta_i}{\beta_0} U_{ss}(x) + \sqrt{\frac{\beta_i}{\beta_0}} U_{sw}(x) + U_{ww}(x) \tag{19}$$

where U_{ss} stands for potential energy of interactions within solute, U_{sw} for those between solvent and water (or generally solvent) and U_{ww} those between solvent molecules. Note that bond and angle terms are not scaled as their scaling was found not to improve the sampling. The ΔE_{ij} term in Equation 18 then equals:

$$\Delta E_{ij}(REST2) = (\beta_i - \beta_j)\left[(U_{ss}(x_j) - U_{ss}(x_i)) + \frac{\sqrt{\beta_0}}{\sqrt{\beta_i} + \sqrt{\beta_j}}(U_{sw}(x_j) - U_{sw}(x_i))\right] \tag{20}$$

where the n-th replica posses coordinates x_n and parameter β_n. Note that the exchange probability does not depend on solvent-solvent interactions.

Furthermore "heating" just a part of the solute whose conformational behaviour is in the interest was shown to improve the sampling efficiency in replica exchange with flexible tempering (REFT) method. The partial solute "heating" can be achieved in similar way as in Equation 19 if only parameters of desired part of solute are scaled.

3 Experimental part

3.1 Motivation - SAM-III riboswitch

X-ray diffraction is a powerful tool to determine a three-dimensional structure of biomacromolecules such as riboswitches at atomic-level resolution. However, this method can be applied only to a substance in form of monocrystal. Thus it cannot be used to study biomacromolecules under their natural conditions which is water solution. This may lead to artificial conformations that might not be relevant for water solutions but are instead enforced by crystal lattice. Furthermore in order to obtain high resolution crystal structure the structure of the molecule must be often modified compared to naturally occurring species.

This issue was several times overcame using MD simulations which can simulate the behaviour of the molecules under their natural conditions and thus reveal the relevancy of crystal structures for such a conditions or find a pathway between two conformationally distinct crystal structures of the same molecule.

New crystal structure of the ligand-free SAM-III riboswitch, in which two crystalographically independent READY states were resolved, reveals an new open questions concerning structural dynamics of ligand-free SAM-III riboswitch.[44] One structure was expected READY state structurally resembling ligand-bound state (ready1, Figure 8B). However, the second one possesses a different three-way junction and completely lacks ligand binding pocket (ready2, Figure 8C). Furthermore, the SAM was successfully soaked into the ready1 structure in the crystal but not in the ready2 structure. The question is whether the ready2 structure might play any role in the riboswitch mechanism. In addition, it is even not clear how stable is such ready2 state and whether it might coexist with another suggested conformations in solution or might exists only in the context of the crystal lattice due to crystal contacts. In this study we aimed to use contemporary computational methods in order to solve these questions.

The MD methods could help us to understand which of the structures are relevant in riboswitch natural conditions and also to understand their relationship. The aim of this study is thus to apply MD methods to SAM-III riboswitch crystal structures (ligand-bound - PDB ID: 3E5C, ready1, ready2, soaked1 and soaked2) in order to determine

their dynamics and stability in water solution and if possible to find the relationship and pathway between ready1 and ready2 structures.

3.2 Methods - SAM-III riboswitch

3.2.1 Classical MD simulations

All classical molecular dynamics simulations were performed using AMBER package[67] version 14 with AMBER ff99bsc0χ_{OL3} force field.[68–70] The crystal structures of the ligand bound (PDB ID: 3E5C), ready1, ready2, soaked1 and soaked2 states of the riboswitch were used as the starting geometries (Figure 8). Note that above-described ready1 and ready2 states were solved as two crystalographically independent monomers within one yet unpublished crystal structure.[44] Soaked1 and soaked2 structures correspond to yet unpublished crystal structures of the same crystal after soaking of SAM, so that soaked1 and soaked2 structures correspond to ready1 and ready2, respectively (whereas soaked2 structure did not bind the ligand). The missing hydrogen atoms were added by tLeaP tool from the AMBER package[67] to every structure, then the structures were neutralized by appropriate number of K^+ counterions (51, 58, 58, 57 and 58 for 3E5C, ready1, ready2, soaked1 and soaked2 structures, respectively) and immersed into the rectangular simulation box with 15 Å thick SPCE waters[71] layer and 150 mmol/l KCl (Joung and Cheatham parameters).[72]

Before running the MD simulation, the systems were minimized and heated as follows. First, the solute hydrogens were minimized, followed by water molecules, counterions and ligand. Then solute was frozen and solvent and counterions molecules were allowed to move and heated up to 298.16 K during 500 ps MD simulation in order to relax total density of the system. Then the nucleobases were allowed to move during five minimizations while a decreasing restrain was applied to sugar-phosphate backbone. After the relaxation the system was heated up to 298.16 K during 100 ps MD run in NVT condition and subsequently the density was finally relaxed in 100 ps run under NpT condition.

The MD simulations were performed at NVT conditions (298.16 K) using periodic boundary conditions and 2 fs time step.

The B-factors were calculated using the ptraj tool from AmberTools14.[67]

3.2.2 Metadynamics simulations

The well-tempered metadynamics[63] was used to study the ready1 to ready2 transition (and vice versa). In particular the G63 and G66 bases exchange in pair with A32 was biased and associated structural changes were monitored. All the WT-MTD simulations were performed using version 4.6.4 GROMACS software[73,74] and plumed plug-in[75] version 2.0.1. Again the AMBER ff99bsc0χ_{OL3} force field, rectangular simulation box (with SPCE water and 150 mmol/l KCl), and periodic boundary conditions were used. The simulations were performed with 2 fs timestep at NVT conditions (298 K). The starting geometries were prepared from the equilibrated structures taken from the classical MD simulation (so the structures were not solvated again but the AMBER coordinates and topology were converted to the GROMACS format).

In both simulations, the following two collective variables were used to describe the G63 G66 exchange: CV(I) was the linear combination of the switching functions whereas the switching function describes H bond:

$$s_{ij} = \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^6}{1 - \left(\frac{r_{ij}}{r_0}\right)^{12}} \tag{21}$$

Where r_{ij} stands for actual distance between atoms i and j (between the hydrogen and the proton acceptor atom in our case) and r_0 for equilibrium distance between the atoms i and j whereas the value of the equilibrium distance was set to 2.5 Å for our purpose.

The linear combination is designed so that in case of A32‑●‑G63 pair formation the value of this CV is negative (specifically close to -2 as there are two hydrogen bonds in the pair) and in case of A32‑●‑G66 pair the combination the value is positive (approaching value of +2).

And CV(II) was the difference between A32-G63 and A32-G66 centre of mass distances (Figure 12). For the CV(II) the UPPER_WALLS and LOWER_WALLS energetic barriers for values above 20 Å and bellow -20 Å were used as we are not interested in a conformation when either G63 either G66 are too far away from the stem. Notable sampling of too high or too low values of such collective variable might result in spurious unfolding of entire riboswitch.

The bias potential was calculated according to WT-MTD scheme (Equations 13 and 16) whereas the Gaussian potential initial height ω, timestep τ_G and bias factor $(T + \Delta T)/T$ were set to 0.5 kJ/mol, 1 ps and 15, respectively and the widths δs for both CV(I) and CV(II) to 0.2.

Moreover the restraint was applied to G31=C67 and U33–A62 base pairs (using the switching function as described in Equation 21 on hydrogen-proton acceptor distance in every H-bond of the pairs, and the LOWER_WALLS energetic barrier for values of the every switching function lower than 0.5) to prevent partial unfolding of P2 and P3 stems upon G63 to G66 exchanging.

The final Free Energy Surfaces were calculated using sum_hills utility from plumed plug-in.[75]

3.2.3   Replica exchange MD with flexible tempering

The replica exchange MD with flexible tempering of solute was used to study the three-way junction dynamics. The H-REMD implemented in the AMBER package[67] version 14 was used to perform the simulations. The REST2 scheme[65] was used to scale the potential energy of the system in order to simulate the heating of desired part of the solute.

The ready1 and ready2 structures were modified in silico, specifically the tetraloops, the bulge and the P3 stem were removed (Figure 13) and these two minimalistic model structures were used for the simulations. Again the AMBER ff99bsc0χ_{OL3} force field, rectangular simulation box (with SPCE water and 150 mmol/l KCl) and periodic boundary conditions were used. The simulations were performed under NVT conditions (at 298.16 K) with 2 fs timestep. The 8 replicas were used for each system whereas only the desired part of the solute was scaled (Figure 13). The β_i scaling factors were set in order to maintain the probability of exchange between adjacent replicas to be in range 0.25±0.03.
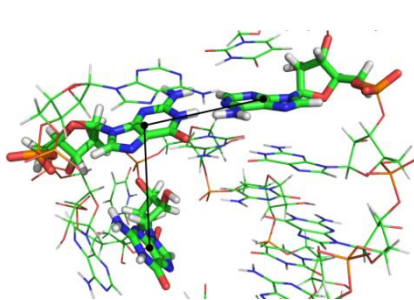
Figure 12 – The centre of mass distances used for CV(II) for WT-MTD
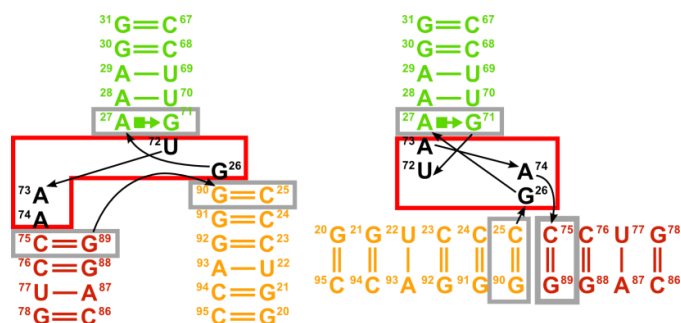
Figure 13 – The parts of the ready1 (left) and ready2 (right) structures used to explore the three-way junction dynamics by H-REMD with flexible tempering, restrained pairs in grey boxes, "hot part" in red box

It is worth noting that REST2 is not directly implemented in AMBER program, so we had to implement, test and apply the in-house program for modification of the topology file to scale the proper parts of the force field. Namely, charges of the hot region are scaled by scaling factors β_i. Van der Waals interactions are scaled, so that all interactions within hot region are scaled by scaling factors β_i, while interactions between hot and cold part are scaled by $\sqrt{\beta_i}$, and interactions within cold region are retained. As some atom types may be present in both hot and cool regions and the constants for nonbonded term are stored for pairs of atom types in the AMBER topology file, some types of nonbonded interactions may belong up to the three described groups (hot-hot, hot-cool or cool-cool). So when needed, parameters were duplicated in topology file, and defined specifically for each group of the interaction. Finally, dihedrals being fully within hot part were scaled by scaling factors β_i, while those having first atom in hot region and last atom in cold region were scaled by square root of the scaling factor $\sqrt{\beta_i}$. Note that the dihedral constants are also defined with respect to atom types, so again some dihedral types in AMBER topology file were duplicated when needed. The bonds and angles were not scaled at all, consistently with REST2 scheme.[65]

Before the REMD simulations the systems were minimized and heated as described in classical MD section.

3.3 Results - SAM-III riboswitch

3.3.1 Overall system dynamics and stability

We observed that classical MD simulations of the ligand bound (PDB ID: 3E5C), ready1 and soaked1 conformations remain stable for entire 1 μs of the simulation time-scale. Contrary, the ready2 and soaked2 structures did not fluctuate around the starting geometries and during 1 μs long classical MD simulations we observed structural changes in regions of bulge between P2 and P3 stems and in three-way junction (Figures 14, 15 and 16). Thus we prolonged the ready2 and soaked2 simulations up to 3 μs, however this prolongation did not bring qualitatively any new information.
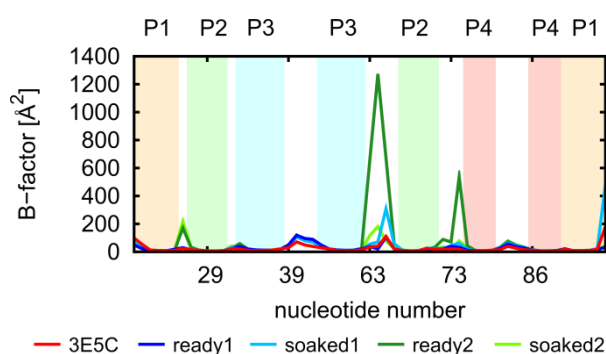
Figure 14 – B factors of particular nucleotides of examined SAM-III structures (ligand bound from 3E5C, ready1, ready2, soaked1, soaked2) during 1 μs of classical MD simulations showing that ligand bound, ready1 and soaked1 structures are stable but ready2 and soaked2 are flexible, especially the bases G26 and A74 from three-way junction and A64 from the bulge (note that the B-factor was calculated always from the trajectory rms-fitted on the starting structure along the nearest stem and if two different stems might be used for rms-fitting, we always used the lowest value of obtained B-factors, so the present B-factors correspond to local flexibility and do not count for flexibility caused by global motion). The tetraloops (on the P3 and P4 stems) and U65 nucleotide are flexible for all the structures. Bases numbering is according to genetic sequence from *Enterococcus faecalis* (consistently with previous studies).[24]

44
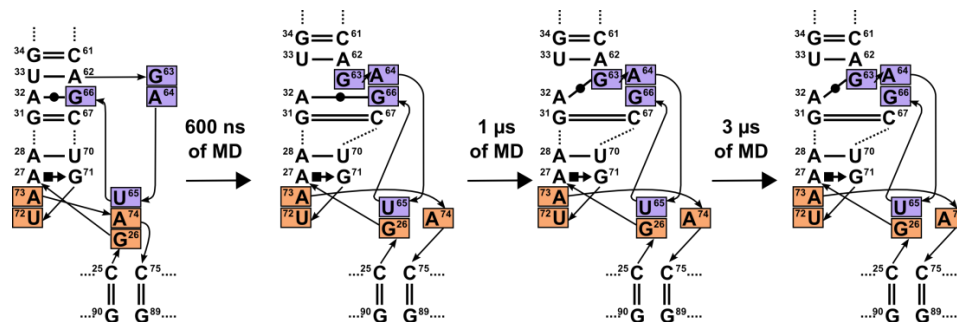
Figure 15 – Structural changes of ready2 structure during 3 μs of classical MD simulation demonstrated on secondary structure scheme. The bulge as well as the G26 and A74 from the three-way junction are flexible and structural changes in this region were observed. The G63 forms stable stacking with A62 and sometimes creates the pair with the A32 instead of the G66. Contrary the A27..A73..U72 stacking remains untouched.
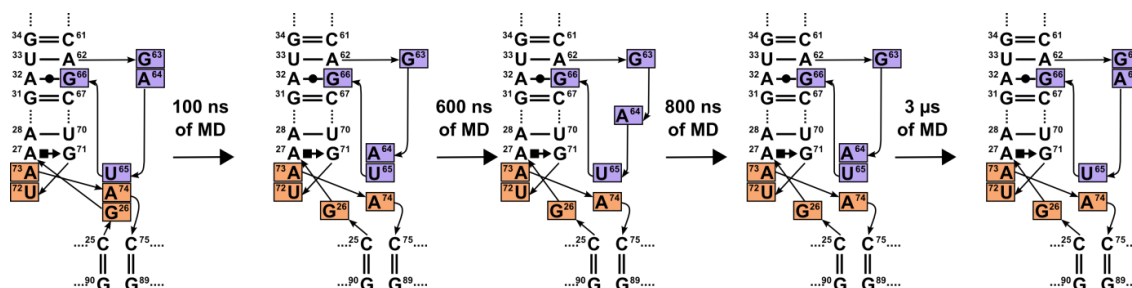

Figure 16 – Structural changes of soaked2 structure during 3 μs of classical MD simulation demonstrated on secondary structure scheme. The bulge as well as the G26 and A74 from the three-way junction are flexible and structural changes in this region were observed. Despite the simulation of the ready2 structure the G63 does not form neither stacking with A62 neither the pair with the A32. The A27..A73..U72 stacking remains untouched.

3.3.2 J3/2 bulge dynamics

As the G63 and G66 positions and interactions are likely the crucial differences between ready1 and ready2 structures and the classical MD simulations suggest that A32—●— G66 base pair is not stable, we explored the G63 vs G66 exchange (and thus the dynamics of the bulge) in both the ready1 and the ready2 structures by MTD simulations, 3 and 2 μs long in case of the ready1 and ready2 structures respectively. Our selected CVs were sufficient to sample G63 vs G66 exchange on microsecond time-scale (Figure 17), however we observed in total only few interconversions, so the simulations were not able to converge during the investigated time. Even though the simulations can provide us some qualitative information about the bulge dynamics.

The most important is that the bulge dynamics does not influence the three-way junction dynamics in our simulations. In the simulation of the ready1 structure the G71..U72 stacking remains stable despite the disruption of the G71○—□G66 and of U72-A64 pairs. In addition the C25..C75 and G89..G90 stacking is never formed. And likewise, in case of the ready2 structure the C25..C75 and G89..G90 stacking remains untouched as well as A27..A73..U72 stacking. The A27■►G71○—□G66 triplet is not formed although the free G66 tends to approach to the G71 but it is obstructed by sugar-phosphate backbone of the A73 and U72 nucleotides. This suggests that the ready1 to ready2 transition is most likely driven by the rearrangement of the three-way junction, while the G63 vs G66 exchange in pair with A32 might only stabilizes the ready1 state via liberating A64 and G66 for interaction with three-way junction in such state (see Figure 8).

From the Free Energy Surfaces obtained by MTD simulations (Figures 18 and 19) one can deduce that when the "ligand-bound-like" ("ready1-like") conformation of the three-way junction is formed, the G63 is preferred to be closer to the A32 than the G66. Contrary when "ready2-like" conformation is formed both the G63 and the G66 are preferred to by nearby the A32 and either A32—●—G63 either A32—●—G66 pair can be formed with the same probability (which is consistent with our classical MD simulations).
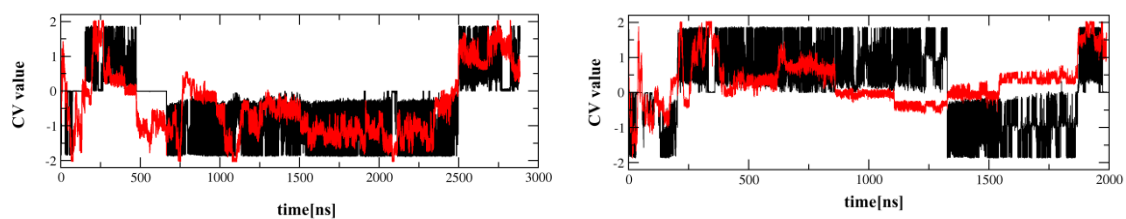
Figure 17 – Evolution of collective variables in time for ready1 (left) and ready2 (right) structure whereas three-way junction remains untouched: CV(I) (in black) = linear combination of coordination functions describing H bonds between A32 and G63 or G66 (-2 = A32-●- G63 pair formed, +2 = A32-●- G66, see Eq 21 for more detail), CV(II) (in red) = difference between A32 G66 distance and A32 G63 distance (in nm, measured between bases centres of masses, see Figure 12)
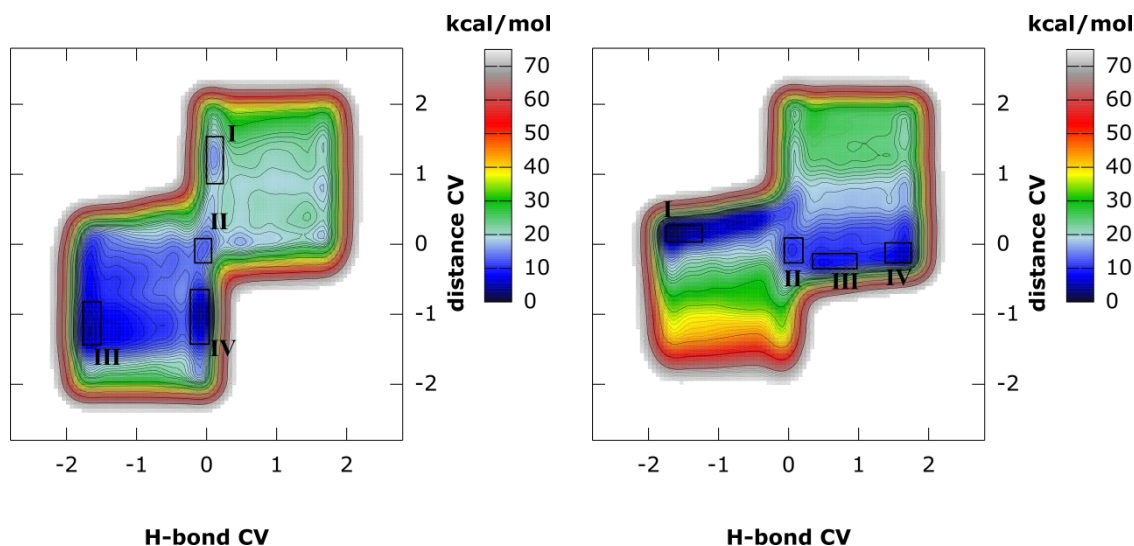

Figure 18 – Free energy surface of G63 and G66 exchange as it depends on CVs (see Figure 12 and Eq 21 for more details) for ready1 structure whereas the three-way junction remains untouched. See Figure 20 for structures I-IV corresponding to energetic minima.

Figure 19 – Free energy surface of G63 and G66 exchange as it depends on CVs (see Figure 12 and Eq 21 for more details) for ready2 structure whereas the three-way junction remains untouched. See Figure 21 for structures I-IV corresponding to energetic minima.
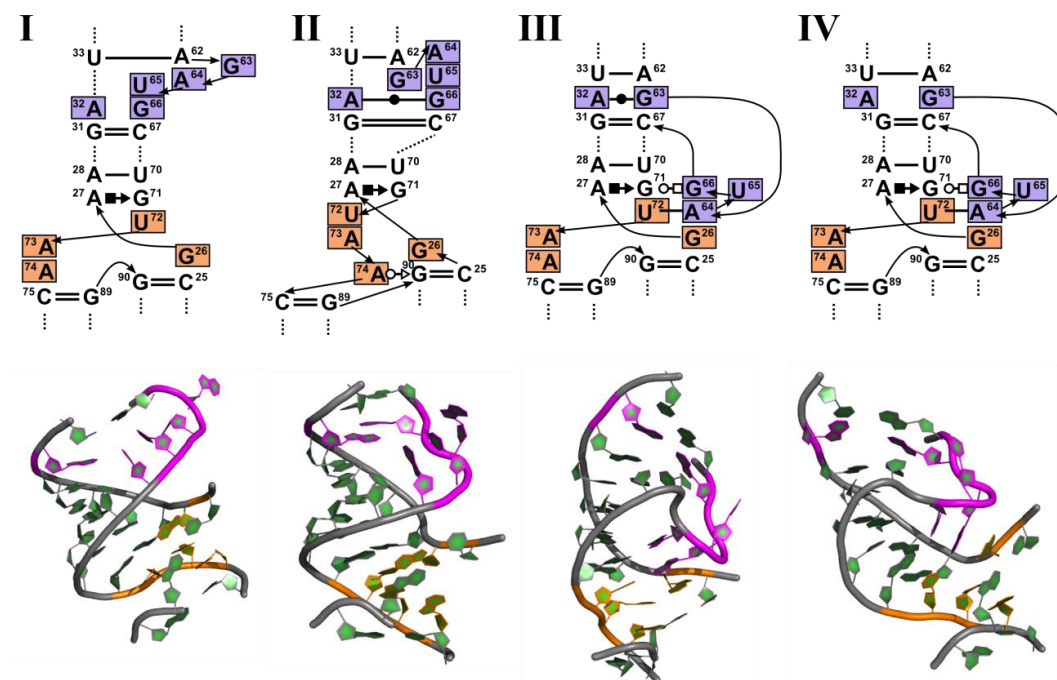
Figure 20 - The secondary structure schemes (at the top) and visualised 3D structures (at the bottom) of the most stable conformations as observed in 3 μs-long WT-MTD which started from ready1 state.
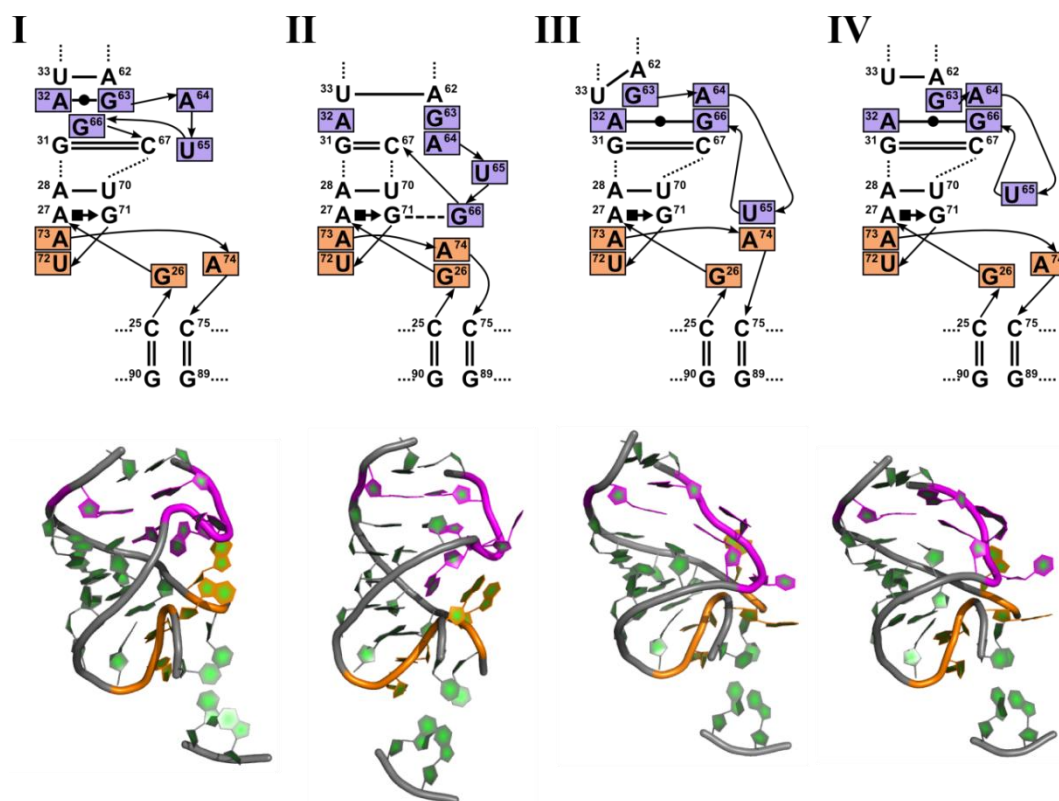
Figure 21 - The secondary structure schemes (at the top) and visualised 3D structures (at the bottom) of the most stable conformations as observed in 2 μs-long WT-MTD which started from ready2 state. Note that conformation II has tendency to form the G66▭O-G71 pair but the G66 is stopped by sugar-phosphate backbone between A73 and A74.

3.3.3 Three-way junction dynamics

As described above, we suggested that the three-way junction conformation is likely the crucial difference between ready1 and ready2 structures and it is not significantly influenced by the bulge dynamics. We thus decided to explore its dynamics by the REMD (REST2) with flexible tempering of the solute (Figure 13). Although the two 1 μs long simulations with 8 replicas did not converge and they did not reach ready1 to ready2 transition path, they provide some partial information about ready1 and ready2 three-way junction stability.

In the REST2 simulation starting from the ready1 structure, two main conformations were observed in unbiased reference replica (Figure 22). None of them represents significant change of neither three-way junction neither P1 and P4 orientation and moreover they are very similar to each other. This suggests that the three-way junction conformation of ready1 structure is stable. The configuration 2 is formed in one replica only (Figure 26) and in the "reference" (unbiased Hamiltonian) replica it is present for 2% of simulation time (Figure 24).

In the REST2 simulation starting from the ready2 structure, two main conformations were found as well (Figure 23). The conformation 1 is very similar to initial conformation but the conformation 2 presents significant structural changes. The conformation 2 is formed in one coordinate following replica only (Figure 27) but it is present in reference (unbiased Hamiltonian) replica for 32% of time (Figure 25). This suggests that such a conformation is at least comparably stable as the conformation seen in the ready2 crystal structure.
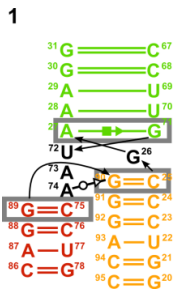
Figure 22 – The most stable conformations found by REMD with flexible solute tempering when starting from ready1 conformation. The changes between conformations 1 and 2 are rather gentle.
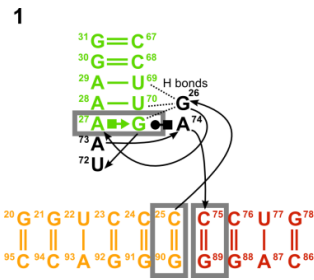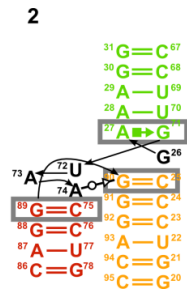
Figure 23 - The most stable conformations found by REMD with flexible solute tempering when starting from ready2 conformation. The conformation 2 contains A27..U72..A73 stacking unlike A27..A73..U72 in conformation 1 and ready2 crystal structure.
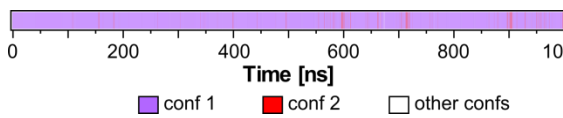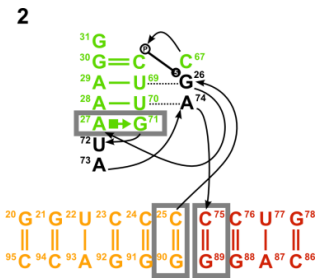
Figure 24 – The states propagation in reference unbiased Hamiltonian replica for simulation which started from ready1 configuration
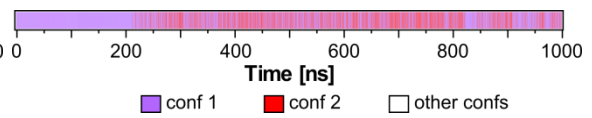
Figure 25 – The states propagation in reference unbiased Hamiltonian replica for simulation which started from ready2 configuration
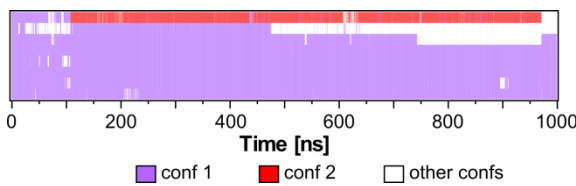
Figure 26 – The states propagation in all "coordinate following" replicas for simulation which started from ready1 configuration

Figure 27 – The states propagation in all "coordinate following" replicas for simulation which started from ready2 configuration

3.4 Motivation - preQ1-I riboswitch

The crystal structures of preQ1-I riboswitch from *T. tengcongensis*[50] in ligand-bound and ligand-free states revealed that in both the states riboswitch is folded in compact structure and first two nucleotides of the SD are sequestered within P2 stem (Figure 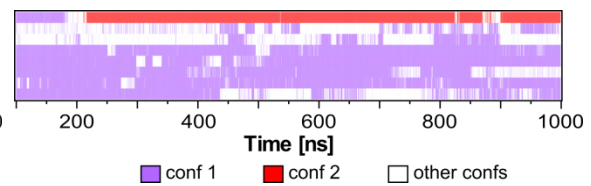11B and C). This was not consistent with experimental observations that in ligand free state in the solution the SD is accessible for binding the anti-SD sequence from 16S ribosomal subunit.[76,77] The hypothesis that L2 loop reorganisation and associated lack of A13..A32 stacking interaction would destabilize the P2 stem in water solution and the observed conformation is the artefact of the crystal packing was confirmed by a MD study.[21] However, the study was performed on the crystal structures where the RNA sequence was cut at position of G33 at 3'-end compared to native sequence in order to determine high resolution crystal structure.

3'-end overhang which is present in riboswitch native sequence is known to stabilize the A-RNA duplex. The MD simulations of structures carrying the overhang would thus provide more relevant information about the P2 stem dynamics. This study thus aims to simulate the dynamics of P2 stem within aptamer carrying the G34 at 3'-end or carrying entire SD sequence, i.e., G34 to U39. In particular, we aim to compare conformational dynamics of these extended models of preQ riboswitch in context of ligand-bound and ligand-free crystal structures and if possible to calculate the energetics of the P2 formation in context of these structures.
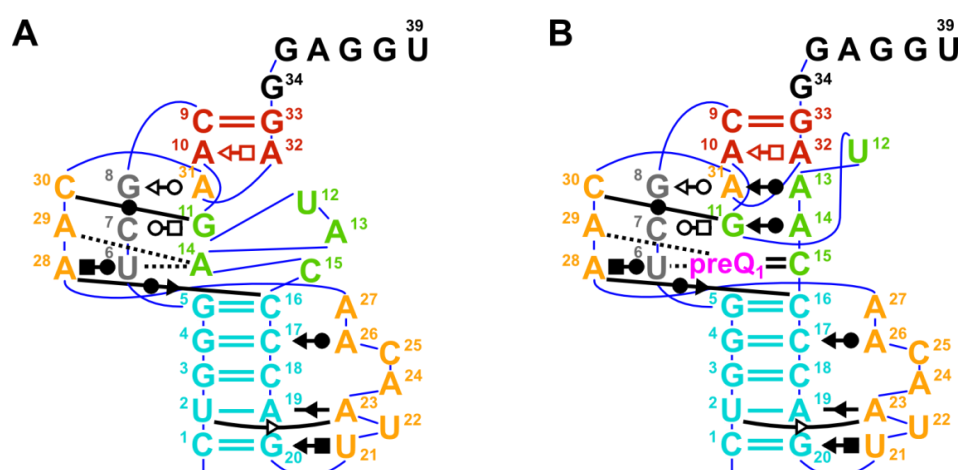
Figure 28 - Secondary structure schemes of conformations of preQ-I riboswitch aptamer from *Thermoanaerobacter tengcongensis* in apo and holo forms (**A**, **B**, PDB IDs: 3Q51, 3Q50, respectively) used as starting geometries for classical MD runs. Either the G34 or the G34 to U39 sequence (in black) were modelled in silico.

3.5 Methods - preQ1-I riboswitch

3.5.1 Classical MD simulations

As in case of the SAM-III system all classical MD simulations were performed using AMBER package[67] version 14 with AMBER ff99bsc0χ_{OL3} force field.[68–70] As the starting geometries the crystal structures of the ligand-free (apo) and ligand-bound (holo) structures of preQ-I riboswitch aptamer from *Thermoanaerobacter tengcongensis* were used (PDB IDs: 3Q51 and 3Q50, respectively) whereas the 3'-overhangs comprising either G34 (referred to as 3G), either G34 to U39 (referred to as SD) were modelled in silico using PyMOL software[78] (Figure 28). Thus in total, four systems were simulated both by classical MD and WT-MTD simulations: apo3G, apoSD, holo3G, and holoSD. In the 3Q51 structure the A13 is not resolved therefore it was modelled in silico as well. The missing hydrogen atoms were added by tLeaP tool from the AMBER package[67] to every structure, then the structures were neutralized by appropriate number of K^+ counterions (32, 37, 31 and 36 for apo3G, apoSD, holo3G and holoSD structures, respectively) and immersed into the rectangular simulation box with 15 Å thick SPCE waters[71] layer and 150 mmol/l KCl (Joung and Cheatham parameters).[72]

Before running the MD simulation, the systems were minimized and heated following the same scheme as for classical MD of SAM-III system. All MD simulations were performed under NVT conditions (298.16 K) using periodic boundary conditions and 2 fs time step.

3.5.2 Metadynamics simulations

In case of preQ1-I system the well-tempered metadynamics[63] was used to study the P2 stem dynamics, in particular the C9=G33 and A10-A32 pairs disruption and the restoration in context of apo3G, apoSD, holo3G and holoSD systems. All the WT-MTD simulations were performed using version 4.6.4 GROMACS software[73,74] and plumed plug-in[75] version 2.0.1. Again the AMBER ff99bsc0χ_{OL3} force field, rectangular simulation box (with SPCE water and 150 mmol/l KCl) and periodic boundary conditions were used. The simulations were performed with 2 fs timestep under NVT conditions (298 K). As the starting geometries, the equilibrated structures taken from the classical MD simulations were used (so the structures were not solvated again but the AMBER coordinates and topology were converted to the GROMACS format).

In all the four simulations, the following two collective variables were used: As CV(I) the sum of switching functions (Equation 21) was used whereas each switching function describes one H-bond of the P2 stem and the equilibrium hydrogen-proton acceptor distance is set to 2.5 Å as in case of MTD of SAM-III system. The sum is designed so that when all the H-bonds of the P2 stem are formed the CV(I) value equals 5, and diminishes to zero when no H-bond is present.

As CV(II) the sum of C9-G33 and A10-A32 centre of mass distances was used. Moreover the restraint was applied to the G11pU12 platform and G11-C30 and C7-G11 base pairs in case of the apo structures, and to A13-A31 and G8-A31 base pairs in case of the holo structures (using the switching function - Equation 21 - on hydrogen-proton acceptor distance in every H-bond of the pairs and H22(G11)...O4(U12) H-bond of the GpU platform, and the LOWER_WALLS energetic barrier for values of the every switching function lower than 0.5) to prevent partial unfolding of the ligand binding pocket involving the L1 and L2 loops and part of the L3 loop.

As in case of MTD of the previous system, the bias potential was calculated according to WT-MTD scheme (Equations 13 and 16) whereas the Gaussian potential initial height ω, timestep τ_G and bias factor $(T + \Delta T)/T$ were set to 0.5 kJ/mol, 1 ps and 15, respectively and the widths δs for both CV(I) and CV(II) to 0.2.

The final Free Energy Surfaces were calculated using sum_hills utility from plumed plug-in.[75]

3.6 Results - preQ1-I riboswitch

3.6.1 Overall stability and P2 dynamics

The classical MD simulations of all the four structures (apo3G, apoSD, holo3G, holoSD) remain stable for 3 µs of entire simulation timescale with only gentle local flexibilities (Figure 29). The local movements occur only in case of bases whose positions are not stabilized by base pairing, namely the U21 and A24 from loop L3 in case of all four simulations, the bulge nucleotide U12 from L2 loop in the simulations of holo form and the bulge U12 and A13 from L2 loop in the simulations of apo form as the A13 lacks the base pairing in absence of the ligand. In case of the holo3G structure, the nucleotide A23 is highly flexible alongside with U22 and A24 as the disruption of the A23-A19 base pair was observed.
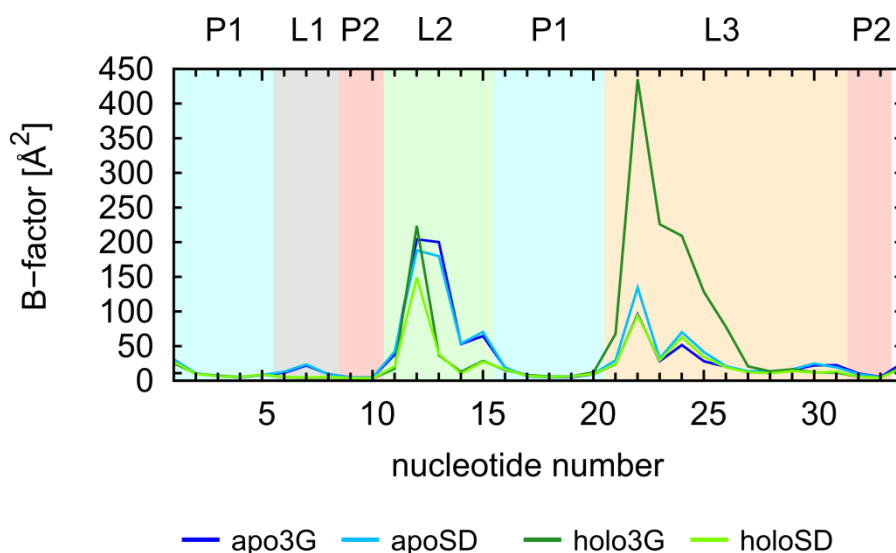
Figure 29 –Relatively low B-factors of particular nucleotides of the preQ1-I riboswitch aptamer structures show that all four structures are more or less rigid and stably fluctuates around the starting configuration during entire 3µs-long classical MD simulations (note that the B-factor was calculated always from the trajectory rms-fitted on the starting structure along the appropriate stem or loop as described above in SAM-III riboswitch section, and we always used the lowest value of obtained B-factors, so the present B-factors correspond to local flexibility and do not count for flexibility caused by global motions). Only the bases whose positions are not stabilized by base pairing are flexible (comprising U12, U22 and A24 in case of holo form and U12, A13, U22 and A24 in case of apo forms).

As the P2 stem dynamics might be crucial for the riboswitch action, it was analyzed in details with its structural context (Figure 30). Consistently with the previous classical MD study[21] in the holo structures (Figure 30C and D) the stem involving A10-A32 base pair as well as adjacent A13-A31 pair and the A13..A32 stacking interaction are highly stable. The stacking interaction may contribute to A10-A32 pair stability as was suggested in the previous study. Even more, we observed considerably smaller fluctuations of the pairs and of the stacking interaction than in the previous study (which lacks the G34 in the simulations).This suggests that P2 stem might be stabilized by G34, which stacks at the top of the P2 stem.

In contrast to the previous study,[21] despite the same time-scale of the simulations, we never observed P2 stem disruption in simulations of the apo forms (Figure 30A and B). The A10-A32 pair stability does not seem to correlate with the formation of G11pU12 platform, which in contrast to A10-A32 base pair seems not to be fully stable. Moreover, we note that the GpU platform was only partially formed comprising only the H22(G11)...O4(U12) interaction whereas the HO2′(G11)…O2P(U12) hydrogen bond was only transiently and rarely formed (Figures 31A and B and 32A and B). Similarly even the stacking interaction between U12 and A32 bases seems not to correlate with A13-A32 pair dynamics. This may again suggest that P2 stem is stabilized by stacking interaction of the G34 and under such condition its dynamics may occur on longer than microsecond timescale.
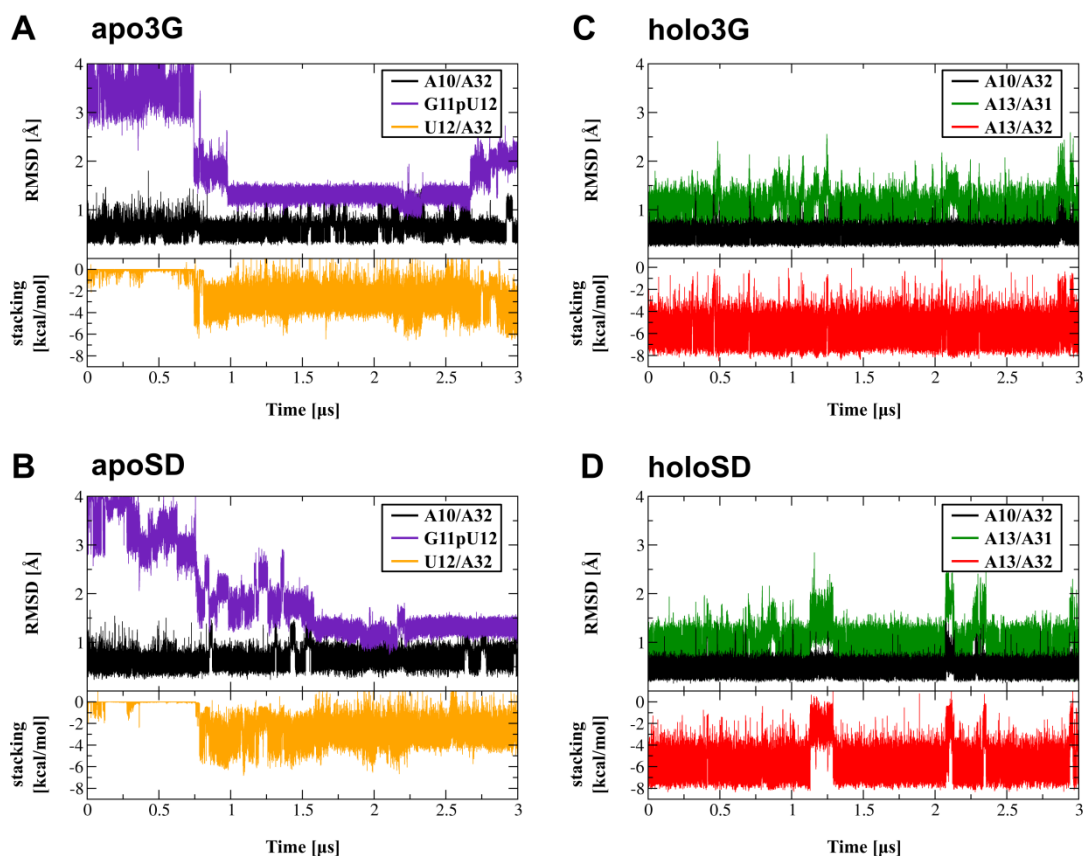
Figure 30 - Time development of RMSD of A10/A32 base pair (in black) of all the four simulations and either the G11pU12 platform (in violet) of the apo structures or the A13/A31 base pair (in green) of the holo structures. The RMSD of GpU platform was calculated with respect to G2655 and U2656 nucleotides of the sarcin-ricin loop (PDB ID: 3DVZ). The lower graphs show the development of stacking interaction energy between U12 and A32 bases (in orange) for the apo structures or between A13 and A32 bases (in red).
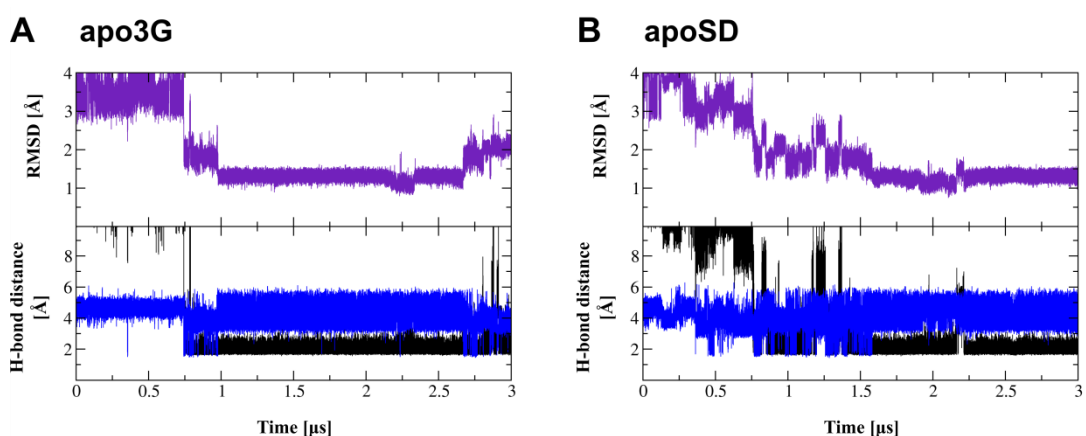
Figure 31 - The RMSD fluctuations of G11pU12 (calculated with respect to G2655 and U2656 nucleotides of the sarcin-ricin loop, PDB ID: 3DVZ) compared with dynamics of the crucial H bond interactions comprising H22(G11)...O4(U12) (in black) and HO2′(G11)…O2P(U12) (in blue).
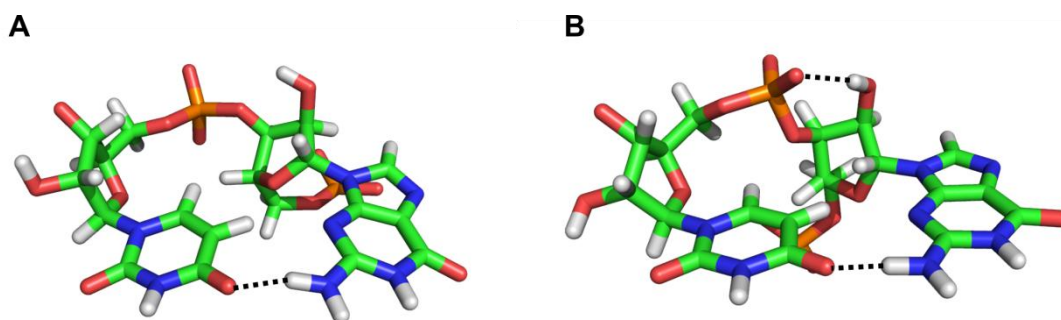
Figure 32 - The GpU platform as observed in our simulations of the apo structures (**A**) compared with the platform from the sarcin-ricin loop (**B**), PDB ID: 3DVZ. Note the difference between intramolecular interaction of 2'-OH of guanine with non-bridging oxygen of the phosphate moiety.

3.6.2 P2 dynamics in metadynamics simulations

As the P2 stem remains stable in all the four classical MD simulations, we attempted to study its disruption and restoration by WT-MTD simulations.

We observed that the chosen CVs were not able to sufficiently sample disruption and restoration of the P2 stem base-pairing despite microsecond time-scale of our MTD simulations (ranging from 1.5 to 2.5 μs; see Figure 33). In particular, the CV(I) (H bonds coordinations) was likely responsible for inability to enforce the restoration of the P2 base pairing. As mentioned above, despite the microsecond time-scale, we found that using these particular CVs the sampling of the formation and restoration of the P2 stem did not converge in our simulations. Moreover, the computed Free Energy Surfaces (Figure 35) are rather counterintuitive as they indicate that the open form of the P2 stem is more stable in case of the holo forms than in apo form. However, it is worth noting that only formation of the proper base-pairing was likely insufficiently sampled, while the sampling of the separation of two stands forming P2 stem and their approaching back to each other controlled by CV(II) was sufficient. This was supported by the fact that in the simulations of the apo forms, the G8-A31 base pair was disrupted in large CV(II) distance but was always restored when the CV(II) decreased below 1 nm (Figures 33 and 34).
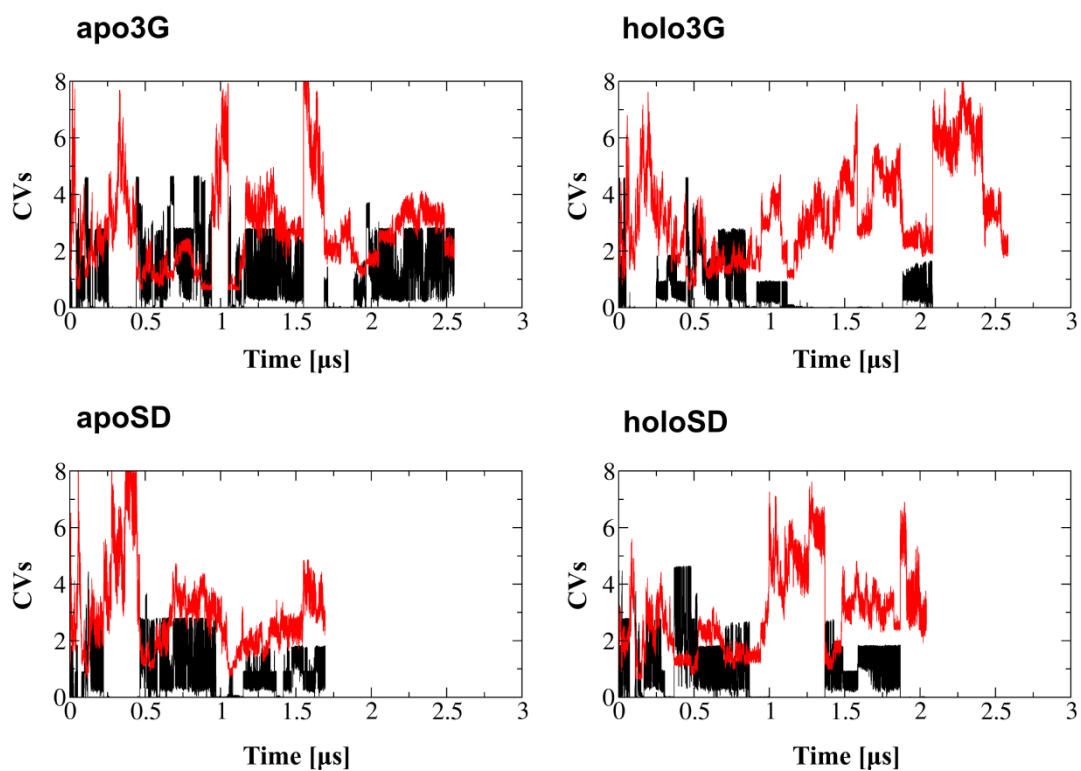
Figure 33 - The development of the CVs (CVs(I and II) in black and red, respectively) in MTD simulations of apo3G, apoSD, holo3G and holoSD structures.
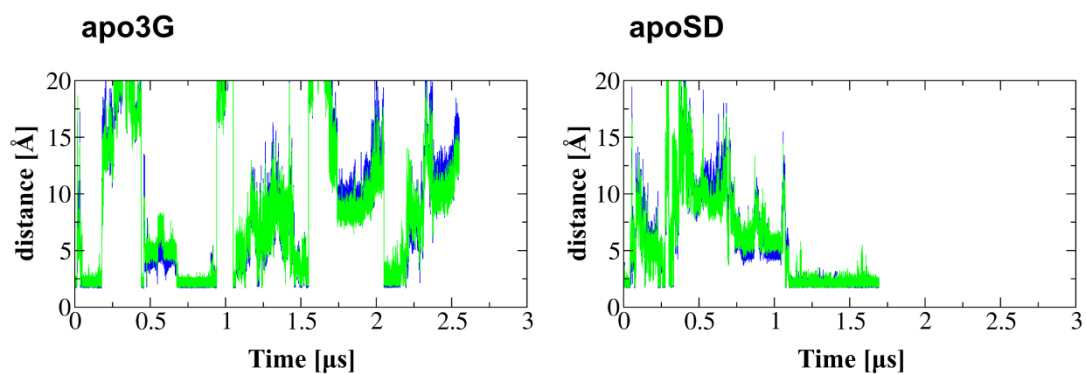
Figure 34 - The development of the distances of H bonds of G8-A31 pair (H22(G8)...N1(A31) in blue, N3(G8)...H61(A31) in green) in simulations of apo forms. Note that formation of the base pair correlates with CV(II) - see Figure 33
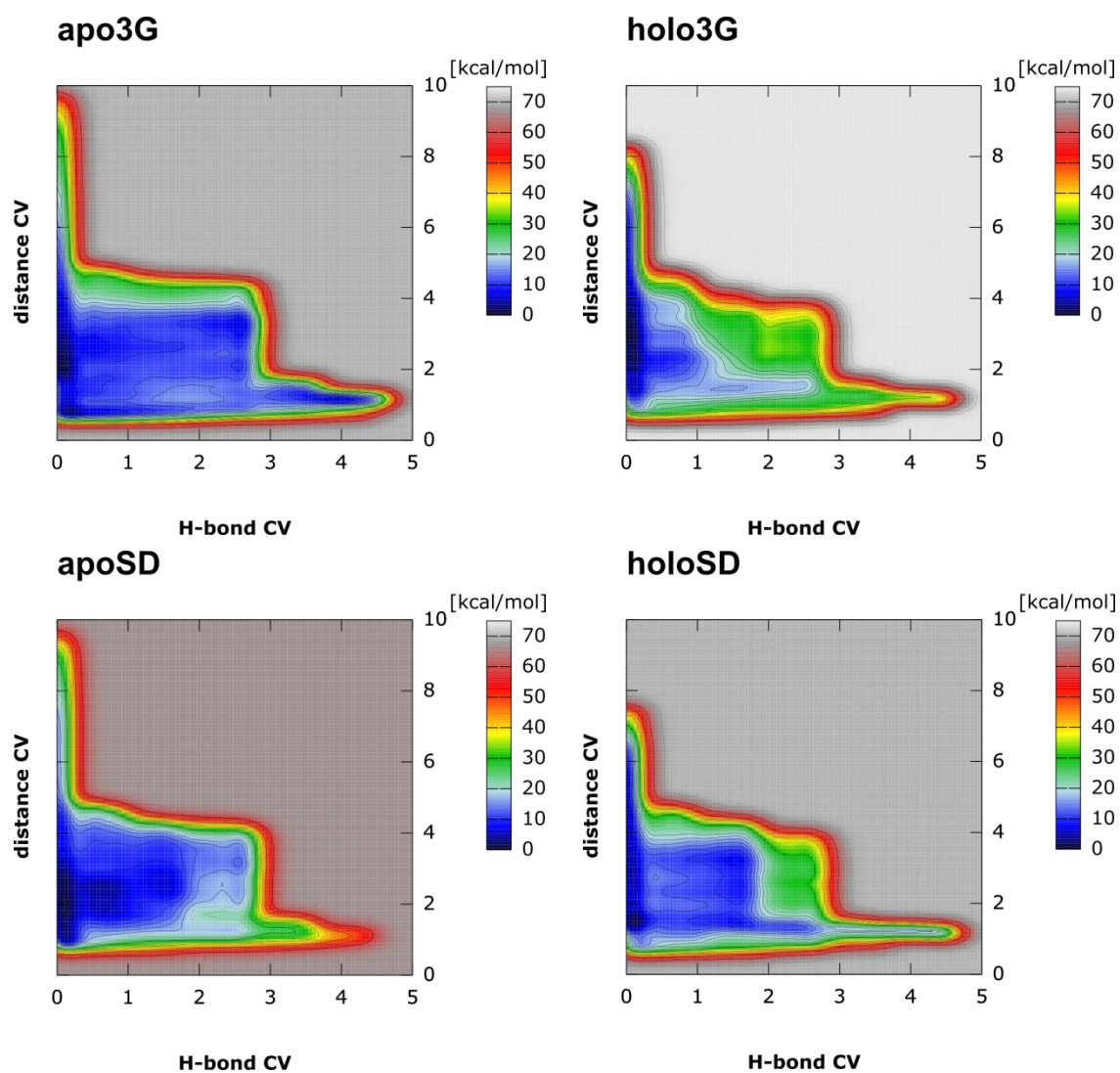
Figure 35 - The Free Energy Surfaces with respect to CVs calculated from MTD simulations of apo3G, apoSD, holo3G and holoSD structures.

3.7 Conclusions

The aim of this work was to study the conformational dynamics of riboswitches in their ligand-bound and ligand-free states using contemporary computational methods including both classical MD simulations as well as enhanced sampling techniques. Two riboswitches were examined: SAM-III riboswitch from *Enterococcus faecalis* and preQ1-I riboswitch from *Thermoanaerobacter tengcongensis*.

In case of the SAM-III riboswitch, we focused on structural dynamics of the structures in their ligand-bound and two crystalographically independent ligand-free READY states (including READY states with soaked ligand) known from crystal structures (PDB ID: 3E5C and unpublished structures).[44] It was found that in the classical MD simulations ligand-bound state riboswitch as well as the ready1 and soaked1 states (which structurally resemble conformation of ligand-bound state) keep the conformation observed in the crystal structures. Contrary, the ready2 and soaked2 states (which do not resemble structure of ligand-bound state and lack the ligand binding pocket) do not keep their crystals conformations (unpublished structures)[44] in the simulations, which indicates that these states are affected by crystal lattice.

As the bulge loop (bases G63 to G66) and the three-way junction (G26 and U72 to A74 bases) structural elements significantly differ between ready1 and ready2 crystal structures and the bulge loop conformational changes were observed in classical MD simulation, we decided to focus on the conformational dynamics of the bulge loop structural element. WT-MTD simulations revealed that even though the bulge loop adopts in simulation starting from one state the conformation of the alternative state (transition from ready1 to ready2 or vice versa) it does not influence the entire conformation of the riboswitch and in particular three-way junction arrangement.

Thus the three-way junction conformational dynamics was examined in context of minimalistic model structures (comprising P1, P2 and P4 stems only) using replica exchange MD with flexible tempering (REFT,[79] the method which scales only selected part of the solute) based on REST2 scaling scheme.[65] The in-house program implementing REST2 method in the AMBER was developed in order to scale chosen part of the solute only. Although the REFT simulations were far from being converged and they did not reach ready1 to ready2 (or vice versa) transition path, they revealed the relative stability of three-way junction arrangements of ready1 and ready2 structures.

The arrangement of ready1 evinces rather gentle changes, which indicates that such a conformation is stable. Contrary significant rearrangement occurred in case of the three-way junction element of ready2 whereas one newly found conformation was found to be likely more stable than the one observed in the crystal.

Overall, our results support the hypothesis that the ready2 conformation observed in crystal structures is an artefact caused by crystal packing whereas ligand-bound and ready1 conformations are fully relevant for riboswitch mechanism of action.

In case of the preQ1-I riboswitch we focused on the P2 stem dynamics in context of both apo and holo forms with either G34 base or complete SD sequence at 3' end of P2 stem. It was found that the P2 stem remains stable in all four structures (apo- and holo-form containing G34 only or complete SD sequence) and the whole structure fluctuates around starting coordinates on 3 microsecond time scale in classical MD simulations.

The attempt to study P2 disruption and restoration by WT-MTD was only partially successful. Our simulations were not sufficiently converged, which preclude us to evaluate thermodynamic stabilities of P2 stem in studied systems. However, the CV(II) was able to enforce the stem bases to move away and approach back. Nonetheless, the CV(I) failed to restore the base pairing, which subsequently contribute to rather limited convergence of our simulation.

Overall, these results suggest that in context of full-length mRNA transcript, the P2 stem is stabilized by 3' overhang and the P2 disruption occurs at longer than microsecond timescale which complements the previous study[21] that lacks the 3' overhang in the simulations.

4 List of abbreviations

ATP   adenosine triphosphate

bp   base pair

c-di-GMP   cyclic diguanylate

CV   collective variable

GTP   guanosine triphosphate

H-REMD   Hamiltonian-replica exchange molecular dynamics

IUB   International Union of Biochemistry

MD   molecular dynamics

MM   molecular mechanics

MTD   metadynamics

NMR   nuclear magnetic resonance

nt   nucleotide

pH-REMD   constant pH-replica exchange molecular dynamics

preQ0   pre-queuosine0 = 7-deaza-7-cyano-guanine

preQ1   pre-queuosine1 = 7-deaza-7-aminomethyl-guanine

QM   quantum mechanics

qRT-PCR   quantitative reverse transcription polymerase chain reaction

RBS   ribosome binding site

REMD   replica exchange molecular dynamics

REST   replica exchange with solute tempering

REFT   replica exchange with flexible tempering

SAH   S-adenosyl-L-homocysteine

SAM   S-adenosyl-L-methionine

SAXS   small-angle X-ray scattering

SD   Shine-Dalgarno

SHAPE   selective 2'-hydroxyl acylation analyzed by primer extension

SiM-KARTS single molecule-kinetic analysis of RNA transient structure

smFRET   single molecule fluorescence resonance energy transfer

TPP   thiamine pyrophosphate

T-REMD   temperature-replica exchange molecular dynamics

UTR   untranslated region

WT-MTD   well-tempered metadynamics

5 References

1. Garst, A. D., Edwards, A. L. & Batey, R. T. Riboswitches: Structures and mechanisms. *Cold Spring Harb. Perspect. Biol.* **3,** 1–13 (2011).

2. Wang, J. X. & Breaker, R. R. Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine. *Biochem. cell Biol.* **86,** 157–168 (2008).

3. Eichhorn, C. D., Kang, M. & Feigon, J. Structure and function of preQ1 riboswitches. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1839,** 939–950 (2014).

4. Kubodera, T., Watanabe, M., Yoshiuchi, K., Yamashita, N., Nishimura, A., Nakai, S., Gomi, K. & Hanamoto, H. Thiamine-regulated gene expression of Aspergillus oryzae thiA requires splicing of the intron containing a riboswitch-like domain in the 5′-UTR. *FEBS Lett.* **555,** 516–520 (2003).

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

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

7. Weinberg, Z., Wang, J. X., Bogue, J., Yang, J., Corbino, K., Moy, R. H. & Breaker, R. R. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol.* **11,** R31 (2010).

8. Hartig, J. S. A group i intron riboswitch. *Chem. Biol.* **17,** 920–921 (2010).

9. Winkler, W. C. An mRNA structure that controls gene expression by binding FMN. *Proc. Natl. Acad. Sci.* **99,** 15908–15913 (2002).

10. Mironov, A. S., Gusarov, I., Rafikov, R., Lopez, L. E., Shatalin, K., Kreneva, R. a., Perumov, D. a. & Nudler, E. Sensing Small Molecules by Nascent RNAA Mechanism to Control Transcription in Bacteria. *Cell* **111,** 747–756 (2002).

11. Nahvi, A., Sudarsan, N., Ebert, M. S., Zou, X., Brown, K. L. & Breaker, R. R. Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9,** 1043–1049 (2002).

12. Gold, L., Brown, D., He, Y. -y., Shtatland, T., Singer, B. S. & Wu, Y. From oligonucleotide shapes to genomic SELEX: Novel biological regulatory loops. *Proc. Natl. Acad. Sci.* **94,** 59–64 (1997).

13. Barrick, J. E. & Breaker, R. R. The distributions, mechanisms, and structures of

metabolite-binding riboswitches. *Genome Biol.* **8,** R239 (2007).

14. Yarnell, W. S. & Roberts, J. W. Mechanism of intrinsic transcription termination and antitermination. *Science (80-.).* **284,** 611–615 (1999).

15. Mandal, M. & Breaker, R. R. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.* **11,** 29–35 (2004).

16. Mandal, M., Lee, M., Barrick, J. E., Weinberg, Z., Emilsson, G. M., Ruzzo, W. L. & Breaker, R. R. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science (80-.).* **306,** 275–279 (2004).

17. Fürtig, B., Nozinovic, S., Reining, A. & Schwalbe, H. Multiple conformational states of riboswitches fine-tune gene regulation. *Curr. Opin. Struct. Biol.* **30,** 112–124 (2015).

18. Poiata, E., Meyer, M. M., Ames, T. D. & Breaker, R. R. A variant riboswitch aptamer class for S -adenosylmethionine common in marine bacteria. *RNA* **15,** 2046–2056 (2009).

19. Mellin, J. R., Tiensuu, T., Bécavin, C., Gouin, E., Johansson, J. & Cossart, P. A riboswitch-regulated antisense RNA in Listeria monocytogenes. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 13132–13137 (2013).

20. Watson, P. Y. & Fedor, M. J. The glmS riboswitch integrates signals from activating and inhibitory metabolites in vivo. *Nat. Struct. Mol. Biol.* **18,** 359–363 (2011).

21. Banáš, P., Sklenovský, P., Wedekind, J. E., Šponer, J. & Otyepka, M. Molecular Mechanism of preQ 1 Riboswitch Action: A Molecular Dynamics Study. *J. Phys. Chem. B* **116,** 12721–12734 (2012).

22. Sudarsan, N., Wickiser, J. K., Nakamura, S., Ebert, M. S. & Breaker, R. R. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev.* **17,** 2688–2697 (2003).

23. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419,** 952–956 (2002).

24. Fuchs, R., Grundy, F. & Henkin, T. The S(MK) box is a new SAM-binding RNA for translational regulation of SAM synthetase. *Nat. Struct. Mol. Biol.* **13,** 226–233 (2006).

25. Sudarsan, N., Lee, E. R., Weinberg, Z., Moy, R. H., Kim, J. N., Link, K. H. & Breaker, R. R. Riboswitches in eubacteria sense the second messenger cyclic di-

GMP. *Science* **321,** 411–413 (2008).

26. Price, I. R., Gaballa, A., Ding, F., Helmann, J. D. & Ke, A. Mn2+-Sensing Mechanisms of yybP-ykoY Orphan Riboswitches. *Mol. Cell* **57,** 1110–1123 (2015).

27. Dann, C. E., Wakeman, C. A., Sieling, C. L., Baker, S. C., Irnov, I. & Winkler, W. C. Structure and mechanism of a metal-sensing regulatory RNA. *Cell* **130,** 878–892 (2007).

28. Ren, A., Rajashankar, K. R. & Patel, D. J. Fluoride ion encapsulation by Mg2+ ions and phosphates in a fluoride riboswitch. *Nature* **486,** 85–89 (2012).

29. Haller, A., Soulière, M. F. & Micura, R. The dynamic nature of RNA as key to understanding riboswitch mechanisms. *Acc. Chem. Res.* **44,** 1339–1348 (2011).

30. Edwards, A. L. & Batey, R. T. Riboswitches: A Common RNA Regulatory Element. *Nat. Educ.* **3,** (2010).

31. McCown, P. J., Liang, J. J., Weinberg, Z. & Breaker, R. R. Structural, functional, and taxonomic diversity of three preq1 riboswitch classes. *Chem. Biol.* **21,** 880–889 (2014).

32. Trausch, J. J., Xu, Z., Edwards, A. L., Reyes, F. E., Ross, P. E., Knight, R. & Batey, R. T. Structural basis for diversity in the SAM clan of riboswitches. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 6624–6629 (2014).

33. Price, I. R., Grigg, J. C. & Ke, A. Common themes and differences in SAM recognition among SAM riboswitches. *Biochimica et Biophysica Acta* **1839,** 931–938 (2014).

34. Weinberg, Z., Regulski, E. E., Hammond, M. C., Barrick, J. E., Yao, Z., Ruzzo, W. L. & Breaker, R. R. The aptamer core of SAM-IV riboswitches mimics the ligand-binding site of SAM-I riboswitches. *RNA* **14,** 822–828 (2008).

35. Breaker, R. R. Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.* **4,** a003566 (2012).

36. Lu, C., Ding, F., Chowdhury, A., Pradhan, V., Tomsic, J., Holmes, W. M., Henkin, T. M. & Ke, A. SAM Recognition and Conformational Switching Mechanism in the Bacillus subtilis yitJ S Box/SAM-I Riboswitch. *J. Mol. Biol.* **404,** 803–818 (2010).

37. Gilbert, S. D., Rambo, R. P., Van Tyne, D. & Batey, R. T. Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nat. Struct. Mol. Biol.* **15,** 177–182 (2008).

38. Lu, C., Smith, A. M., Fuchs, R. T., Ding, F., Rajashankar, K., Henkin, T. M. & Ke, A. Crystal structures of the SAM-III/S(MK) riboswitch reveal the SAM-dependent translation inhibition mechanism. *Nat. Struct. Mol. Biol.* **15,** 1076–1083 (2008).

39. Grundy, F. J. & Henkin, T. M. The S box regulon: A new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria. *Mol. Microbiol.* **30,** 737–749 (1998).

40. Cornish-Bowden, A. Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *Nucleic Acids Res.* **13,** 3021–3030 (1985).

41. Leontis, N. B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7,** 499–512 (2001).

42. Lu, C., Smith, A. M., Ding, F., Chowdhury, A., Henkin, T. M. & Ke, A. Variable sequences outside the sam-binding core critically influence the conformational dynamics of the SAM-III/SMK box riboswitch. *J. Mol. Biol.* **409,** 786–799 (2011).

43. Wilson, R. C., Smith, A. M., Fuchs, R. T., Kleckner, I. R., Henkin, T. M. & Foster, M. P. Tuning riboswitch regulation through conformational selection. *J. Mol. Biol.* **405,** 926–938 (2011).

44. Price, I. R., Grigg, J. C. & Ke, A. unpublished data.

45. Fuchs, R. T., Grundy, F. J. & Henkin, T. M. S-adenosylmethionine directly inhibits binding of 30S ribosomal subunits to the SMK box translational riboswitch RNA. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 4876–4880 (2007).

46. Smith, A. M., Fuchs, R. T., Grundy, F. J. & Henkin, T. M. The SAM-responsive SMK box is a reversible riboswitch. *Mol. Microbiol.* **78,** 1393–1402 (2010).

47. Schuster, P., Fontana, W., Stadler, P. F. & Hofacker, I. L. From Sequences to Shapes and Back: A Case Study in RNA Secondary Structures. *Proc. R. Soc. London. Ser. B Biol. Sci.* **255,** 279–284 (1994).

48. Urbonavicius, J., Qian, Q., Durand, J. M. B., Hagervall, T. G. & Björk, G. R. Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J.* **20,** 4863–4873 (2001).

49. McCarty, R. M. & Bandarian, V. Biosynthesis of pyrrolopyrimidines. *Bioorg. Chem.* **43,** 15–25 (2012).

50. Jenkins, J. L., Krucinska, J., McCarty, R. M., Bandarian, V. & Wedekind, J. E.

Comparison of a PreQ1 riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *J. Biol. Chem.* **286,** 24626–24637 (2011).

51. Suddala, K. C., Rinaldi, A. J., Feng, J., Mustoe, A. M., Eichhorn, C. D., Liberman, J. A., Wedekind, J. E., Al-Hashimi, H. M., Brooks, C. L. & Walter, N. G. Single transcriptional and translational preQ1 riboswitches adopt similar pre-folded ensembles that follow distinct folding pathways into the same ligand-bound structure. *Nucleic Acids Res.* **41,** 10462–10475 (2013).

52. Roth, A., Winkler, W. C., Regulski, E. E., Lee, B. W. K., Lim, J., Jona, I., Barrick, J. E., Ritwik, A., Kim, J. N., Welz, R., Iwata-Reuyl, D. & Breaker, R. R. A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat. Struct. Mol. Biol.* **14,** 308–317 (2007).

53. Klein, D. J., Edwards, T. E. & Ferré-D'Amaré, A. R. Cocrystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat. Struct. Mol. Biol.* **16,** 343–344 (2009).

54. Kang, M., Eichhorn, C. D. & Feigon, J. Structural determinants for ligand capture by a class II preQ1 riboswitch. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E663–E671 (2014).

55. Liberman, J. A., Salim, M., Krucinska, J. & Wedekind, J. E. Structure of a class II preQ(1) riboswitch reveals ligand recognition by a new fold. *Nat. Chem. Biol.* **9,** 353–355 (2013).

56. Liberman, J. A., Suddala, K. C., Aytenfisu, A., Chan, D., Belashov, I. A., Salim, M., Mathews, D. H., Spitale, R. C., Walter, N. G. & Wedekind, J. E. Structural analysis of a class III preQ1 riboswitch reveals an aptamer distant from a ribosome-binding site regulated by fast dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E3485–E3494 (2015).

57. Rinaldi, A. J., Lund, P. E., Blanco, M. R. & Walter, N. G. The Shine-Dalgarno sequence of riboswitch-regulated single mRNAs show ligand-dependent accessibility bursts. *Nat. Commun.* **7,** 1–10 (2016).

58. Banáš, P., Jurečka, P., Walter, N. G., Šponer, J. & Otyepka, M. Theoretical studies of RNA catalysis: hybrid QM/MM methods and their comparison with MD and QM. *Methods* **49,** 202–216 (2009).

59. Lewars, E. G. in *Computational Chemistry* 85–173 (2011).

60. van Gunsteren, W. F. & Berendsen, H. J. C. Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. *Angew.*

Chemie Int. Ed. English **29,** 992–1023 (1990).

61. Case, D. A., Berryman, J. T., Betz, R. M., Cerutti, D. S., T.E. Cheatham, I., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Luchko, T., *et al.* Amber 2015 Reference Manual. (2015).

62. Laio, A. & Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports Prog. Phys.* **71,** 126601 (2008).

63. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100,** 1–4 (2008).

64. Nymeyer, H., Gnanakaran, S. & Garcia, A. E. Atomic Simulations of Protein Folding, Using the Replica Exchange Algorithm. *Methods Enzymol.* **383,** 119–149 (2004).

65. Wang, L., Friesner, R. A. & Berne, B. J. Replica exchange with solute scaling: A more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B* **115,** 9431–9438 (2011).

66. Liu, P., Kim, B., Friesner, R. a & Berne, B. J. Replica exchange with solute tempering: a method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 13749–13754 (2005).

67. Case, D. A., Berryman, J. T., Betz, R. M., Cerutti, D. S., T.E. Cheatham, I., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Luchko, T., *et al.* Amber14 and AmberTools14. (2014).

68. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117,** 5179–5197 (1995).

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

70. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E., Laughton, C. A. & Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids:

Improving the Description of α/γ Conformers. *Biophys. J.* **92,** 3817–3829 (2007).

71. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **91,** 6269–6271 (1987).

72. Joung, I. S. & Cheatham, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **112,** 9020–9041 (2008).

73. Hess, B., Spoel, D. van der & Lindahl, E. GROMACS User Manual version 4.6.7. *Www.Gromacs.Org* (2014).

74. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., Van Der Spoel, D., Hess, B. & Lindahl, E. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29,** 845–854 (2013).

75. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185,** 604–613 (2014).

76. Kang, M., Peterson, R. & Feigon, J. Structural Insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA. *Mol. Cell* **33,** 784–790 (2009).

77. Rieder, U., Lang, K., Kreutz, C., Polacek, N. & Micura, R. Evidence for pseudoknot formation of class I preQ1 riboswitch aptamers. *ChemBioChem* **10,** 1141–1144 (2009).

78. DeLano, W. L. The PyMOL Molecular Graphics System. *Schrödinger LLC www.pymol.org* http://www.pymol.org (2008).

79. Moors, S. L. C., Michielssens, S. & Ceulemans, A. Improved replica exchange method for native-state protein sampling. *J. Chem. Theory Comput.* **7,** 231–237 (2011).