# Structural basis for functional versatility in riboswitches

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Caroline Wetherington Reiss

Dissertation Director: Dr. Scott A. Strobel

May 2018

## Acknowledgments

It's extremely hard to believe that it's been almost six years since I began graduate school. I couldn't have made it through graduate school without the incredible scientific mentorship I received. My advisor, Scott Strobel, despite being incredibly busy, somehow always makes time to talk about project ideas, discuss experiments, or read manuscripts when I ask. Scott fosters a lab environment where good science and having a life outside the lab are both very important, which has helped keep me motivated over the last five years. Joining the Strobel lab was one of the best decisions I made in graduate school.

Thank you also to my thesis committee, Ron Breaker and Yong Xiong. Ron always has valuable advice and new ideas about riboswitches. Thank you also to Maddie Sherlock, who, with Ron, did much of the beautiful discovery work that my dissertation relied on. Yong Xiong has been invaluable as a source of X-ray crystallography knowledge and I've felt like his lab has been a second lab-family to me when I'm on science hill. Also, thank you to Jimin Wang and Michael Strickler, who make X-ray crystallography possible and accessible for me.

I'd also like to acknowledge the Strobel lab. We help each other a lot and we also laugh a lot, which makes lab both a productive and fun place to be. In particular, thanks to Dave Hiller (the only person left from when I joined the lab), Chad Torgerson (who joined just a year after me), Katie Smith (who taught me crystallography), Tanya Berbasova (who I could always work nights with), and Andrew Knappenberger (who I've worked with extensively even though he's only been here a year).

On a personal note, I would never have been able to get to or finish graduate

school without my family and friends. My mom and dad brought me up in an unbelievably fun and loving home and have always encouraged me to do what makes me happy, no matter how weird. I laugh a lot, but I have never laughed so hard as when we're on our New Hampshire family vacations together. My older sister, Eliza, was a huge role model for me throughout my whole life academically, athletically, and personally. To Eliza: I still haven't really figured out how you accomplish so much and still arrive places on time. Sorry the Myers-Briggs test says your Star Wars personality type is a Tatooine moisture farmer. My 10 month old nephew, Dylan, is extremely smiley and drooly and I want to hold him constantly. My brother-in-law, Mike, who has been part of our family for over ten years now taught me the rules for "American football". My boyfriend, Brady, was the first person I met on the first day of graduate school orientation. We won a trivia competition together that night. I love you and I'm excited to see where we go together next. My best friends from elementary and middle school, Liz, Marie, Kaleigh, Sarah, and Erin, have been extremely supportive. I'm so lucky to have a group of girlfriends that are incredibly smart, talented, and confident and help me be the same. Also, a quick acknowledgment to my cats: Bernadette (19 years old), Buzz (RIP), Linus (7), and Dudley (5) - my love of animals was one of the first things that made me interested in biology, even though I ended up veering toward the less visible side of science. Linus and Dudley keep me sane by reminding me that I have two animals to come home to at the end of lab. For my grandparents Rita (Ri), Albert (Al), and Emma (Grammy), who were loving, fierce, and persistent. Also for my grandpa, Gomp, who still sends me a card every birthday and whispers "dishwasher".

# Contents

## Abstract

Structural basis for functional versatility in riboswitches

Caroline Wetherington Reiss

2018

Since the discovery of tRNA in 1957, it has been demonstrated in the literature that RNA performs catalysis, uses cofactors, binds ligands with high affinity and stringency, and can readily evolve new functions under selective pressure. The versatility of RNA is exemplified by a class of riboswitches known as the *ykkC* RNA motif, which is a single RNA scaffold that has evolved to bind multiple, dissimilar ligands. In my dissertation, I present the structural basis of how this class of RNAs recognizes guanidine, phosphoribosyl pyrophosphate (PRPP), or guanosine tetra- and pentaphosphate ((p)ppGpp) with high affinity and specificity through changes in only a few critical residues in the ligand binding pocket. The RNA uses a combination of altering the precise location of the binding pocket and also manipulating a long-range base pair located in an S-turn motif to change base pairing in the binding pocket, all while maintaining a similar overall structure. I also present the structure of a second class of guanidine riboswitch known as the mini-*ykkC* RNA, which has a dramatically different tertiary structure from the guanidine-I riboswitch. This RNA utilizes the same types of interactions at the level of the binding pocket to recognize the guanidinium cation, but the overall structure is dramatically different. Finally, I present data suggesting that the specificity of the ZMP/ZTP riboswitch for Z nucleotides over adenosine nucleotides is partially accomplished through the use of the binding pocket magnesium ion. Together, these data showcase the versatility of RNA.

# 1 Introduction

## 1.1 Functional RNAs and the importance of structure

The "central dogma" of molecular biology, first described by Francis Crick in 1958, describes the flow of information into function [1]. DNA, which stores genetic information, is transcribed into messenger RNA, which is translated into proteins. The central dogma is inaccurate by omission, since it presents protein as the only functional polymer and makes RNA out to be simply the messenger.

Transfer RNA (tRNA) was the first functional RNA identified not in complex with a protein. It was shown to be an intermediate in protein synthesis in a set of experiments showing that, during translation in a system with free [$^{14}$C]-leucine, a soluble RNA could be isolated containing the $^{14}$C label. This labeled leucine would then transfer from the RNA to the protein being newly synthesized [2; 3]. After the discovery and sequencing of tRNA, a number of labs dedicated themselves to determining the structure of tRNA. In 1965, the cloverleaf secondary structure of tRNA was first predicted [4; 5]. In the 1970s, two labs published the crystal structure of yeast phenylalanine transfer RNA [6; 7; 8]. These structures showed for the first time the intricate three-dimensional structure of a functional RNA to atomic resolution.

These structures began the process of changing the way science thought about RNA structure and function. Unlike the comparatively plain duplex structure of DNA, RNA forms not only canonical basepairs, but also specific and complex tertiary interactions. In the case of tRNA, the secondary structure predicted through base pairing and biochemical data was correct, but this secondary structure also folded onto itself to form

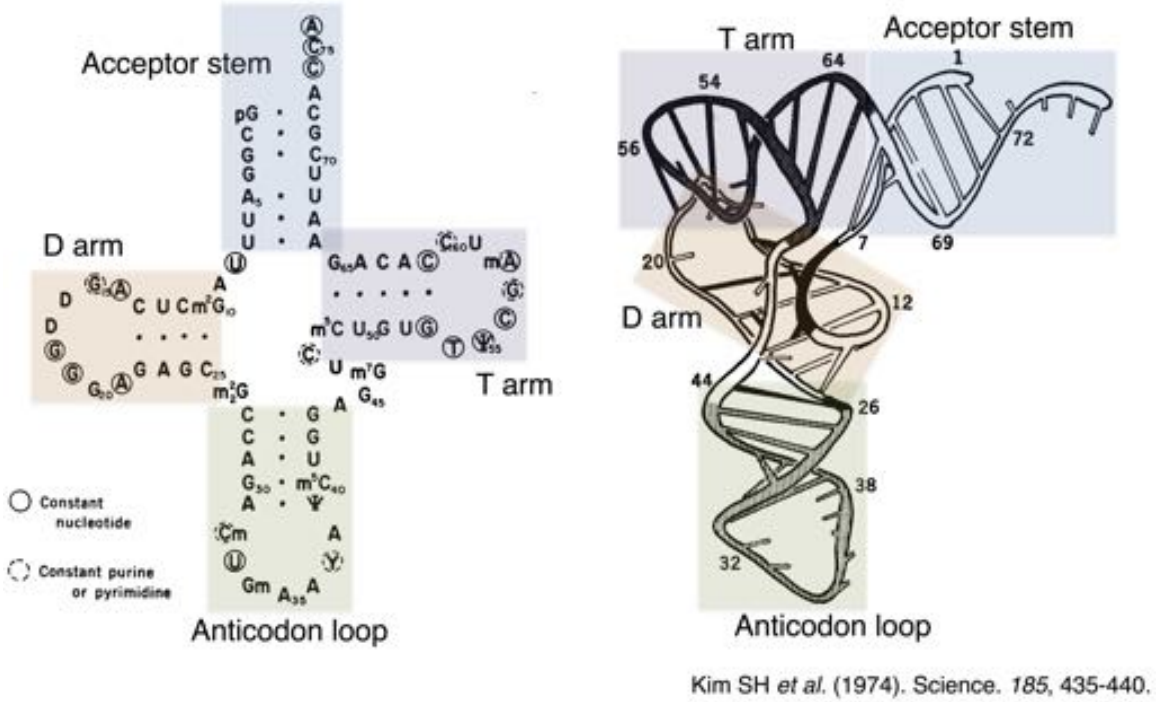


Figure 1.1: Comparison of yeast phenylalanine tRNA secondary structure to tertiary structure [7]

tertiary structure. I've displayed the secondary and tertiary structures of tRNA next to each other, as they were depicted in 1974, to illustrate the information gained from the crystal structures (Figure 1.1). Notice that the helices are arranged in a way that is still impossible to predict through bioinformatic methods. The central dogma paints RNA as simply the messenger in between code and function, but the tRNA structures (1) made it obvious that RNA could adopt a complex structure and (2) lended a structural explanation to how RNA functioned independently of protein. In this case, tRNA was acting in concert with other machinery, including the ribosome, to perform protein translation, a process central to life for all organisms.

Since the structure of tRNA was solved, numerous other functional RNAs with a huge variety of activities have been identified or created in the laboratory. In 1981, a land-

mark paper by Cech et al. demonstrated that an intervening sequence in the Tetrahymena ribosomal precursor RNA could splice itself out of the RNA and form a circular product in the complete absence of protein [9]. Just one year later, Kruger et al. from the Cech lab would used the word "ribozyme" for the first time, defining it as "an RNA molecule that has the intrinsic ability to break and form covalent bonds" [10]. Next came RNase P, which processes precursor tRNAs [11; 12]. The following years were a boom in functional RNA discovery, with current known examples ranging from small RNAs like the hammerhead ribozyme to massive ribonucleoprotein complexes with RNA active sites, like the ribosome [13].

## 1.2 RNA evolvability as a requirement of the RNA world

The discovery of functional RNAs led to the RNA world theory, which is an early-life theory positing that, before our current world of DNA, RNA, and protein, RNA performed the functions of all three [14; 15; 16; 17]. Although it is difficult to know whether RNA came before protein and DNA, we can speculate on the plausibility of this hypothesis based on what we know about RNAs that exist today. For an early RNA to propagate and evolve in a Darwinian fashion, it must have experienced a cycle of replication, mutation, and selection. Although an early RNA could not have evolved without mutation and selection, it must have experienced sufficiently robust (i.e. high fidelity) replication, or else it would face functional degeneration due to the rapid accumulation of mutations. This concept has been explored computationally in studies showing that mutation and fidelity would need to strike a balance in order for an early RNA to be successful [18; 19]. Structured

RNAs may have an advantage if they can tolerate mutation without negative functional effects, since they can explore a larger sequence space with neutral phenotypic effects, possibly allowing the RNAs to stumble upon new or improved function [18; 19; 20]. This has been referred to in the literature as "evolvability" [20; 21; 18; 19]. Computational approaches aside, the evolvability of RNA has been demonstrated many times over in the laboratory using *in vitro* evolution experiments. Ribozymes made by *in vitro* evolution include RNA-dependent RNA polymerases [22; 23] and self-replicating RNAs [24], which demonstrate the plausibility of the RNA world theory. Systematic evolution of ligands by exponential enrichment (SELEX) has been used to evolve RNA aptamers that exhibit high affinity binding to many ligands, including T4 DNA polymerase ($K_d$ as low as 12 nM), theophylline ($K_d$ = 100 nM, 10,000-fold selective over caffeine), and serotonin ($K_d$ as low as 1.2 μM) [25; 26; 27]. These RNAs can be selected from a random pool or from an existing RNA scaffold. RNAs evolved from a scaffold appear to maintain their overall tertiary structure and mutate nucleotides in the binding pocket [26; 27]. The evolvability of RNA has also been observed in nature. Much of my dissertation seeks to compare three-dimensional snapshots of structurally- or functionally-related RNAs to understand how RNA achieves evolvability at atomic-level resolution.

## 1.3 Introduction to riboswitches

Riboswitches are conserved regulatory RNAs that typically exist in the 5′ untranslated region of genes, mainly in bacteria [28; 29]. They alter the expression of gene(s) downstream in response to a small molecule ligand or ion in a concentration-dependent manner. They

have two main components: the aptamer and the expression platform. The aptamer is highly conserved and exhibits complex structure, which serves to specifically recognize the small molecule ligand [30; 31]. The ligand is usually a central metabolite, a signaling molecule, or a toxin. Examples of these include coenzyme $B_{12}$ (AdoCbl), flavin mononucleotide (FMN), the nucleobases guanine and adenine, glycine, *S*-adenosylmethionine (SAM), 5-aminoimidazole 4-carboxamide mono- or triphosphate (ZMP/ZTP), cyclic-di-GMP, the fluoride ion, and the guanidinium cation [32; 33; 31; 34; 35; 36; 37; 38; 39; 40]. The expression platform is less conserved and it propagates ligand-induced changes in the structure of the aptamer, often through helical switching, which subsequently causes a change in the expression level of the downstream gene(s). The ligand is generally related to the downstream gene(s). Riboswitches with a single aptamer and single expression platform can either be ON switches or OFF switches. In an ON switch, gene expression increases in the presence of increasing ligand concentration. For example, the fluoride riboswitch turns on the expression of genes coding for fluoride channels in response to increasing fluoride ion concentrations, which relieves the toxic build-up of the fluoride ion in bacterial cells [38; 41; 42]. In an OFF switch, gene expression is reduced in the presence of increasing ligand concentration. For example, the guanine riboswitch turns off the expression of genes coding for xanthine import and phosphoribosylation as well as *de novo* purine biosynthesis in response to increasing concentrations of guanine. Tandem riboswitches have more complex switching depending on the type of logic gate employed [43; 44; 45]. The two most common mechanisms of riboswitch regulation occur at the level of transcription and translation [31; 46; 47].

### 1.3.1 Transcriptional control by riboswitches

In a transcriptional riboswitch, the expression platform contains a terminator stem, which is a long ($\geq$ about 8bp), GC-rich helix that causes RNA polymerase to pause. A run of uridines immediately following the terminator stem destabilizes the duplex of template DNA and newly transcribed RNA, causing the paused RNA polymerase to fall off, terminating transcription [48; 49]. Intrinsic termination is sometimes facilitated by the protein factor, NusA, especially in cases with imperfect terminator stems or interrupted U-tracts [50; 51; 52; 53; 54]. Ligand binding in the riboswitch aptamer domain induces helical switching to either destabilize formation of the terminator stem (ON switch) or stabilize formation of the terminator stem to cause attenuation (OFF switch) [28; 55; 30; 46; 47]. A generic example of a transcriptional ON switch is shown in Figure 1.2A.

### 1.3.2 Translational control by riboswitches

In a translational riboswitch, the Shine-Dalgarno (SD) sequence plays an important role [33]. The Shine-Dalgarno sequence is a short, purine-rich sequence in bacteria (consensus sequence AGGAGG) that base pairs to the pyrimidine-rich sequence near the $3'$ end of 16S ribosomal RNA. The SD sequence sits approximately 8 nucleotides upstream of the AUG start codon. Through base pairing with the ribosome, the SD sequence helps recruit the ribosome for translation and plays an important role in translation intiation [56; 57]. The expression platform of a translational riboswitch can switch between two states, one where the SD sequence is accessible and one where it is sequestered [33]. In ON switches, the SD sequence is normally sequestered (default OFF). Ligand binding induces helical switching

to make the SD sequence accessible. In OFF switches, ligand-induced helical switching causes the sequestration of the previously accessible SD sequence (default ON). In this mechanism of control, riboswitches affect translation from an mRNA that has been transcribed [33; 47]. A generic example of a translational ON switch is shown in Figure 1.2B.

## 1.4 Riboswitch variants

Riboswitches are grouped into classes based on conservation within the aptamer domain. The aptamer structure acts as a conserved recognition motif for the ligand, so it is tempting to think of a single riboswitch class as corresponding to a single ligand. However, this is often not true. A riboswitch variant is a subtype of a conserved motif that utilizes the same structural scaffold to change gene expression through helical switching, but with altered ligand specificity [58].

Variant riboswitches are the result of divergent evolution. The structural elements of the aptamer remain the same, but nucleotide(s) in the binding pocket have been changed, resulting in a change in ligand specificity. Variant riboswitches do not exist upstream of the same genes that the main subtype does, rather the genes downstream are related to the new ligand. Variant riboswitches exemplify RNA evolvability, since they represent the acquisition of new function through minimal mutations. The RNA retains the same overall fold, but as a result of mutation and selection pressure, it gains a different specificity and controls different genes. The shape of the fitness landscape for a particular motif determines the number of mutations required to recognize a new ligand with high affinity and stringency, suggesting that specific motifs may be quicker to adapt than others.

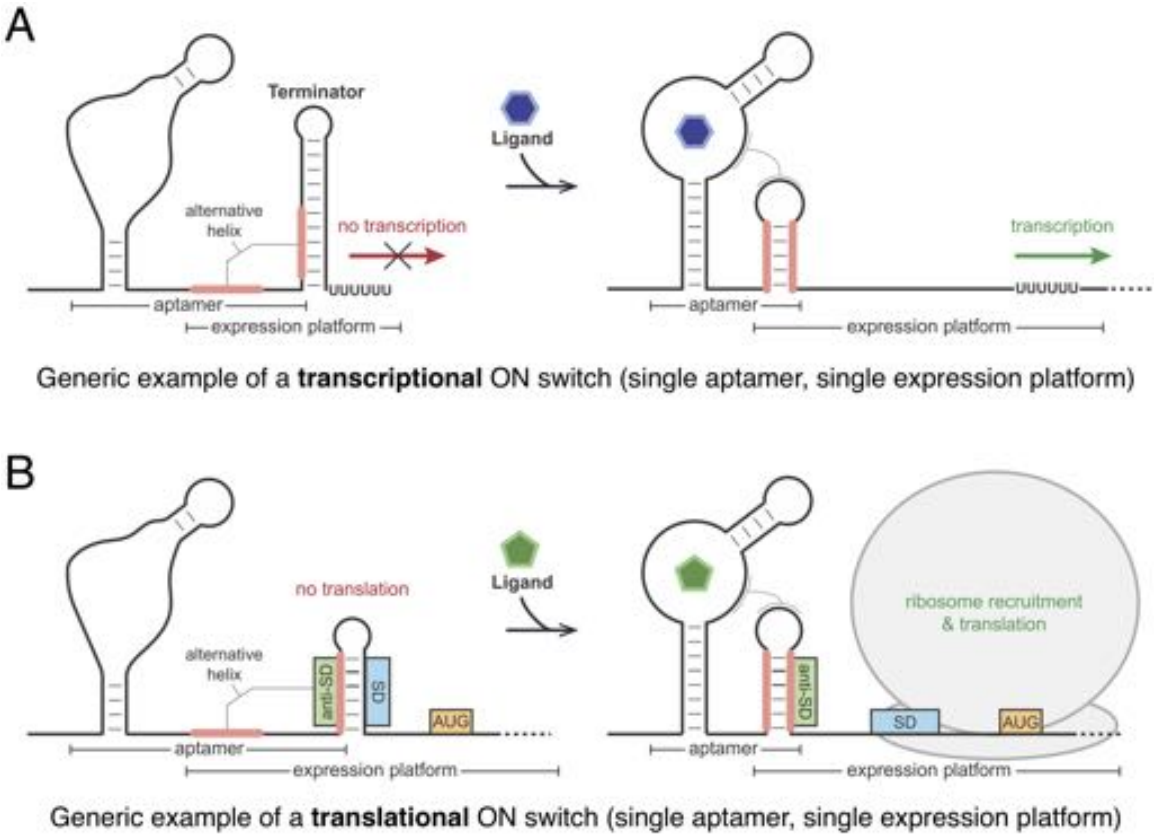


Figure 1.2: A. A generic example of an ON switch functioning at the transcriptional level, where there is a single aptamer domain and a single expression platform. Binding of the ligand (blue hexagon) to the aptamer disrupts the formation of the terminator stem by stabilizing a mutually exclusive, alternative helix (pink bars), allowing transcription to proceed. B. A generic example of an ON switch functioning at the translational level, where there is a single aptamer domain and a single expression platform. Binding of the ligand (green pentagon) to the aptamer disrupts the formation of a helix that sequesters the Shine-Dalgarno (SD) sequence by stabilizing a mutually exclusive, alternative helix (pink bars), which frees the Shine-Dalgarno sequence to participate in ribosome recruitment and initiation of translation of the downstream gene.

### 1.4.1 The purine riboswitch: the first case of a riboswitch variant class

One of the very first classes of riboswitch identified was the guanine riboswitch, published in 2003 [28]. In the genome of *Bacillus subtilis* alone, five different transcriptional riboswitches adhering to this RNA motif were found. Together, they controlled 17 genes related to purine biosynthesis and transport. Just one year later, Mandal et al. demonstrated that some examples of the guanine riboswitch class actually rejected guanine as a ligand. Instead, these examples bind adenine with high nanomolar affinity. The only difference in conservation between the guanine and adenine riboswitch sequences occurs at position 74. In guanine riboswitches, this position is a C; in adenine riboswitches, it is a U. Mandal et al. speculated that the C in guanine riboswitches formed a Watson-Crick base pair with the ligand and that a specificity switch was achieved by altering this to a U, which would base pair with adenine [34]. Future crystal structures confirmed this hypothesis [59]. Additionally, the switches had a different logic: guanine binding triggered the formation of a transcription terminator stem (OFF switch), while adenine binding excluded transcription terminator stem formation (ON switch). This is consistent with the genes downstream: guanine controls genes related to purine biosynthesis and import while adenine controls genes related to purine efflux. Presumably, one of these two riboswitches evolved first and a bacteria was able to repurpose the same scaffold to detect a different ligand in an example of divergent evolution. The ability of this RNA element to detect two different ligands with specificity using only a single mutation was the first published example of riboswitch evolvability, albeit through a simple change in Watson-Crick base pairing.

### 1.4.2 Other riboswitch variants

Since the discovery of the purine riboswitch, a large number of riboswitch variants have been identified. Another example is the cyclic-di-GMP riboswitch, which recognizes the bacteria second messenger, cyclic-di-GMP and controls a variety of genes related to the cell's lifestyle, including many implicated in virulence [37]. A subset of these riboswitches were discovered to bind cyclic-GMP-AMP and reject cyclic-di-GMP due to a G20A mutation that alters specificity at the $\alpha$ site [60; 61; 62]. In 2017, a pipeline was described by Weinberg et al. to search for these variants and they reported a number of novel variant classes [58]. They searched for riboswitch variants using the following three criteria that are characteristic of a riboswitch variant:

1. **Sequence alignments**: to locate variable positions
2. **Atomic-resolution crystal structures**: to determine which bases contact the known ligand
3. **Genes located downstream**: to identify correlations between gene context and sequence variation

This pipeline identified known examples of riboswitch variants in addition to uncovering new variant subclasses found within the purine, glycine, cyclic-di-GMP, FMN, and Ni/Co riboswitch classes. These variant subclasses all reject the ligand of the main subclass. In the case of a new purine riboswitch variant, Weinberg et al. saw conformational changes induced by $2'$-deoxyguanosine in a ligand-dependent manner. These data demonstrate that riboswitch variants are common and hint at the natural plasticity of RNA.

## 1.5 Some riboswitch classes respond to the same ligand

Riboswitch variants show that the same RNA scaffold can modestly mutate to respond to a different ligand. On the other hand, riboswitches of entirely different structures have also convergently evolved to recognize the same ligand. This phenomenon is most impressive in the *S*-adenosylmethionine (SAM) riboswitches, of which there are six published classes [63; 64; 36; 65; 66; 67; 68]. SAM is one of the most common enzyme substrates in life and is involved in a massive number of biochemical processes, including methionine and cysteine metabolism, DNA restriction, and polyamine synthesis [69]. SAM is also compelling as a possible substrate for the last universal common ancestor (LUCA), since it is ubiquitous throughout life and chemically resembles a nucleotide [69]. For these reasons, it makes sense that SAM would be recognized by multiple different, widespread riboswitches. Some SAM riboswitch structures share similarities in their conserved core (e.g. SAM-I and SAM-IV), but it does not appear that they are phylogenetically related, making the SAM clan a robust example of convergent evolution in riboswitches [70; 71; 72; 73]. SAM is not the only ligand for which multiple classes of riboswitch have evolved. Others include *S*-adenosylhomocysteine (2 classes if you include the SAM/SAH riboswitch), cyclic-di-GMP (2 classes), pre-$Q_1$ (3 classes), 2′-deoxyguanosine (2 classes), magnesium (2 classes), and guanidine (3 classes) [74]. For this to be a common occurrence, RNA must be capable of finding many unique solutions to the same problem. In many cases, these RNA structures are dramatically different in fold, gene control mechanism, and stringency of ligand recognition. These are examples of convergent evolution in riboswitches and they illustrate the flexibility of RNA, which can use many different architectures to solve the same problem.

## 1.6 Unpacking the versatility of RNA structure

In the past, sequence alignments and structural studies have been used to understand the evolutionary relationships between proteins. Divergently evolved proteins, like those that confer antibiotic resistance, were studied to determine how mutations could confer the structural changes needed to recognize and degrade new antibiotics [75; 76; 77; 78]. Structural studies of convergently evolved proteins, like the serine proteases chymotrypsin and subtilisin, uncovered that both proteins use the same catalytic triad, despite having different overall protein folds [79]. I seek to perform this same, detailed unpacking of RNA by solving X-ray crystal structures of riboswitches related through divergent evolution (variants) or convergent evolution (distinct structural classes that bind the same ligand).

# 2 Structure of the guanidine-I riboswitch aptamer

This section is adapted from Reiss et al., 2017, an article published in the journal *Structure* [80].

## 2.1 Background

The *ykkC/yxkD* RNA was first identified as a riboswitch by bioinformatics in 2004 [29]. It was dubbed an orphan riboswitch, because the ligand was unknown [29]. Ligand identification proved elusive for many years due to the wide variety of poorly studied genes under its control [81]. Many of the genes downstream were transporter proteins, suggesting that the ligand could be a toxic small molecule [40]. A precedent for this was the *crcB* RNA, which was also found upstream of uncharacterized transporter proteins [38]. It was eventually discovered that the *crcB* RNA bound the toxic anion, fluoride, which in turn revealed that many of the downstream genes encoded fluoride ion channels, including Fluc [38; 82; 41; 83; 42; 84]. In order to identify the ligand of the *ykkC* RNA, Nelson et al. screened approximately 2,000 toxic compounds and growth conditions for expression of a *lacZ* reporter under riboswitch control. The only condition that induced reporter expression contained guanidine hydrochloride. This screen and subsequent testing by additional reporter assays, in-line probing, and transcription termination determined that guanidine induced gene expression through direct binding to the *ykkC* RNA motif with micromolar affinity [40]. Nelson et al. also showed that the *ykkC* RNA was specific for guanidine over guanidino-containing compounds, including agmatine and *L*-arginine. Although guanidine is a fragment within key metabolites like arginine, guanidine had not previously been

linked to any metabolic pathway in bacteria, suggesting that there was more to learn about guanidine biochemistry.

Genes associated with this riboswitch include genes annotated as urea carboxylases, urea carboxylase-associated proteins, and a variety of transporters, including multidrug resistance efflux proteins [40]. The guanidine riboswitch exists in the 5′ untranslated region of these genes. Nelson et al. examined a gene from *Oleomonas sagaranensis* previously characterized as a urea carboxylase, yet the catalytic efficiency is 40 times better for guanidine than urea, suggesting that it is a guanidine carboxylase that has been mis-annotated as a urea carboxylase. The product of the reaction catalyzed by guanidine carboxylase is carboxyguanidine, which is in turn suspected of being the actual substrate for genes annotated as allophanate hydrolases that are controlled by the riboswitch (these would then be carboxyguanidine hydrolases). They also investigated a transporter (annotated as a SugE protein) in Clostridiales controlled by the guanidine-I riboswitch. They found that intrinsic fluorescence of tryptophan residues increases in a guanidine-dependent manner ($K_d$ = 1 mM), providing preliminary evidence that this transporter specifically recognizes guanidine [40]. Very recently, Kermani et al. demonstrated that most members of the small multidrug resistance (SMR) family of proteins are actually highly specific guanidinium/proton antiporters. They call them Gdx proteins, since they export guanidinium from the cell [85]. Nelson et al. also demonstrated that guanidine binding to the aptamer turns on gene expression in *Bacillus subtilis* and *Desulfotomaculum ruminis*. Transcription termination assays in *D. ruminis* show that upregulation of gene expression occurs in a guanidine-dependent manner via destabilization of a terminator stem [40]. This is comparable to switching mechanisms used in other riboswitches (For review see [47; 31]. This,

combined with data regarding downstream gene function, suggests that guanidine turns on expression of genes that either enzymatically act on guanidine or transport it out of the cell.

To provide evidence that the riboswitch controlling these genes is specifically recognizing guanidine rather than another guanidino-containing compound, I set out to determine the crystal structure of the guanidine-I riboswitch bound to guanidine via X-ray crystallography. Specific recognition of guanidine by the RNA poses a challenge because the ligand is a common functional group in many other metabolites. The binding pocket must exclude larger guanidino-containing compounds while still maintaining sufficient guanidine affinity. Arginine and agmatine both contain guanidino groups and are present at mid to high micromolar concentrations in bacteria [86; 87; 88; 89]. These compounds do not bind the riboswitch, suggesting they are excluded from the binding site [40]. The riboswitch also has no apparent affinity for urea, which is produced by the urea cycle in bacteria and differs from guanidine by only a single functional group, where one amine (C-NH2) in guanidinium is a carbonyl group (C=O) in urea [40].

The consensus motif for the guanidine riboswitch generated by bioinformatics provided less structural information than usual due to a particularly high degree of sequence conservation. Three helices, P1, P1b, and P2, were identified by covariation analysis. Of the remaining 51 nucleotides (out of approximately 90-100 nt in the aptamer) that do not exhibit covariation or form variable hairpins, 35 are greater than 97% conserved. The predicted secondary structure had a large, highly conserved internal loop between P1 and P1b and a long, highly conserved tail at the $3'$ end of the RNA [40]. Such a high degree of conservation suggested that the riboswitch contained additional structural elements including many sequence-specific tertiary interactions.

## 2.2 Results

### 2.2.1 Structure determination of the guanidine-I riboswitch

I targeted the aptamer domain of the guanidine-I riboswitch from *Sulfobacillus acidophilus* for crystallization. This riboswitch controls the expression of a protein annotated as urea carboxylase-associated protein 2, though the gene may actually be a guanidine carboxylase-associated protein. The crystallization construct includes 94 nucleotides from the native sequence, plus an added G on the 5′-end to aid in efficient transcription. The non-conserved P1b helix (nt 23-26 and 30-33) was mutated from the wild type sequence CUAG/CUAG to CGUC/GGUG to introduce a tandem GU wobble motif. This tandem GU motif is an ideal binding site for iridium hexamine, a metal complex with anomalous signal for use in structure determination [90]. Initial phases were generated using a three-wavelength MAD dataset on a co-crystal of the riboswitch and iridium hexamine, a SAD dataset collected on a co-crystal of the riboswitch and iridium hexamine, and a native dataset isomorphous to the SAD dataset (Table 2.1). Three iridium hexamine sites were identified and used for phasing. A model of the *S. acidophilus* guanidine riboswitch was built into the unbiased density and refined using the high remote wavelength dataset from MAD collection.

### 2.2.2 Overall architecture of the *S. acidophilus* guanidine-I riboswitch

The structure of the guanidine-I riboswitch is dominated by a long continuous set of coaxially stacked helices that includes P1, P1a, P1b, and P2. The large internal loop between P1 and P1b forms a helix, which I termed P1a, that extends the helical stack. An additional helix, P3, forms between conserved nucleotides in the extended tail at the 3′ end

Table 2.1: X-ray statistics for the guanidine-I aptamer

| | | MAD + guanidine and [Ir($NH_3$)$_6$] | | SAD + guanidine and [Ir(NH3)6] | Native + guanidine |
|---|---|---|---|---|---|
| **PDB ID** | | **5T83** | | | |
| **Data collection:** | | | | | |
| Beamline | | 8.2.2 at ALS | | 19-BM at APS | 19-BM at APS |
| Space group | | P 43 21 2 | | P 43 21 2 | P 43 21 2 |
| **Unit cell:** | | | | | |
| a, b, c (Å) | | 49.1, 49.1, 246.3 | | 48.3, 48.3, 246.7 | 48.3, 48.3, 246.6 |
| $\alpha$, $\beta$, $\gamma$ (°) | | 90, 90, 90 | | 90, 90, 90 | 90, 90, 90 |
| | **High remote** | **Inflection** | **Peak** | | |
| Wavelength (Å) | 0.999946 | 1.10532 | 1.10482 | 1.10062 | 1.10062 |
| Resolution (Å) | 40.00-2.70 (2.80-2.70) | 40.00-2.80 (2.85-2.80) | 40.00-2.80 (2.85-2.80) | 40.00-2.92 (2.97-2.92) | 40.00-2.93 (2.98-2.93) |
| $R_{merge}$ | 0.074 (0.712) | 0.093 (0.729) | 0.107 (0.784) | 0.104 (>1.000) | 0.165 (>1.000) |
| I/$\sigma$I | 35.2 (2.5) | 29.1 (1.3) | 30.1 (1.3) | 23.0 (0.9) | 18.3 (0.94) |
| $CC_{1/2}$ in highest resolution shell | 0.928 | 0.831 | 0.832 | 0.447 | 0.448 |
| CC* in highest resolution shell | 0.981 | 0.953 | 0.953 | 0.786 | 0.787 |
| Completeness (%) | 99.8% (100.0%) | 99.8% (100.0%) | 99.9% (100.0%) | 99.8% (99.8%) | 99.6% (100.0%) |
| Redundancy | 6.9 (6.7) | 6.6 (4.2) | 6.9 (4.3) | 8.0 (6.3) | 13.0 (11.9) |
| Total reflections | 107588 | 92522 | 97493 | 96712 | 91329 |
| Unique reflections | 15638 | 14071 | 14090 | 12135 | 7022 |
| **Refinement:** | | | | | |
| Resolution () | | 40.00-2.71 | | | |
| No. reflections | | 8451 | | | |
| $R_{work}/R_{free}$ | | 0.20/0.23 (0.39/0.44) | | | |
| **No. of atoms:** | | | | | |
| Total | | 1985 | | | |
| RNA | | 1917 | | | |
| Ligand | | 4 | | | |
| Cations | | 46 | | | |
| Water | | 18 | | | |
| **B-factors** | | | | | |
| Overall (Å$^2$) | | 88.9 | | | |
| RNA (Å$^2$) | | 87.7 | | | |
| Ligand (Å$^2$) | | 81.8 | | | |
| Cations (Å$^2$) | | 129 | | | |
| Water (Å$^2$) | | 116.1 | | | |
| **Root-mean-square deviations:** | | | | | |
| Bond lengths (Å) | | 0.007 | | | |
| Bond angles (°) | | 1.815 | | | |

of the sequence. Helices P1a and P3 were not previously identified by bioinformatics, because the very high level of sequence conservation and limited canonical pairing prevented conclusions based upon covariation analysis [81; 91; 92]. The P1a and P3 helices form sequence-specific tertiary interactions, consistent with this high degree of conservation. The guanidine-I riboswitch consensus motif has been updated to reflect the secondary structure observed in the crystal structure (Figure 2.1A,B). The secondary structure of the crystallized construct and a ribbon diagram of the crystal structure are shown in Figure 2.1B and C, respectively.

### 2.2.3 Guanidinium cation recognition by the riboswitch

The guanidine binding pocket is formed at an interface between the P3 and P1a helices. Electron density for the guanidinium cation can be seen in the unbiased electron density map initially generated from phasing (Figure 2.2A). After refinement, the modeled guanidine has similar B-factors to the residues surrounding it.

The nucleotides surrounding the guanidine ligand are highly conserved. Only a small tunnel is left solvent-accessible by the RNA and this is blocked by a metal ion that is visible in the crystal structure (Sr-7) (Figure 2.2D).

The ligand is exhaustively recognized by the RNA. The pKa of guanidine is 13.6 [93] and under physiological conditions is present almost exclusively as the guanidinium cation, which has three-fold symmetry and a positive charge stabilized by resonance (Figure 2.1D). The guanidinium has six hydrogen bond donors, two on each nitrogen. All potential hydrogen bonds are satisfied by the RNA. Guanidinium donates hydrogen bonds to three nucleotides, G45, G90, and G73, one nucleotide for each of its three edges. On one

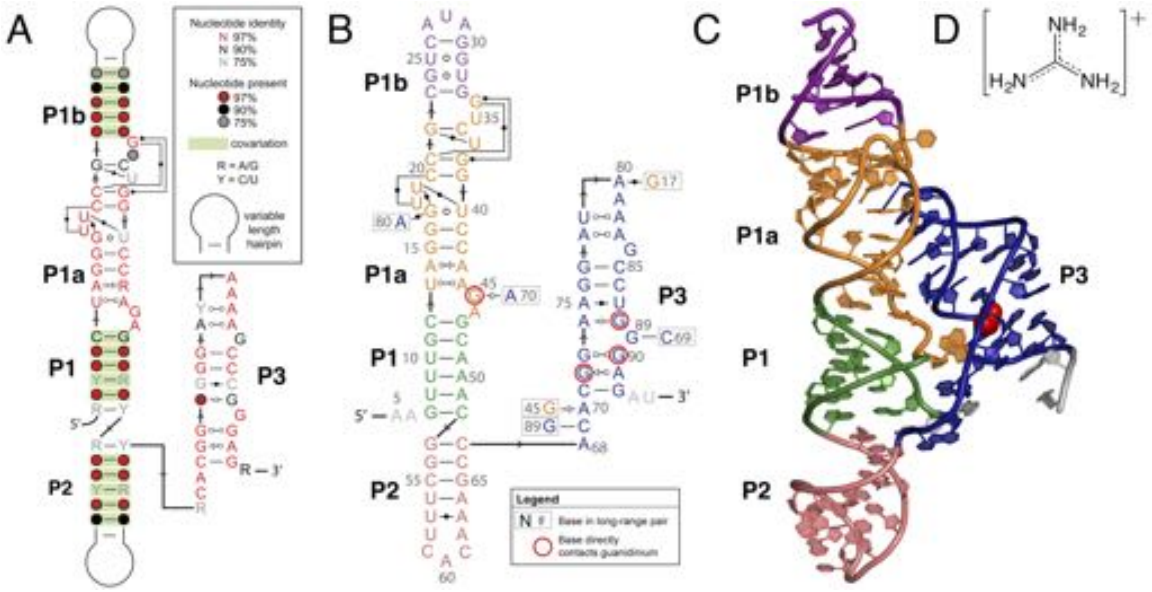


Figure 2.1: A. Updated consensus sequence of the guanidine-I riboswitch. Conservation is indicated according to the key in the upper righthand corner of the consensus diagram. B. Secondary structure of the guanidine-I riboswitch from *S. acidophilus*. Helices P1, P1a, P1b, P2, and P3 are colored in green, orange, purple, pink, and navy, respectively. Nucleotides circled in red directly contact the ligand via the base. C. Cartoon model of the guanidine-I riboswitch crystal structure from *S. acidophilus*, where the helices are colored as in part B. Guanidine is depicted as red spheres. D. Chemical structure of guanidine in the cation form, which is stabilized by resonance and is the predominant form at physiological pH.

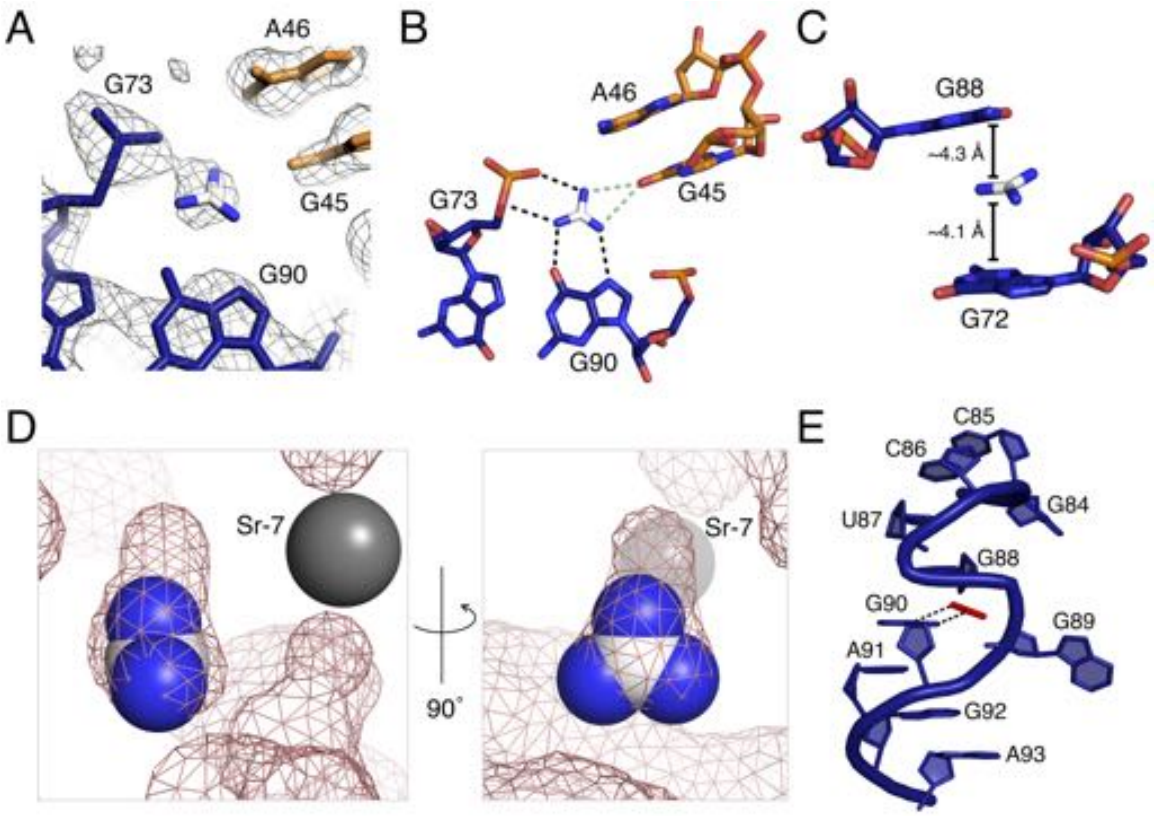


Figure 2.2: Binding pocket electron density and proposed contacts made by the ligand to the riboswitch. (A) Unbiased electron density map contoured at 1$\sigma$. Color scheme is the same as in 1B and C. For guanidinium, carbon is white and nitrogen is blue. (B) Top-down view of the binding pocket. Black dashed lines represent hydrogen bonds. The green dashed lines represent two mutually exclusive H-bonds. Oxygen atoms are represented in red and nitrogen in blue. (C) A side view of the binding pocket showing two guanines sandwiching the guanidinium ligand. (D) Guanidinium binding cavity. Guanidinium is shown in the sphere representation using the same color scheme as in A. The grey sphere is a strontium ion in the crystal structure. The pink mesh is a surface view of the riboswitch, including metal ions and excluding the ligand. (E) An S-turn involving two bases located in the binding pocket. Guanidinium is shown in red. The dashed lines are two H-bonds from guanidinium to G90. See Figure 2.3 for additional hydrogen bonding contacts.

edge, the ligand donates two hydrogen bonds to the Hoogsteen face of G90 in the P3 helix in a bifurcated manner. One amine donates to the N7 imino of G90 and another donates to the O6 carbonyl oxygen. On the second edge, two hydrogen bonds are made between the two guanidinium amines and the 5′ bridging and pro-$R_p$ non-bridging phosphate oxygens of G73 (Figure 2.2B). The proximity of the positively charged guanidinium cation and the negatively charged phosphate of G73 is also likely to provide an ionic stabilizing effect on ligand binding. On the third guanidinium edge, two nitrogens are approximately equidistant from the carbonyl oxygen of G45 (3.3 $Å^2$ and 3.4 $Å^2$). They could both donate a hydrogen bond; however, they appear to be mutually exclusive of each other based upon the narrow bond angle that would be formed between them (39.1°). I predict that both are alternatingly sampled upon ligand binding (Figure 2.2B). Alternatively, since the lone pairs on O6 may not be in the correct plane to accept the predicted hydrogen bonds, the interaction with O6 may be mainly electrostatic.

In addition to hydrogen bonding interactions on the guanidinium edges, there are also contacts with the top and bottom faces of the ligand. The guanidinium is positioned to potentially engage in π-cation interactions with G72 and G88 (Figure 2.2C). In this interaction, positively charged species form noncovalent interactions with the electron-rich π orbitals of an aromatic system, like the purine bases [94]. The distances between the guanidinium and the centers of the 6-membered rings in G72 and G88 are 4.1 Å and 4.3 Å respectively. The ϑ angles between guanidinium and the centers of the 6-membered rings of each base are 21.0° for G72 and 44.5° for G88. The distances and angles are consistent with cation-π interactions found in proteins and RNA [95; 96].

Surface models demonstrate that the guanidinium binding pocket is almost com-

pletely enclosed (Figure 2.2D). A calculation of the solvent accessible surface area (SASA) in PyMol shows that the SASA of guanidinium in complex with the riboswitch is 0.30 Å$^2$, compared to 62.68 Å$^2$ of total surface area in solution. Expressed another way, 99.5% of the guanidinium ligand is buried within the riboswitch. The surface model demonstrates that there is little room in the binding pocket for anything larger than guanidinium to bind (Figure 2.2D). There is a small pocket of extra space next to one guanidinium $NH_2$ that would potentially be large enough to accommodate one additional atom. Consistent with this, methylguanidine, hydroxyguanidine, and aminoguanidine turn on expression in a *B. subtilis* guanidine riboswitch reporter system, but it is a weak response relative to that of guanidine [40]. The small size of the binding pocket supports the finding that other common metabolites containing guanidino groups, including arginine, are too large and are not able to efficiently bind the riboswitch [40].

The architecture of the binding site is formed using an S-turn motif involving G90 and G88 (Figure 2.2E). This motif is characterized by two successive bends in the RNA backbone, resulting in the eponymous "S" shape [97; 98]. A single base, in this case G88, is oriented with the ribose in the reverse direction from the rest of the strand and adopting a C2′-endo conformation [97; 99]. This allows G88 to be positioned above the ligand to form a putative $\pi$-cation interaction with guanidinium. Nucleotide G90 is positioned in the normal orientation to hydrogen bond with the guanidinium. G89 is bulged out and forms a Watson-Crick (WC) base pair with highly conserved C69, cinching the two strands of the P3 helix together near the binding pocket (Figure 2.3A).

The nucleotides interacting with guanidinium form base-pairing networks with each other and with other nucleotides in the P3 helix. G90 and G73 form a trans Watson-

Crick/Hoogsteen base pair, which orients G90 and G73 to form their respective hydrogen bonds with guanidinium (Figure 2.3B). G45 and A46 flip out from helix P1a and insert into P3 (Figure 2.3C). The 2′-hydroxyl of A46 forms a single hydrogen bond with a non-bridging phosphate oxygen on C71, which likely plays a role in stabilizing its position in the P3 helix (Figure 2.3C). G45 forms a sugar-edge/sugar-edge base pair in the trans orientation with A70 as well as a hydrogen bond between its N1 imino and the carbonyl oxygen of G72, possibly to stabilize G45 in the flipped-out position (Figure 2.3D). A sheared base pair is formed between A91, which stacks on G90, and G72 (Figure 2.3D). This positions G72 to form the predicted $\pi$-cation interactions with guanidinium (Figure 2.2C). The N2 of G72 donates a hydrogen bond to the pro-$R_P$ non-bridging phosphate oxygen of G90.

### 2.2.4 Equilibrium binding analysis of binding site mutants

To further probe the importance of bases near the guanidinium, I used equilibrium dialysis to measure the $K_d$ for individual point mutants at four conserved residues within the binding pocket (Figure 2.4). Under these conditions, the guanidinium (a chaotropic agent) is present in trace amounts, eliminating the possibility of denaturing effects on the RNA. The wild type *S. acidophilus* guanidine riboswitch and the crystal construct with an engineered iridium hexamine binding site have an equilibrium binding affinity of 20 $\mu$M, which is comparable to the affinity measured by in-line probing for guanidine-I riboswitches from other bacteria [40]. In contrast, G45A and G90A both exhibited less than 10% binding at the highest RNA concentration possible in the assay. This corresponds to binding defects of more than 100-fold. Mutation of either G45 or G90 is predicted to disrupt a hydrogen bond between a guanidinium $NH_2$ and a carbonyl oxygen on the RNA base. I also

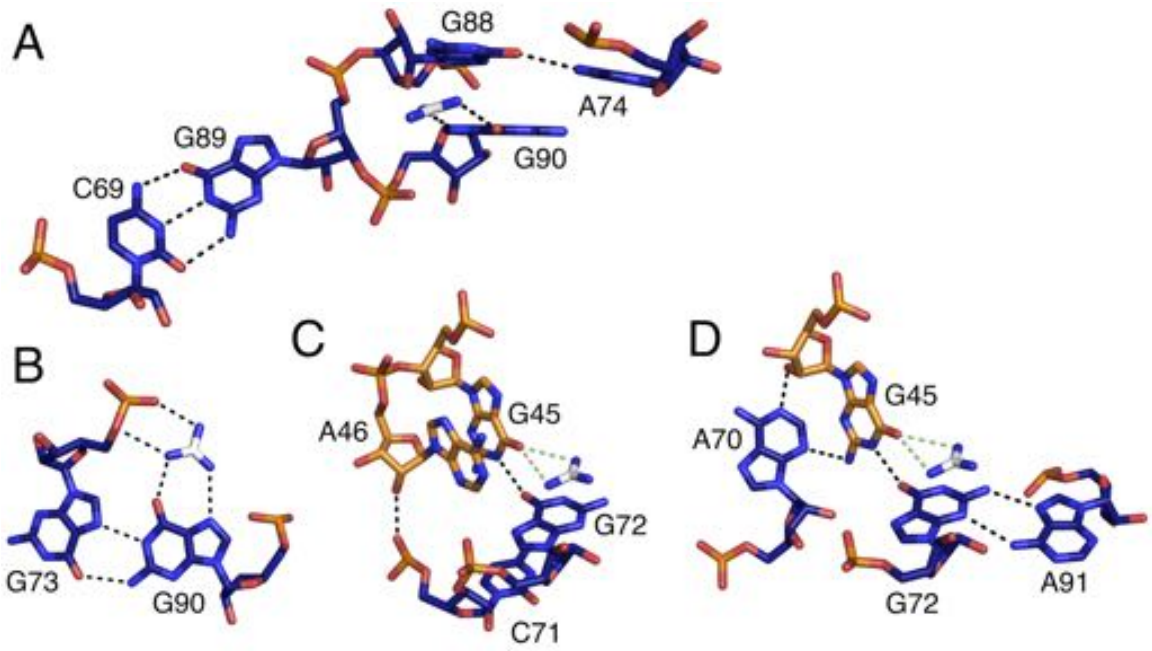


Figure 2.3: Base pairing network surrounding the guanidinium binding pocket. (A) View of the C69-G89 base pair relative to the guanidinium. G89 is located within the S-turn. The color scheme is the same as in 1B and C. (B) Base pair between G73 and G90, which both interact with guanidinium. (C) Hydrogen bonding network involving G45, A46, C71, and G72. (D) Hydrogen bonding network involving A70, G45, G72, and A91 (same POV as 1C with different nucleotides shown).

measured binding to the G72C mutation to explore the effects of possible cation-$\pi$ interactions. G72C showed less than 10% binding at the highest concentration range tested. The A46C mutant showed a 40-fold binding defect, consistent with the structural observation that guanidinium H-bond donors are too distant to form strong hydrogen bonds with the N1 of A46 (3.5 Å).

### 2.2.5 Tertiary interactions and metal ions observed

The main tertiary contacts within the aptamer are formed by docking the P3 helix into P1a, an interaction that is critical for creating the binding pocket. P3 specifically recognizes the P1a minor groove through A-minor interactions. A highly conserved series of two base triples and a base quadruple in P1a may open the minor groove, allowing contact with a highly conserved run of A′s in the hairpin loop of P3. A81, A82, and A83 form A-minor interactions with the P1a helix (Figure 2.5A and 4B). A81 and A82 form type II A-minor interactions with G17 and G16, respectively, while A83 forms a type I A-minor interaction with the G15-C42 base pair (Figure 2.5B).

Two highly conserved base triples and a highly conserved base quadruple distort the P1a minor groove. To form internal base triples from only two RNA strands, a number of nucleotides bulge out and reinsert into either the major or minor groove of neighboring base pairs. The G22-C36 base pair in P1a ($\geq$90% conserved) stacks directly below the C23-G33 base pair at the bottom of helix P1b (not pictured). U35 is flipped outward. G34 ($\geq$97% conserved) bulges and re-inserts into the major groove of P1a. G34 forms a base triple with the C21-G38 WC base pair (both $\geq$97% conserved) (Figure 2.5C). Non-conserved nucleotide U37 bulges out and forms a weak dinucleotide platform with C36

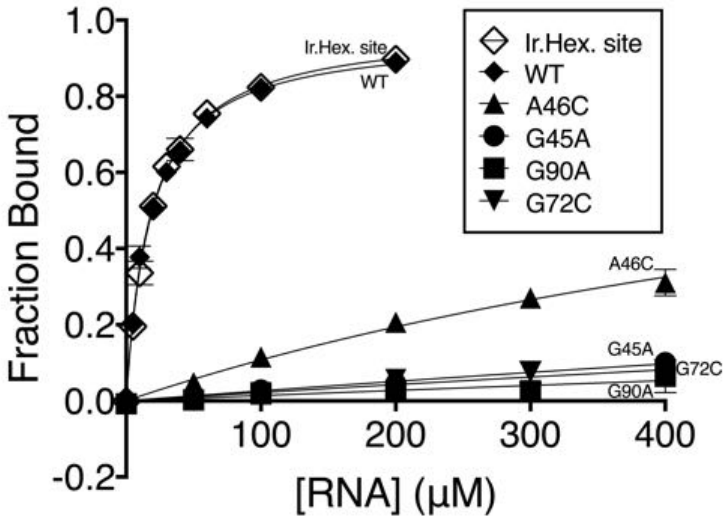


Figure 2.4: Binding affinity of [$^{14}$C]-guanidinium to the wild type riboswitch and point mutants by equilibrium dialysis. All point mutants are done in the background of the wild type RNA. WT, G45A, G90A, A46C, and G72C are represented by closed diamonds, circles, squares, triangles, and inverted triangles, respectively. The crystal construct, which contains an iridium hexamine binding site, is labeled "Ir.Hex. site" in the figure and represented by open diamonds. The data are fit to a hyperbolic curve. For the point mutants, $B_{max}$ is constrained to 1.0 (see experimental procedures section for more information). Fraction bound is shown with standard error bars with two replicates for each data point.

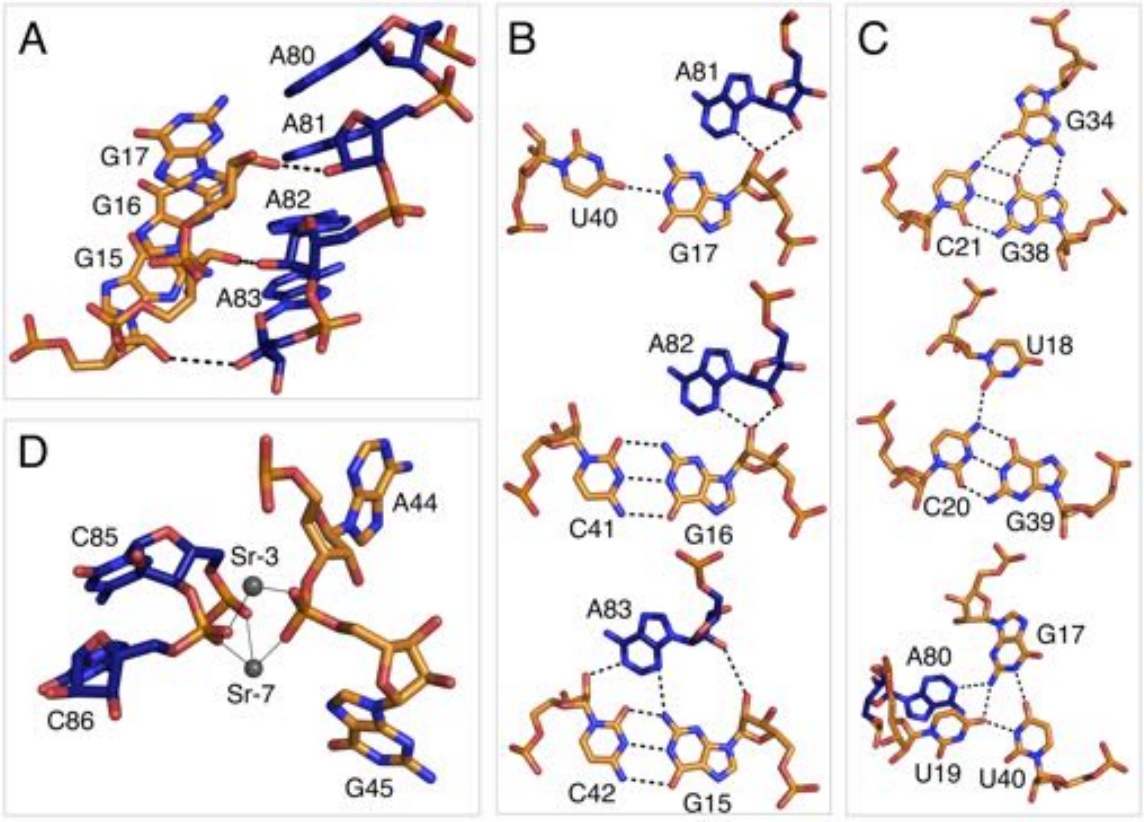


Figure 2.5: Long-range tertiary interactions involving highly conserved bases. (A) Side view of the highly conserved A patch (A80-A83) and adjacent bases (G15-G17) involved in A-minor interactions. The color scheme is the same as in 1B and C. (B) Detailed view of A-minor interactions between helix P1a and the A patch in P3. (C) Highly conserved base triples and a base quadruple in helix P1a. (D) Strontium ions Sr-3 and Sr-7 in the crystal structure coordinating P1a and P3 phosphate oxygens. Thin, black lines indicate inner sphere coordination.

(not pictured). The G39-C20 WC base pair (≥97% conserved) forms a base triple with U18 (≥97% conserved), in which the primary amine of C20 donates a hydrogen bond to the O1 of U18 (Figure 2.5C). U19 bulges into the minor groove and engages in a base quadruple.

A base quadruple is formed by U40, G17, U19, and A80 (Figure 2.5C). U40 and G17 engage in a pseudo-GU wobble with only one hydrogen bond between the N1 imino of G17 and the O4 of U40. The O4 of U19 accepts two hydrogen bonds, one from the N3 imino of U40 and one from the primary amine of G17. The primary amine of G17 also donates a hydrogen bond to the N1 imino of A80. This base quadruple is further stabilized by stacking of U19 and A80. The set of base triples distorts the P1a helix, allowing a base quadruple to form that specifically interacts with A80 to help anchor the A-rich hairpin loop into the minor groove of P1a.

Two metal ions also appear to stabilize the tertiary docking of P3 into P1a. The two metals (Sr-3 and Sr-7) facilitate the close approach of the phosphate backbones of the two helices by coordinating between the non-bridging phosphate oxygens of G45, C86, and C85 (Figure 2.5D). Both of these metal ions are $Sr^{2+}$ in the crystal structure, since the crystallization condition contains $SrCl_2$. Under cellular conditions, we expect that these are magnesium ions that make similar inner-sphere coordinations. One metal ion (Sr-7 in the crystal structure) coordinates to the pro-$S_P$ oxygen of G45 (2.3 Å), the pro-$S_P$ oxygen of C85 (2.3 Å), and the pro-$R_P$ oxygen of C86 (2.3 Å). The bond distances are consistent with inner sphere coordination. The other metal ion (Sr-3) coordinates to the pro-$R_P$ oxygen of G45 (2.3 Å) and the pro-$S_P$ oxygen of C86 (2.2 Å), again with bond distances consistent with inner sphere coordination. The distance of Sr-3 to the pro-$S_P$ oxygen of A44 is 3.3

Å suggestive of water-mediated coordination to this phosphate. As previously noted, Sr-7 appears to close the guanidinium binding pocket, thus limiting the size of the ligand (Figure 2.5D).

## 2.3 Discussion

Compounds containing guanidino groups, including arginine and agmatine, are prevalent in bacterial cells. In this environment, the guanidine riboswitch faces the challenge of specifically recognizing guanidinium while excluding larger compounds with a guanidino functional group. The riboswitch accomplishes this by recognizing every possible surface of guanidinium, including all three edges and both faces, leaving almost no solvent-exposed surface area around the ligand. One clear pattern in the recognition of guanidinium is the presence of electronegative oxygen atoms in the binding pocket, which act as hydrogen bond acceptors for the $NH_2$ groups in guanidinium. The only non-oxygen hydrogen bond acceptor interacting with guanidinium is the N7-imino on G90. Despite possessing a three-fold axis of symmetry, the three identical $NH_2$ groups in guanidinium are recognized in an asymmetric fashion. Above and below the two-fold plane of symmetry, the recognition is pseudo-symmetric. Guanine bases above and below guanidinium (G72 and G88) are predicted to form πcation interactions. In protein-RNA interactions, the guanidino group of arginine commonly interacts with RNA bases via π-cation interactions, suggesting that the guanidinium ion alone could use the same mechanism of binding to RNA [95; 100; 96]. The residues involved in guanidine binding within the crystal structure are the same ones that undergo a conformational change upon guanidine addition. In-line probing data shows six nucleotides have reduced strand cleavage upon guanidine binding [40]. Four of these nucleotides (G45, A46, G72, and G73 in *S. acidophilus*) are in the immediate vicinity of guanidinium. The other two nucleotides are A70 and C71. As previously mentioned, A70 contacts G45, which directly contacts guanidinium. C71 forms a Watson-Crick base pair

with G92 at the base of the P3 helix, suggesting that guanidine binding stabilizes the P3 helix.

Two distinct subtypes of the *ykkC* RNA have been identified [40]. Subtype 1, which is the guanidine-I riboswitch presented here, is the most common. Subtype 2 does not bind guanidine and is divided further into subclasses. Some of the ligands for the subtype 2 subclasses are known and discussed in the next chapter. The two subtypes differ at nucleotides 46, 72, and 88-91, all of which cluster in the binding site, further suggesting that these nucleotides are crucial for specific recognition of guanidine in subtype 1 and possibly rejection of subtype 2 ligands. Some subtype 2 variants also possess conserved 5′ and 3′ tails that are not conserved in subtype 1.

The P3 helix, which was not predicted by bioinformatics, is formed at the 3′ end of the riboswitch. This helix appears to be stabilized by guanidine binding and is most likely where switching of gene expression occurs. A potential terminator stem in the *S. acidophilus* riboswitch (Figure 2.6) includes nucleotides 84 through 94 of the aptamer. Formation of this helix would sequester six nucleotides that are more than 97% conserved, including G90 in the binding pocket and G89 in the S-turn. In the guanidinium-bound state, nucleotides 84-94 form eight base pairs with other nucleotides in the aptamer, four of which are canonical GC pairs. In the unbound state, assuming that the P3 helix does not form, residues 84-94 could form eleven base pairs within the terminator stem, six of which are canonical GC pairs. Guanidine-dependent switching is likely to occur by stabilization of the P3 helix, allowing it to outcompete terminator stem formation in the presence of the ligand. To test if the predicted expression platform is actually used during switching, future studies with *in vitro* transcription termination could be performed to determine the site of

termination. Mutations could also be made to destabilize the terminator stem followed by compensatory mutations to rescue any defect, although this may be difficult due to the very high degree of conservation in the switching region. The same mutation experiments could be performed using a reporter system in bacteria to examine the switching mechanism *in vivo*.

The structure of this riboswitch reveals the basis for specific binding of guanidinium to RNA. It also reveals two previously unpredicted helices, P1a and P3. Both of the unpredicted helices form complex tertiary interactions crucial for riboswitch folding and guanidinium recognition. P3 contains the binding pocket and a portion of the predicted terminator stem, providing a possible mechanism for switching through stabilization of the P3 helix upon guanidinium binding. This allows guanidinium binding to modulate the expression of downstream genes through the riboswitch. The extensive and selective recognition of guanidinium by the riboswitch provides further evidence that free guanidine is a biologically relevant compound encountered by bacteria. Other biochemical data shows that genes controlled by this riboswitch act on guanidine as a substrate or remove it from the cell. Collectively, these data suggest that guanidine is biologically relevant compound in bacteria whose concentration must be managed to avoid toxicity.

This structure also made sense of the consensus motif differences observed between subtype 1 and the subtype 2 variant riboswitches. Nucleotides that are different between the two classes clustered around the binding pocket of guanidine, suggesting that the variant RNAs bind a different ligand, but in the same area as the subtype 1 binding pocket.

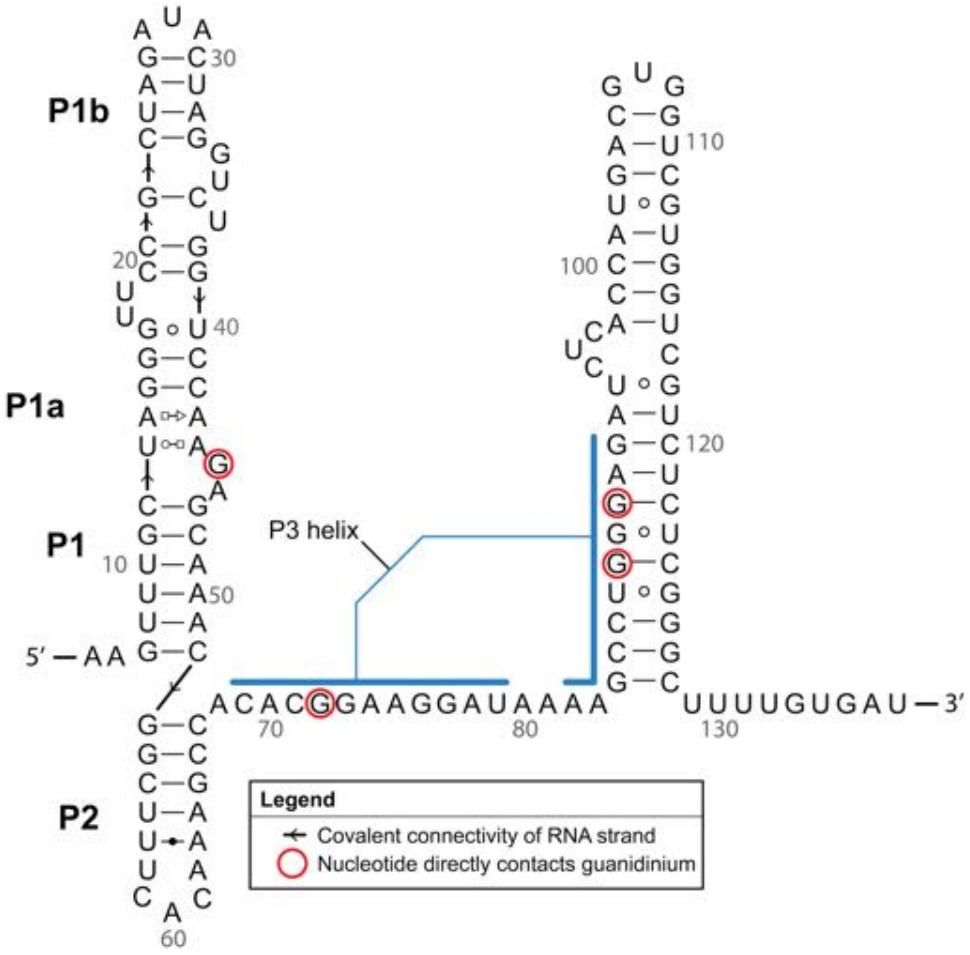


Figure 2.6: Proposed secondary structure of the wild type *Sulfobacillus acidophilus* guanidine riboswitch in the absence of ligand. Thick blue lines indicate nucleotides that form helix P3 in the ligand-bound state based on the crystal structure. Red-circled bases directly contact guanidinium in the ligand-bound state. The base pair notation is the same as previously published (Leontis and Westhof, 2001). The AUG start codon for the controlled gene is located 28 nucleotides downstream of the proposed terminator stem (not shown).

# 3 *ykkC* RNA variants: structures of the PRPP and ppGpp aptamers

This story was a close collaboration with Andrew Knappenberger in the Strobel Lab. Andrew solved the PRPP riboswitch crystal structure and performed some of the biochemical experiments. He and I participated equally in preparing the following story, which was adapted from a manuscript currently under review (Knappenberger*, Reiss*, and Strobel, under review).

## 3.1 Background

RNA has diverse functional capabilities, which has driven speculation that the first organisms may have been RNA-based [47; 15; 17; 16; 101; 102; 14]. For this hypothesis to be plausible, RNA must be adaptable; i.e., capable of acquiring new functions through mutation. In the field of evolutionary biology, this trait is described as evolvability [103; 20]. Evolvability is the propensity of a system to produce a mutated genotype that yields a beneficial phenotype under new selective pressures [104; 103; 18; 20]. Often, this occurs through mutation of an existing gene through divergent evolution. For example, bacterial β-lactamases demonstrate significant evolvability through mutations in the ω-loop. This loop determines substrate specificity, but mutation or outright deletion of the loop does not dramatically affect the overall structure of the protein [105; 75; 76; 77]. This locus of evolvability allows the protein to adapt to the selective pressures of novel antibiotics. The concept of evolvability has also been studied in RNA, including a notable paper by Draghi

et al. [19]. This study found that adaptation rate is hastened when the build-up of some phenotypically neutral mutations occurs and the web of accessible phenotypes becomes broad. The speed at which an organism adapts is determined by this, as well as the ruggedness of the fitness landscape, which is related to the number of mutations required to reach a new fitness maximum.

Variant riboswitches yield insight into the evolvability of RNA. Riboswitch variants are naturally occurring riboswitches with a conserved overall fold but altered ligand specificity. Examples include the guanine/adenine riboswitches and the cyclic-di-GMP/cyclic-GMP-AMP riboswitches [106; 28; 31; 107; 59; 108; 37]. Bioinformatic and structural studies of the guanine/adenine and cyclic-di-GMP/cyclic-GMP-AMP aptamers showed that the altered specificity occurs simply by changing base pairing between the RNA and ligand. Recently, the *ykkC* RNA motif was identified as binding to multiple, chemically dissimilar ligands, which makes this specific scaffold a compelling target for structural studies of RNA evolvability (Sherlock et al., in press) [40; 45].

The *ykkC* RNA was discovered in 2004 and its ligand(s) remained unknown for over a decade [29; 40]. In 2017, Nelson et al. published two pivotal discoveries regarding this motif: (1) the majority of these RNAs bind specifically to the guanidinium cation and (2) the *ykkC* riboswitch class can be divided into at least two subclasses. Subclass 1, which has approximately 1500 known examples, is the major class now known as the guanidine-I riboswitch. Subclass 2 was defined as all variants of this motif that do not recognize guanidine. The class 2 variants are overall quite similar to guanidine-I riboswitches. They retain the same overall fold, but possess a few characteristic differences at nucleotides crucial for guanidine binding. Notably, most of these differences are centered around a classic S-turn

motif that forms the binding pocket of the guanidine-I riboswitch. A similar overall architecture with key differences in binding pocket nucleotides is a signature characteristic of a riboswitch variant [58].

Variant *ykkC* RNAs (aka subtype 2 *ykkC* RNAs) are found upstream of a variety of genes, although two major groups are apparent. One major group regulates amino acid synthesis and transport genes, which are upregulated during the stringent response (Sherlock et al., in press). The other regulates *de novo* purine biosynthesis, which produces purine nucleotides from smaller metabolites under conditions where intact nucleobases are not available [45; 109; 110]. These riboswitches were designated as *ykkC* subtype 2a and 2b, respectively. When compared to guanidine riboswitches, subtypes 2a and 2b harbor systematic changes to residues directly involved in guanidine binding, which led to the suggestion that they may have different ligand specificity. For example, where guanidine riboswitches have a conserved adenosine residue (A46 in the guanidine riboswitch), subtypes 2a and 2b have a pyrimidine (C49 in the present study) [80; 40; 45] (Sherlock et al., in press). For clarity, differences in conservation at this position and subsequently discussed positions are summarized by subtype in Table 3.1. Sorting the entire *ykkC* class by the identity at position 46/49 alone results in a strikingly complete segregation of guanidine-related gene contexts from those that are incongruent with mitigation of guanidine toxicity [40]. Alignment of subtype 1, 2a, and 2b sequences also shows an extension of conservation at both the $5'$ and $3'$ ends of the 2a and 2b aptamer subtypes. These key differences in conserved residues and gene contexts suggested that these *ykkC* variants have altered ligand specificity while retaining the same overall architecture [80; 40].

Table 3.1: Summary of conservation in *ykkC* RNA subtypes 1, 2a, and 2b at key nucleotides

| | | | Conservation at position:* | | |
|---|---|---|---|---|---|
| **Subtype** | **Ligand** | **Downstream genes** | **46/49** | **69/75** | **89/96** |
| Subtype 1 | Guanidine | Guanidine carboxylases, guanidine transporters, etc. | A | C | G |
| Subtype 2a | (p)ppGpp | *de novo* purine biosynthesis | Y | C | G |
| Subtype 2b | PRPP | Amino acid synthesis genes and ABC transporters | Y | C | N |

*the position number in the *S. acidophilus* guanidine-I riboswitch aptamer is displayed first [80], followed by the equivalent position number in the *T. mathranii* PRPP riboswitch aptamer. Y indicates the position is conserved as a pyrimidine (C or U) and N indicates the position does not display conservation.

Subtype 2a and 2b *ykkC* riboswitches do not retain the ligand specificity of their parent riboswitch. Using transcription termination and in-line probing assays, Sherlock et al. found that neither subtype is responsive to guanidine. Instead, subtype 2a is responsive to guanosine tetra/pentaphosphate ((p)ppGpp, hereafter referred to as ppGpp), an alarmone that regulates the stringent response [111; 112; 113]. Subtype 2b is responsive to phosphoribosyl pyrophosphate (PRPP), a precursor in purine biosynthesis. Like the guanidine riboswitch, both function as ON switches. The consensus motifs for subtypes 2a and 2b are remarkably similar to each other, even relative to other *ykkC* RNAs (Sherlock et al., in press) [45]. The most apparent difference is a highly-conserved guanosine (G96 in this study) in subtype 2b that is not conserved in subtype 2a (Table 3.1). This residue is equivalent to G89 in the guanidine-I riboswitch and is a conserved part of its S-turn motif. In the guanidine-I structure, G89 flips out to form a Watson-Crick base pair with conserved C69 (C75 in this study) [80]. Although bioinformatic data suggest that variation in G96 is central to the structural differences between subtype 2a and 2b riboswitches, its precise role in this context remains uncertain.

Unlike guanidine, the biological roles of PRPP and ppGpp are both well-documented. PRPP is an activated form of ribose 5-phosphate, and a major macromolecular building block of RNA [114]. It is a central metabolite used in biosynthesis of purine and pyrimidine nucleotides, the amino acids histidine and tryptophan, nicotinamide adenine dinucleotide, thiamine diphosphate, flavins, and pterins [115; 116; 117]. The centrality of PRPP within metabolism makes it an appealing target for regulation. ppGpp is an alarmone that initiates the stringent response, a global reaction to nutrient starvation in bacteria [111; 118; 119; 120]. Amino acid starvation triggers synthesis of ppGpp and ppGpp binds to a variety of effector molecules to initiate sweeping changes in the cell's transcriptional profile, including a reduction in tRNA and rRNA synthesis and transcription of amino acid biosynthesis genes [118; 121; 122; 123]. Consistent with a role in the stringent response, the ppGpp riboswitch turns on transcription of amino acid biosynthesis and transport genes in response to alarmone binding (Sherlock et al., in press).

A common ancestral RNA likely diverged to recognize guanidine, PRPP, and ppGpp in spite of the chemical and structural diversity among these ligands. PRPP and ppGpp are more similar to each other than either is to guanidine, which reflects the greater similarity in their aptamers (Figure 3.1). While guanidine harbors a single delocalized positive charge, PRPP and ppGpp harbor multiple separate loci of negative charge. Guanidine is small and achiral with three-fold rotational symmetry, while PRPP and ppGpp are larger, chiral, asymmetric molecules. PRPP and ppGpp both contain ribose sugars and pyrophosphate moieties, but ppGpp has an entire guanine base that PRPP lacks. Bioinformatic evidence suggests that the 2a and 2b aptamers represent an especially concise solution to a central biophysical problem: biologically relevant switching entails recognition

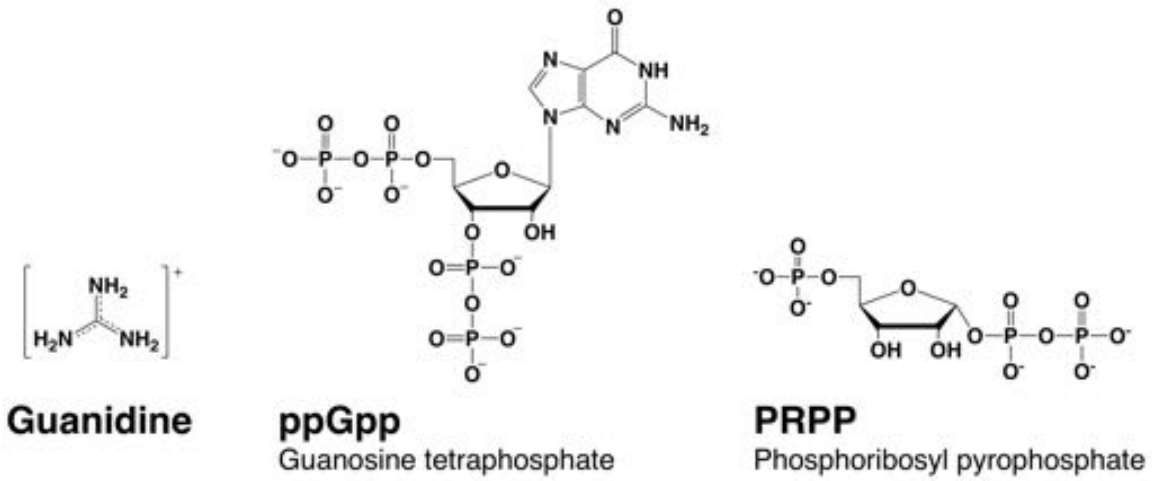


Figure 3.1: Chemical structures of guanidine, phosphoribosyl pyrophosphate (PRPP), and guanosine tetraphosphate (ppGpp), which are the ligands for *ykkC* subtypes 1, 2a, and 2b, respectively.

of a cognate ligand and rejection of structurally similar alternatives. We set out to determine how three RNA elements with a common scaffold could recognize such dissimilar ligands with high specificity. Central questions include how a polyanionic macromolecule differentially recognizes two distinct small polyanions, and how the presence or absence of the guanine base changes the RNA's recognition strategy. To address these questions of molecular recognition by RNA, we report the atomic resolution structure of a native *ykkC* 2b riboswitch in complex with PRPP via X-ray crystallography. We also convert this construct into a ppGpp-binding aptamer with a single G96A mutation and present the structure of the mutant bound to ppGpp. The sequence of the G96A aptamer is consistent with the consensus motif for native subtype 2a riboswitches. This structural and biochemical information reveals how the *ykkC* RNA differentiates between ppGpp and PRPP. This study showcases the functional plasticity of RNAs and the evolvability of RNA function from a single structural scaffold.

## 3.2 Results

### 3.2.1 The structure of the wild type PRPP aptamer and a single point mutant ppGpp aptamer

To understand the basis of ligand recognition by the PRPP riboswitch, we determined the crystal structure of the aptamer domain of the *ykkC* 2b riboswitch from *Thermoanaerobacter mathranii* at 2.5 Å resolution in the presence of its native ligand PRPP (Table 3.2). PRPP is an activated metabolic intermediate. As a result, it is highly unstable. It degrades on a time course of minutes to hours via several mechanisms in the presence of divalent metal ions, acidic or basic pH, and/or elevated temperatures [114; 124; 125; 126; 127; 128]. However, binding to the PRPP riboswitch aptamer domain protects PRPP on a time scale of hours to days (Figure 3.2). The stabilizing effect of the aptamer permitted crystals of the intact complex to be observed after two days. Once formed, unfrozen crystals disappeared after approximately five to ten days, underscoring the need for prompt crystallization and cryogenic preservation in this study. This construct crystallizes in the presence of $BaCl_2$, so the divalent metal ions, including those associated with PRPP, are $Ba_{2+}$ ions. The structure was solved by molecular replacement using the guanidine-I aptamer as an initial model. After model building and refinement, the model fit the data with an $R^{work}$ of 0.216 and an $R_{free}$ of 0.253.

Like its parent aptamer, the PRPP riboswitch contains two adjacent helical stacks (Figure 3.2.1). P3 forms a large portion of the binding pocket, and a conserved loop at the end of P3 docks into P1a. This allows conserved nucleotides from P1a to participate in ligand recognition. P1, P1a, P1b, and P2 together form one continuously stacked region

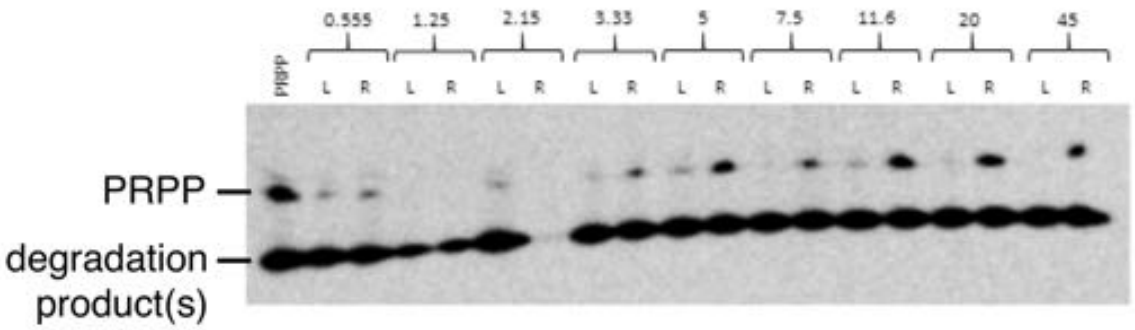


Figure 3.2: Autoradiograph of a representative PAGE gel from dissociation constant determination for the PRPP aptamer (see also Figure 3.6). The top band represents intact PRPP and the bottom band represents degradation products. The leftmost lane is purified PRPP, which has already experienced degradation. Each pair of lanes thereafter represents one dialysis cassette. The concentration of aptamer added to the cassette in micromolar is depicted above each pair. "L" is the side of the cassette to which PRPP ligand was initially added. "R" is the side of the cassette to which RNA aptamer was initially added. After overnight incubation at room temperature, material recovered from dialysis cassettes was electrophoresed to separate intact PRPP from degradation products. PRPP incubated in the presence of aptamer experienced significant protection relative to unbound PRPP, permitting the use of this unstable metabolite in the present structural and biophysical studies.

Table 3.2: X-ray statistics for the PRPP and ppGpp aptamers

| RNA | Wild type | G96A mutant |
|---|---|---|
| **Data Collection:** | | |
| Beamline | 24-IDE at APS | 24-IDE at APS |
| Space group | P 1 21 1 | P 1 |
| **Unit cell:** | | |
| a, b, c (Å) | 34.0, 89.6, 136.4 | 53.3, 62.1, 125.9 |
| $\alpha$, $\beta$, $\gamma$ (°) | 90.0, 96.9, 90.0 | 91.2, 89.6, 101.9 |
| Wavelength (Å) | 0.97918 | 0.97918 |
| Resolution (Å) | 40.00 - 2.49 | 40.00 - 3.10 (3.15 - 3.10) |
| $R_{merge}$ | 0.100 (1.068) | 0.100 (1.222) |
| I/$\sigma$I | 16.431 (1.27) | 8.36 (0.67) |
| $CC_{1/2}$ in highest-resolution shell | 0.176 | 0.253 |
| CC* in highest-resolution shell | 0.547 | 0.635 |
| Completeness (%) | 97.0 (99.6) | 93.7 (94.7) |
| Redundancy | 3.6 (3.9) | 2.2 (2.1) |
| Total Reflections | 769056 | 49095 |
| Unique Reflections | 27671 | 28567 |
| **Refinement:** | | |
| Resolution () | 37.35 - 2.49 | 36.54 - 3.10 |
| Number of reflections | 26309 | 26642 |
| $R_{work}/R_{free}$ | 0.216/0.253 | 0.248/0.304 |
| **Number of atoms:** | | |
| Total | 4545 | 9041 |
| RNA | 4418 | 8829 |
| Ligand | 44 | 144 |
| Cations | 31 | 39 |
| Water | 52 | 21 |
| **B factors:** | | |
| Overall ($Å^2$) | 83.9 | 144.6 |
| RNA ($Å^2$) | 84.1 | 145.3 |
| Ligand ($Å^2$) | 73.9 | 123 |
| Cations ($Å^2$) | 117.9 | 119 |
| Water ($Å^2$) | 62.3 | 95.6 |
| **Root-mean-square deviations:** | | |
| Bond lengths (Å) | 0.0086 | 0.018 |
| Bond angles (°) | 1.827 | 1.852 |

adjacent to P3. However, unlike the guanidine aptamer, the PRPP aptamer has structured tails at the 5′ and 3′ ends that are not conserved in the guanidine riboswitch. The ends pair to form an additional short helix that we have termed P0. P0 stacks with P3 and extends the binding pocket for recognition of the larger PRPP ligand [129]. The overall architecture of the PRPP aptamer reveals that it is a rather conservative adaptation of the guanidine aptamer with key differences that allow for PRPP recognition.

Although PRPP is unstable in solution, it has high occupancy in this crystal structure. PRPP is modeled in with an occupancy of 1, and its B factors refined similarly to those of nearby residues. The quality of the fit between the electron density data and this model shows that a combination of protection by the riboswitch and a vast molar excess of ligand permitted a high degree of aptamer saturation when data were collected.

PRPP is a potentially challenging ligand for RNA to recognize; it has three negatively charged phosphate groups and lacks a moiety resembling a nucleobase. PRPP is known to interact with two divalent metal ions per molecule in solution. The 5-phosphate associates weakly with one metal and the pyrophosphate moiety more strongly coordinates a second metal (see Thompson et al., 1978) [130]. In the current model, these two metals are present in the complex with the riboswitch, suggesting that the mono- or dimetal complex is the entity recognized in vivo (Figure 3.4A, B). One metal (M1) associates with the 5-phosphate, and the second metal (M2) associates with the pyrophosphate. Both metals form contacts bridging PRPP and the RNA aptamer. A third metal ion, M3, is also near the 5-phosphate but is only coordinated by the aptamer. The three phosphate groups are major elements of recognition via interactions with nucleobase amines and divalent metal ions.

The 5-phosphate of PRPP experiences recognition by a metal ion and the amino

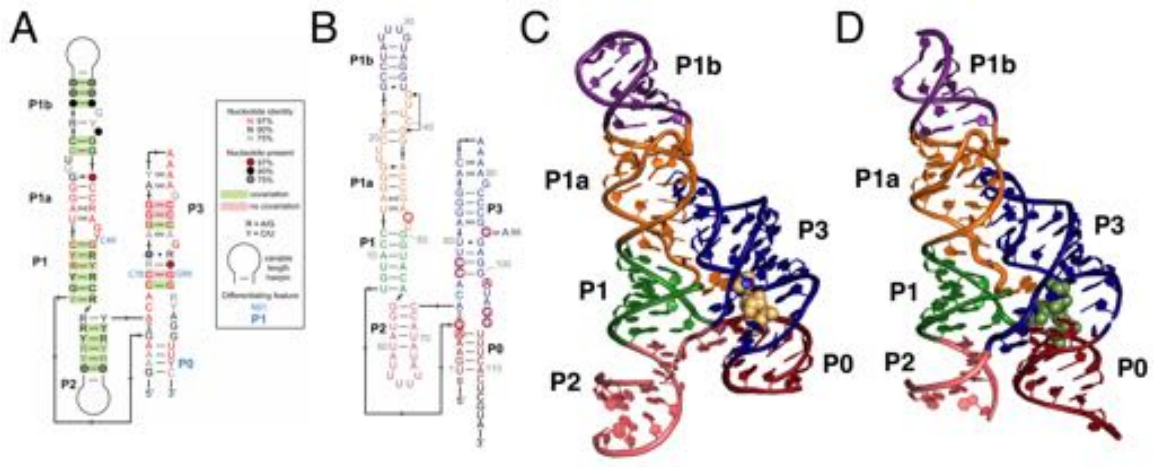


Figure 3.3: Overall structure of the PRPP riboswitch and its G96A mutant, which is a ppGpp aptamer. (A) Consensus sequence of the PRPP riboswitch. The secondary structure has been updated to show structural information gained from the present study. The sequence is depicted as in Sherlock et al. (unpublished) (see key). Nucleotides noted in blue are important bioinformatic differences between PRPP riboswitches and guanidine riboswitches. Base pair notation is as published by Leontis and Westhof. (B) Secondary structure of the PRPP riboswitch aptamer from *T. mathranii*. Nucleotides are colored by paired region. Labels for paired elements are indicated in bold. Sequence numbering is indicated in gray. Nucleotides whose bases directly contact PRPP are circled in red. (C) Crystal structure of the PRPP riboswitch. The RNA is depicted as a cartoon and PRPP is depicted as purple spheres. Nucleotides are colored by paired region as in B. (D) Crystal structure of the G96A mutant. The RNA is depicted as a cartoon and ppGpp is depicted as green spheres. Nucleotides are colored by paired region as in B.

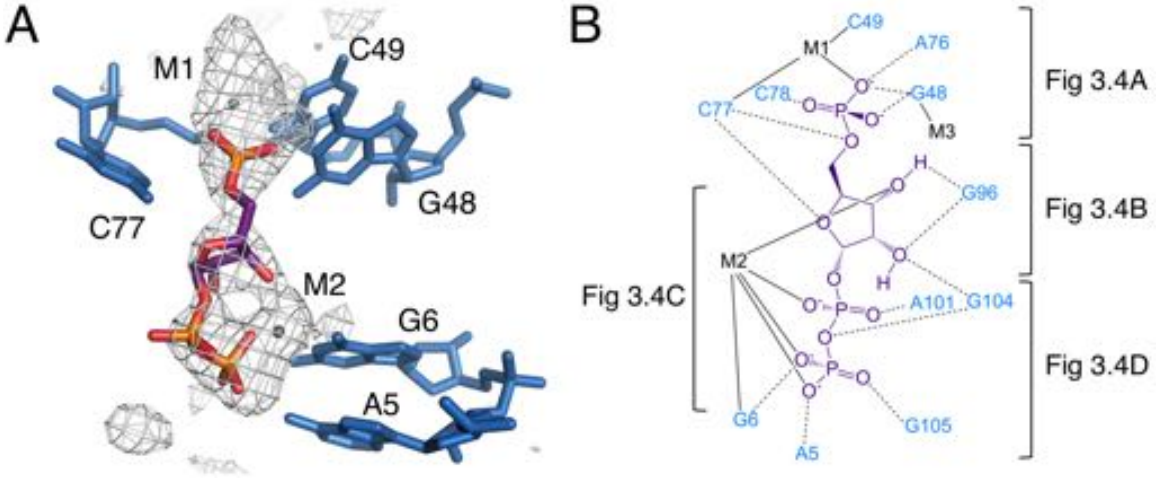


Figure 3.4: The binding pocket of the PRPP riboswitch. (A) Crystal structure of the ligand-binding site in chain A. PRPP is depicted as sticks and colored by element with purple carbons. Nucleotides are depicted as teal sticks. Metal ions are depicted as gray spheres. Individual nucleotides and metals are labeled. An Fo-Fc map contoured at 2.5$\sigma$ is shown as a gray mesh. The map was calculated using an otherwise complete model lacking PRPP, M1, and M2. (B) Ligand interaction map. The map is colored essentially as in A. All RNA and metal contacts to PRPP are shown. Dashed gray lines indicate hydrogen bonds. Solid gray lines indicate coordination to a metal ion.

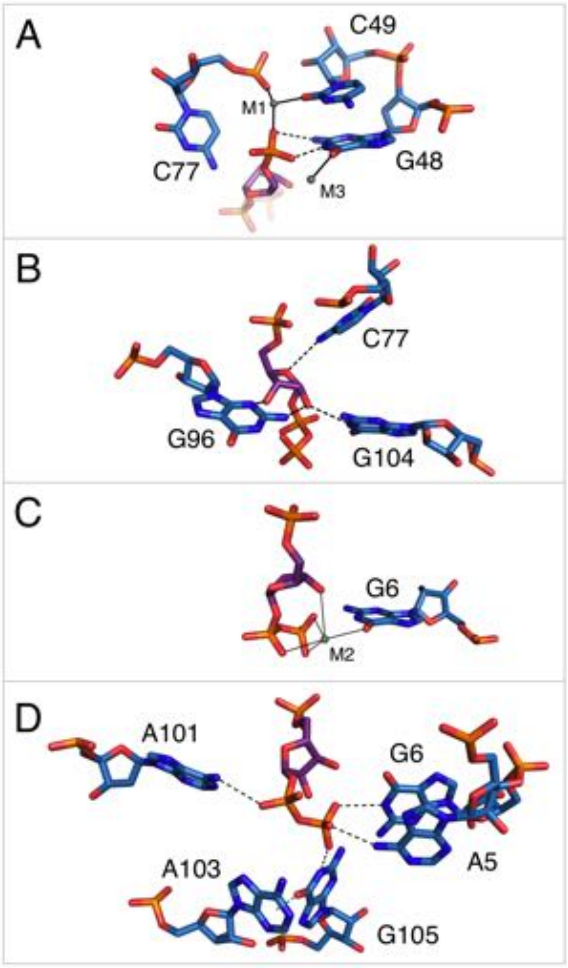


Figure 3.5: Notable contacts to the PRPP ligand. PRPP is depicted as sticks and colored by element with purple carbons. Nucleotides are depicted as sticks and colored by element with teal carbons. Dashed black lines indicate hydrogen bonds. A dashed green line shows the lone pair-$\pi$ interaction between A103 and G105. Solid black lines indicate coordination to a metal ion. (A) Contacts among the 5-phosphate of PRPP, residues G48, C49, C77, and metal ions M1 and M3. (B) Hydrogen bonds between the ribose of PRPP and residues C77, G96, and G104. (C) Coordination of metal M2 by PRPP and residue G6. (D) Recognition of the pyrophosphate group of PRPP by residues A5, G6, A101, and G105.

groups of conserved nucleotides (Figure 3.5A). The N1 and N2 of G48 form hydrogen bonds with two phosphoryl oxygens, while the N4 of C78 hydrogen bonds to the third non-bridging phosphoryl oxygen. The 5-phosphate also coordinates M1, which is held in place by coordination interactions with C77 and C49. The residue equivalent to C49 is conserved as an adenosine in the guanidine-I riboswitch, but is a pyrimidine in PRPP and ppGpp riboswitches. The identity of residue 49 was used as a marker to distinguish between these two variants [45]. The O6 of G48 coordinates M3, but M3 is too distant from the 5-phosphate to be coordinated by it.

The ribose moiety of PRPP also makes extensive interactions with the RNA aptamer (Figure 3.5B). The sugar face of G96 forms hydrogen bonds with the 2- and 3-hydroxyl groups. The N4 of C77 donates a hydrogen bond to the ribose oxygen, and the N1 group of G104 donates a hydrogen bond to the 2-hydroxyl group. These three residues are all highly conserved in the consensus sequence of this aptamer. At 2.5 Å resolution, conclusive determination of the ribose sugar pucker is not possible, but a C2-endo pucker is the most likely conformation in this complex and it fits the electron density data well. This conformation avoids a steric clash between the 2-hydroxyl and the $\beta$ phosphate. It allows for the 3-hydroxyl group to coordinate M2. This conformation is also consistent with previously reported structures of PRPP in complex with macromolecules [131; 132; 133].

The P0 region of the aptamer extends below P3 and permits a suite of interactions with the pyrophosphate group of PRPP (Figure 3.5C,D). The $\beta$ phosphate of PRPP is more extensively recognized than the $\alpha$ phosphate. The O6 of G6 coordinates M2, which forms several interactions with the pyrophosphate group (Figure 3.5C). The observed coordination interactions are consistent with previously documented barium inner sphere coordina-

tion geometry [134]. The N6 group of the weakly conserved A101 ($\geq$ 75% conserved as a purine) contacts a non-bridging oxygen of the $\alpha$ phosphate (Figure 3.5D). The N6 group of A5 and the N1 groups of G6 and G105 make direct contacts with non-bridging oxygens of the $\beta$ phosphate. An abrupt deformation in the local backbone conformation positions A103 under G105, allowing a lone pair-$\pi$ interaction to form between the O6 atom of G105 and the six-membered ring of A103 [135; 136; 137; 138; 139]. The present results show that the PRPP aptamer recognizes its ligand through a shifted and extended helical ligand-binding region, allowing for the retention of bound metal ions and extensive hydrogen bond donation to phosphate groups.

Some of the conserved bases in the binding pocket only make one hydrogen bond to PRPP. The hydrogen bond donors and acceptors not making contacts to PRPP are involved in other hydrogen bonding interactions within the RNA that shape the binding pocket.

The intracellular PRPP concentration in bacteria is estimated to be in the millimolar range [114; 140; 141; 142; 143; 144; 145]. However, enzymes and protein regulatory elements that sense PRPP concentrations in bacteria typically have micromolar dissociation ($K_d$) or Michaelis constants ($K_M$) [114; 146; 147]. Sherlock and colleagues recently found that the $T_{50}$ (the ligand concentration that produces half-maximal effect) of a PRPP riboswitch in transcription termination assays is 90 $\mu$M. The $K_d$ of the riboswitch aptamer domain for PRPP was determined (Table 3.3, see also Figure 3.6) by equilibrium dialysis using radiolabeled [$\beta$-$^{33}$P]-PRPP. This assay yields a $K_d$ of 2.0 $\pm$ 0.3 $\mu$M. There are two notable differences between the present experimental system and that employed by Sherlock et al.. First and most important, the present study examines binding affinity in an isolated

aptamer domain, while Sherlock et al. focused on the ability of the full riboswitch to terminate transcription. The full system is governed by the kinetics of ligand association and RNA folding, while the present experimental system only measures the thermodynamics of ligand binding. Also, in this study, [$\beta$-$^{33}$P]-PRPP was used in trace quantities and the amount of intact PRPP remaining in each sample was carefully measured to deconvolute the counts obtained from intact PRPP and the counts obtained from breakdown products. Sherlock et al. used unlabeled PRPP and could not quantify the amount of degradation, likely resulting in some underestimation of PRPP's ability to terminate transcription. Lastly, binding was measured here in the presence of 20 mM magnesium to encourage RNA folding, while Sherlock et al. performed transcription termination at 5 mM magnesium. The present data show that the affinity of the complex is at least of low micromolar affinity, placing it well within the range observed for complexes of PRPP with protein elements [146; 147]. In the future, binding could be measured closer to physiological magnesium concentrations to see how this affects the affinity.

### 3.2.2 A single mutation in the PRPP aptamer generates a ppGpp aptamer

In parallel with structural inquiries into the PRPP riboswitch, crystallization of native ppGpp aptamers was pursued. However, crystallization was unsuccessful with the subset of ppGpp aptamers tested. Considering the evident plasticity of the *ykkC* motif and the overt similarity between the consensus sequences of *ykkC* RNA subtypes 2a and 2b, a specificity switch of the PRPP aptamer to a ppGpp aptamer was pursued via mutation as an alternative strategy.

Close examination of the consensus motifs of the PRPP and ppGpp riboswitch

aptamers revealed that the ppGpp aptamer consensus sequence was almost entirely a subset of the PRPP aptamer consensus sequence, with the PRPP aptamer generally having more stringent requirements than the ppGpp aptamer. The most salient difference between the two consensus sequences is at position 96. In the PRPP aptamer, this position is $\geq 97\%$ conserved as a guanosine, but this conservation is lost in the ppGpp aptamer. The dramatic difference in conservation at this site suggested that it may be a lynchpin in differential recognition of PRPP and ppGpp.

We mutated position 96 in the *T. mathranii* PRPP aptamer from guanosine to adenosine, generating the G96A mutant. The wild-type aptamer shows poor affinity for ppGpp ($K_d$ = 91 $\pm$ 3 $\mu$M) and 46-fold greater affinity for PRPP ($K_d$ = 2.0 $\pm$ 0.3 $\mu$M) (Table 3.3). Conversely, the G96A mutant has affinity for ppGpp virtually identical to that of wild-type for PRPP ($K_d$ = 1.8 $\pm$ 0.1 $\mu$M) and PRPP binding is abolished in the mutant up to 400 $\mu$M RNA (estimated $K_d$ = 1600 $\pm$ 200 $\mu$M). The G96A mutant has at least 900-fold higher affinity for ppGpp than PRPP. The G96A mutation thus strikingly resulted in approximately a 40,000-fold switch in ligand specificity. The mutant's affinity for ppGpp is well within the range of native aptamers tested (data not shown).

### 3.2.3 Co-crystal structure of the generated ppGpp aptamer and its ligand

Having shown that the G96A mutant is a ppGpp aptamer, we solved its crystal structure in the presence of ppGpp to 3.1 Å resolution. The crystallization conditions that reproducibly gave rise to co-crystals of the wild-type PRPP aptamer did not yield comparable results for co-crystals of the G96A mutant. However, the G96A mutant was found to crystallize in a separate condition that also produced crystals of wild-type aptamer. The crystallization

Table 3.3: Dissociation constants for PRPP and ppGpp to wild type and G96A *T. mathranii* PRPP riboswitch aptamer

| **RNA** | **$K_d$ for PRPP** | **$K_d$ for ppGpp** | **Fold specificity for PRPP over ppGpp** | **Magnitude of specificity switch** |
|---|---|---|---|---|
| WT | $2.0 \pm 0.3$ μM | $91 \pm 3$ μM | 46 | |
| G96A | $1600 \pm 200$ μM | $1.8 \pm 0.1$ μM | approx. 0.001 | approx. 40,000 |

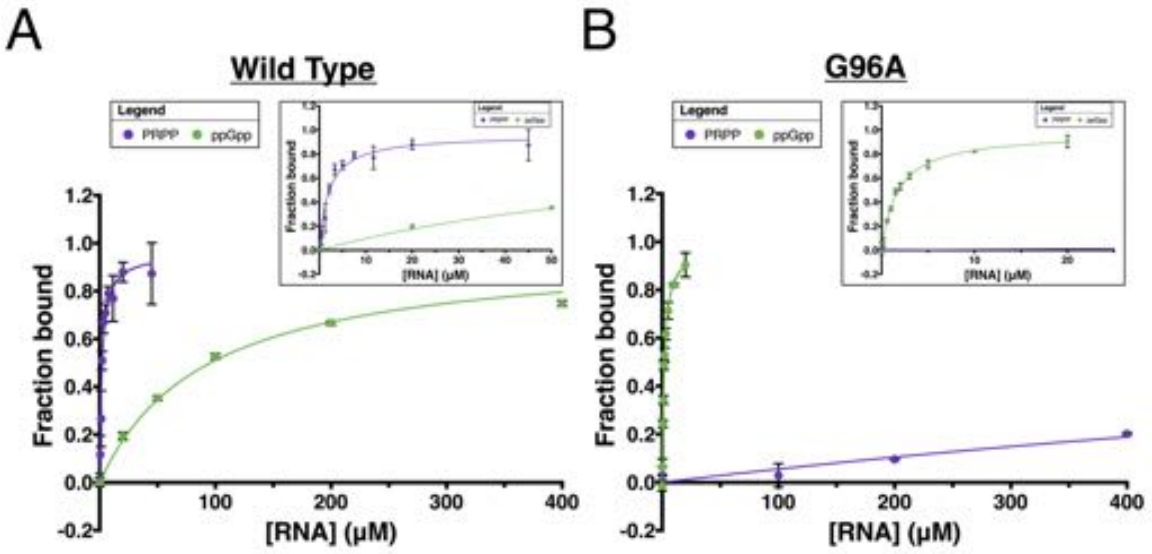


Figure 3.6: Data from equilibrium dialysis experiments and fits used to calculate dissociation constants. All experiments were performed in triplicate and the standard deviation for each data point is shown using black error bars. (A) wild type RNA binding to PRPP (purple) and ppGpp (green). The ppGpp complex did not quite reach saturation, so the $B_{max}$ was constrained to equal the $B_{max}$ value from the fit for ppGpp binding to the G96A mutant. (B) G96A mutant binding to PRPP (purple) and ppGpp (green). PRPP reached just 20% bound at 400 μM RNA, so the $B_{max}$ was constrained to equal the Bmax value from the fit for PRPP binding to wild type RNA. All floating $B_{max}$ values were within error of 1.0, as expected.

reagent used for G96A lacks barium, which was the most prominent divalent metal ion in the wild type crystallization condition. Potassium chloride, sodium chloride, and magnesium chloride were present in the crystallization drops. $K^+$ and $Mg^{2+}$ ions are observed in the mutant crystal structure. The best mutant crystal diffracted to a resolution of 3.1 Å and its structure was solved by molecular replacement using chain A of the PRPP riboswitch as an initial model. The asymmetric unit contained four aptamer molecules. Molecular replacement and refinement revealed robust density for the electron-dense pyrophosphate groups of ppGpp as well as its guanine base. In the initial solution and throughout refinement, the quality of the electron density was worse in chain D compared to chains A-C. The model of chain D is consistent with that of chains A-C, but is excluded from discussion in the text.

Overall, the architecture of the G96A mutant is very similar to that of the wild-type aptamer (Figure 3.2.1D). Notably, the $2F_O$-$F_C$ map generated directly by molecular replacement showed no electron density in the former location of the ribose and phosphate of G96. Additional lack of electron density for the ribose of G95 and the phosphate of G97 immediately suggested that the G96A mutation caused major conformational rearrangement in this region.

The ppGpp ligand has two pyrophosphate groups, meaning there is some symmetry to the ligand. The locations of the 5′ and 3′ pyrophosphate groups of ppGpp are easily inferred from the available electron density data. The electron density for the guanine base, parts of the sugar, and the 3′ pyrophosphate is high, while predictably lesser electron density surrounds the 4′ and 5′ carbons (Figure 3.7). Positioning of ppGpp such that the region of lower density surrounds the 4′ and 5′ ribose carbons results in the 5′

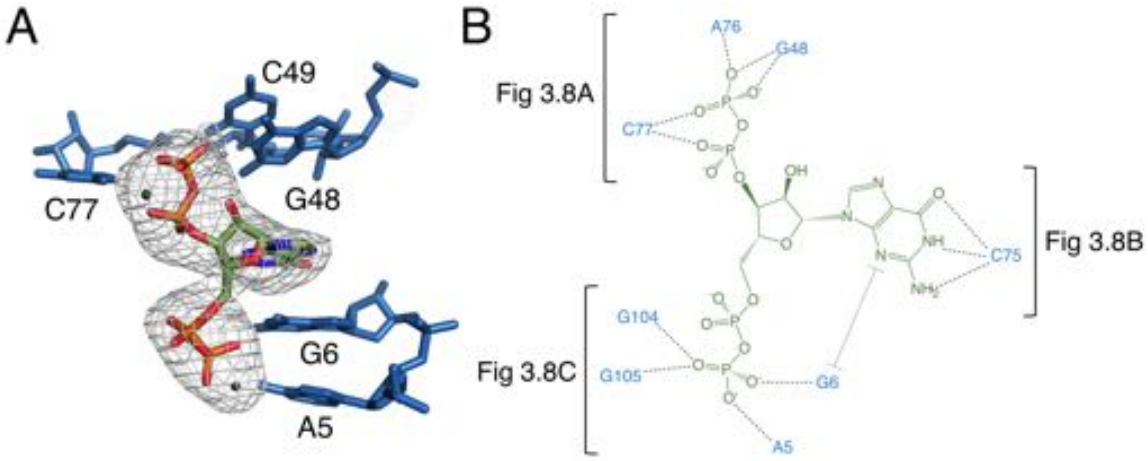


Figure 3.7: The binding pocket of the G96A mutant in complex with ppGpp. A: Crystal structure of the ligand-binding site in chain A. ppGpp is depicted as sticks and colored by element with green carbons. Nucleotides are depicted as teal sticks. Metal ions are depicted as gray spheres. Individual nucleotides are labeled. An Fo-Fc map contoured at 3.0$\sigma$ is shown as a gray mesh. The map was calculated using an otherwise complete model lacking ppGpp and nearby metals. (B) Ligand interaction map. The map is colored essentially as in A. All RNA contacts to ppGpp are shown. Dashed gray lines indicate hydrogen bonds. The grey bracket indicates base stacking. Black brackets indicate interactions shown in individual panels of Figure 5.

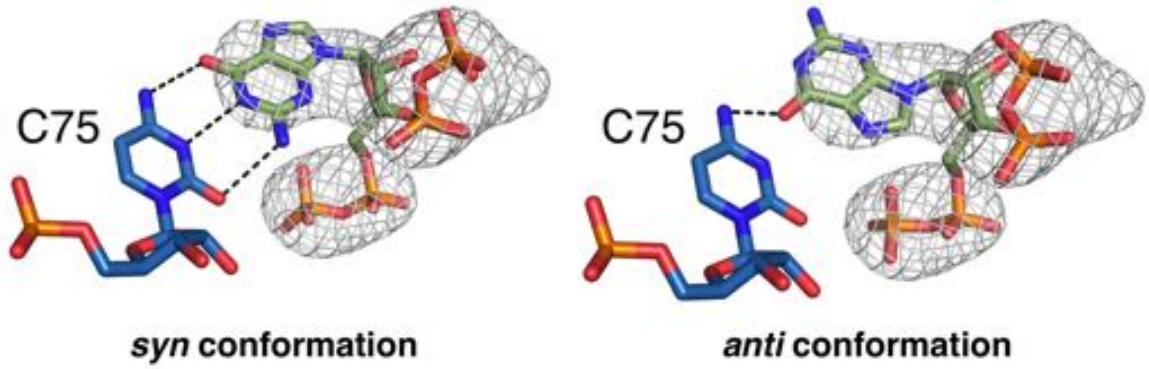


Figure 3.8: A comparison of ppGpp modeled in the *syn* conformation and the anti conformation. ppGpp has green carbons and is colored by element. The Fo-Fc map was generated using an otherwise complete model lacking ppGpp and is shown contoured at 3.0 $\sigma$ as a grey mesh.

pyrophosphate of ppGpp occupying the former position of the pyrophosphate of PRPP (Figures 3.4A and 3.7A). This means that the 5′ pyrophosphate is oriented down toward P0 in the ppGpp structure, but the 5-phosphate is oriented up toward P3 in the PRPP structure (Figures 3.2.1C/ 3.4A and 3.2.1D/ 3.7A). Several metal or water ions appear near the pyrophosphate moieties. These were first assigned as magnesium ions or water molecules, and subsequently assigned as potassium ions due to implausibly low B factors after refinement. The positioning of these entities is highly variable among the molecules in the asymmetric unit, suggesting that they do not make essential contributions to ligand recognition. We expect that the pyrophosphates are able to directly coordinate magnesium ions. This is not observed in the crystal structure, but this may simply be a consequence of the resolution only going to 3.1 Å. The divalent ions observed in the wild type PRPP structure were much more prominent, due to the comparatively very large electron density of barium ions relative to magnesium ions..

### 3.2.4 Assignment of ppGpp conformation in the G96A mutant

The guanine base of ppGpp is modeled in the *syn* conformation (Figure 3.7A). At 3.1 Åresolution, it is essential to inform this decision with the expected behavior of the chemical constituents in addition to the available electron density data. The shape of the electron density appears visibly more consistent with the *syn* conformation than the *anti* conformation (Figure 3.8). The chemical environment is also more plausible if the base is in the *syn* conformation. In the *syn* conformation, the guanine base of the ligand forms three hydrogen bonds with C75 in a Watson-Crick base pair. In the anti conformation, the Hoogsteen face of the guanine base would form two hydrogen bonds with the Watson-Crick face of C75.

The Hoogsteen/Watson-Crick pair would rely on protonation of the N3 of C75 (pKa = 4.45 in solution) [148]. Refinement of the ligand in the anti conformation tended to yield steric clashes or very short hydrogen bonds between the O6 of ppGpp and the N4 of C75, while simultaneously yielding unusually long hydrogen bonds ($> 3.5$ ) between the N7 of ppGpp and the N3 of C75. Modeling a Watson-Crick base pair (*syn* conformation) is consistent with a recent study showing that the equivalent of a C75U mutant in a native ppGpp riboswitch confers specificity to adenine-containing ligands over guanine-containing ligands (Sherlock et al., in press). The *syn* conformation of ppGpp has been previously observed in a 2.0 ÅX-ray crystal structure of an *E. coli* lysine decarboxylase, LcdI, which is inhibited by ppGpp [149]. The asymmetric unit of the LcdI crystal structure is a pentamer containing five ppGpp binding sites (one at each monomer-monomer interface), all of which are in the *syn* conformation. Finally, a structural overlay of the wild-type and G96A structures at C75 shows that in the *syn* conformation, the base of ppGpp in the G96A structure occupies the same position as the base of G96 in the wild type structure (Figure 3.7C). Together, this body of evidence strongly supports modeling the guanosine moiety of ppGpp in the *syn* conformation.

### 3.2.5 The guanine base and pyrophosphate groups of ppGpp are heavily recognized by the mutant aptamer

The 3′ pyrophosphate of ppGpp consistently sits in a pocket lined with hydrogen bond donors (Figure 3.9A). The N4 of C77, the N1 and N2 of G48, and the 2′OH of A76 are all seen making hydrogen bonds to phosphate oxygens. While the position of the 5′-pyrophosphate of ppGpp is relatively invariable, the 3′-pyrophosphate occupies a slightly

different position in each molecule of the asymmetric unit. Consistent with this model, the 3′ pyrophosphate atoms have slightly higher B factors than the rest of the ligand (about 138 $Å^2$ for the 3′ pyrophosphate compared to about 119 $Å^2$ for the 5′ pyrophosphate). In chain A, the 3′-β-phosphate has one oxygen that accepts a hydrogen bond from the N1 of G48, a second oxygen that accepts a hydrogen bond from the N4 of C77, and a third, unrecognized oxygen. The recognition strategy is slightly different for chains B and C.

The guanine base of ppGpp is buried in the RNA and is a focal point of ligand recognition. In the PRPP aptamer, the highly conserved C75 forms a Watson-Crick base pair with G96. The G96A mutant ppGpp aptamer recognizes its ligand through a similar Watson-Crick base pair between the G of ppGpp and C75 (Figure 3.9B). The guanine base of the ligand is also recognized via stacking with G6. Such extensive recognition of ppGpp's nucleobase suggests a likely mechanism for the mutant's observed discrimination for ppGpp over PRPP. In the native PRPP aptamer, C75 is in the same location near the binding pocket, poised to form this interaction with ppGpp. However, the highly conserved G96 is also available to form this base pair and its spatial proximity to C75 raises its effective concentration, making it potentially able to outcompete ppGpp for this base pairing interaction. This model is consistent with the observation of low-affinity ppGpp binding ($K_d$ = 91 $\pm$ 3 μM) in the wild type PRPP aptamer and explains why a single mutation at position 96 renders this aptamer capable of recognizing ppGpp with high affinity. The ribose is not recognized by the aptamer, leaving the guanine base and pyrophosphates as the major points of recognition.

Recognition of the 5′-pyrophosphate of ppGpp is extensive; its phosphate oxygens accept several hydrogen bonds from amino groups of conserved nucleobases (Fig-

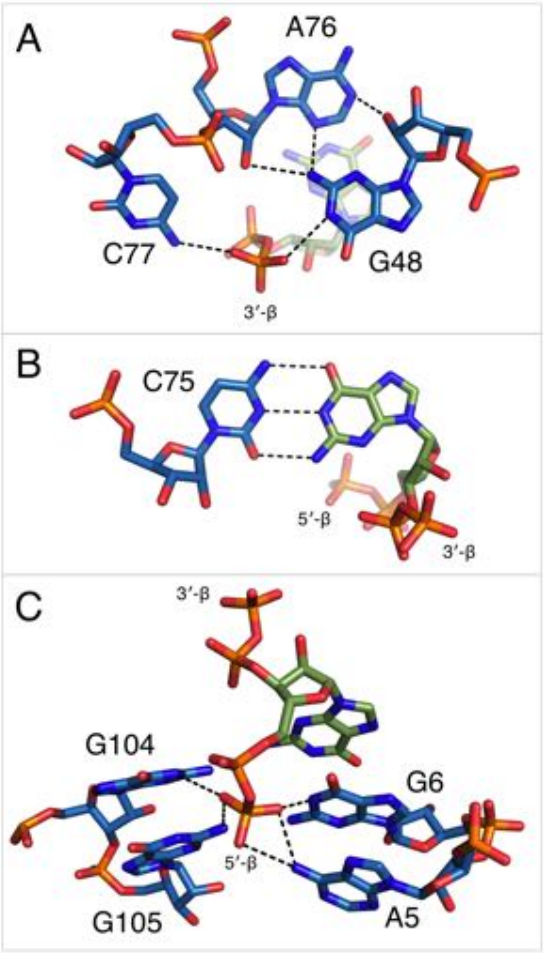


Figure 3.9: Notable contacts to the ppGpp ligand. ppGpp is depicted as sticks with green carbons and is colored by element. RNA is depicted as sticks with teal carbons and is colored by element. Dashed black lines indicate hydrogen bonds. (A) hydrogen bonds donated from amino groups in C77 and G48 to the 3′-β-phosphate of ppGpp, as well as the hydrogen bond network that constructs this part of the binding pocket. The 2′OH of A76 is close enough in chains B and C to form an additional hydrogen bond to the 3′-β-phosphate. B: The Watson-Crick base pair between C75 in the RNA and the ppGpp ligand. B: hydrogen bonds donated from amino groups in A5, G6, G104, and G105 to the 5′-β-phosphate of ppGpp.

ure 3.9C). The 5′-β-phosphate has three oxygen atoms that can accept hydrogen bonds from the aptamer. One of these oxygens accepts a hydrogen bond from the N6 group of A5. The second oxygen can accept hydrogen bonds from the N6 of A5 and N1 and N2 of G6, although it is not expected that these would all form simultaneously. The third oxygen can accept hydrogen bonds from the N1 and N2 of G105 and the N1 of G104. As with the previous oxygen, it is not expected that these would all form simultaneously. The 5′-α-phosphate appears to be unrecognized, consistent with its similar position to the poorly recognized α phosphate of PRPP in the native structure.

Nucleotide A74 appears to play a conserved structural role in the PRPP and ppGpp aptamers. In all three *ykkC* classes, it forms a noncanonical base pair with G6, which directly contacts PRPP and ppGpp, suggesting that it plays a role in positioning G6 (Figure 3.10). In the guanidine-I riboswitch, this nucleotide is not conserved. However, A6 in the guanidine-I crystal structure flips out to form the same non-canonical base pair with A68 (equivalent to A74) that is observed in the present study [80]. The lack of conservation at this position does not support a role in guanidine recognition, but this conserved interaction is observed in all three aptamers [40].

The guanidine-I and PRPP *ykkC* aptamers each have an S-turn motif in the P3 helix. In the guanidine-I aptamer, the orientation of G88 is reversed relative to its stacking partners and G89 flips out of the helix. These are classic features of the S-turn. The guanidine-I riboswitch also possesses a cross-strand purine stack, a characteristic backbone kink on the opposite strand from the S-turn, and stabilizing hydrogen bonds, all of which were first observed in the S-turn of the conserved sarcin-ricin loop in the 23S rRNA [150]. In the PRPP riboswitch, a similar S-turn motif exists at the equivalent position

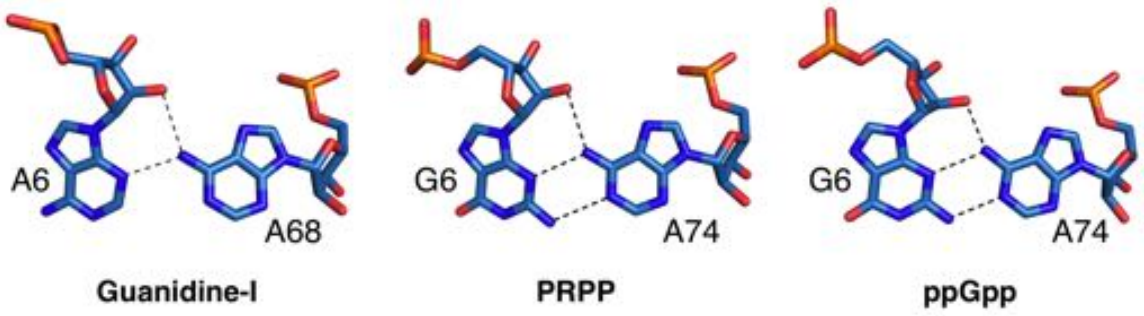


Figure 3.10: Conserved interactions between P1 and P3. Nucleotides are labeled and depicted as sticks and colored by element with teal carbons. Dashed lines indicate hydrogen bonds. Left: Nucleotides A6 and A68 of the guanidine aptamer. Middle: Nucleotides G6 and A74 of the PRPP aptamer. Right: Nucleotides G6 and A74 of the G96A mutant, a ppGpp aptamer.

(Figure 3.11B). Equivalent to G89 in guanidine-I, G96 flips out and base pairs with C75 while also hydrogen bonding to PRPP. Notably, the cross-strand purine stack is absent in the PRPP riboswitch, but other S-turn characteristics are preserved. Conversely, the S-turn motif is abolished in the G96A mutant, and no contacts are observed between A96 and other nucleotides. Even more significantly, G95 does not possess the reverse ribose orientation that defines an S-turn. Rather, this region resembles a standard A-form helix with a single nucleotide bulge. The guanine of ppGpp replaces the flipped out guanosine of the former S-turn motif (Figure 3.11A), revealing that the S-turn is a key center of functional plasticity in the *ykkC* RNAs.

## 3.3 Discussion

Taken together, the present structural and biochemical data shed light on the functional plasticity of RNA as a whole and of the *ykkC* motif in particular. Just as residue C49 was previously used to distinguish guanidine aptamers from other *ykkC* RNAs, here we show that G96 is the residue that differentiates PRPP and ppGpp aptamers. The sequence space of the *ykkC* motif is clearly amenable to binding dissimilar ligands. The existence of *ykkC* RNAs with other gene contexts and unknown ligand specificity further reinforces the diversity of functions that these RNAs achieve with very small variations in consensus sequence [40; 45], (Sherlock et al., in press). Three-dimensional structural models of the wild-type and G96A mutant aptamers reveal that the mechanism of specificity switching is recruitment of C75 as a primary effector of ligand recognition (Figure 3.11A). The presence or absence of the S-turn motif governs whether an RNA base or the ppGpp base can pair with C75, and therefore controls the specificity of the aptamer.

The wild-type aptamer featured in the current study binds PRPP at a location very near, but distinct from the binding pocket of the guanidine-I riboswitch. The P0 region, which is not present in the guanidine-I riboswitch, recognizes a portion of the larger ligand; metal ion M3 binds in the location where its parent motif binds guanidine (Figure 3.11; see also Figure 3.4). In the S-turn of the sarcin-ricin loop, the bulged G re-inserts into its helix to form a base triple. In an overlay of the S-turns of the sarcin-ricin loop, the guanidine-I riboswitch, and the PRPP riboswitch, the guanidino group of the bulged guanosine in the sarcin-ricin loop overlays well with the guanidinium cation and both overlay with metal M3 in the PRPP riboswitch (Figure 3.12A,B) [150]. The common binding site of M3

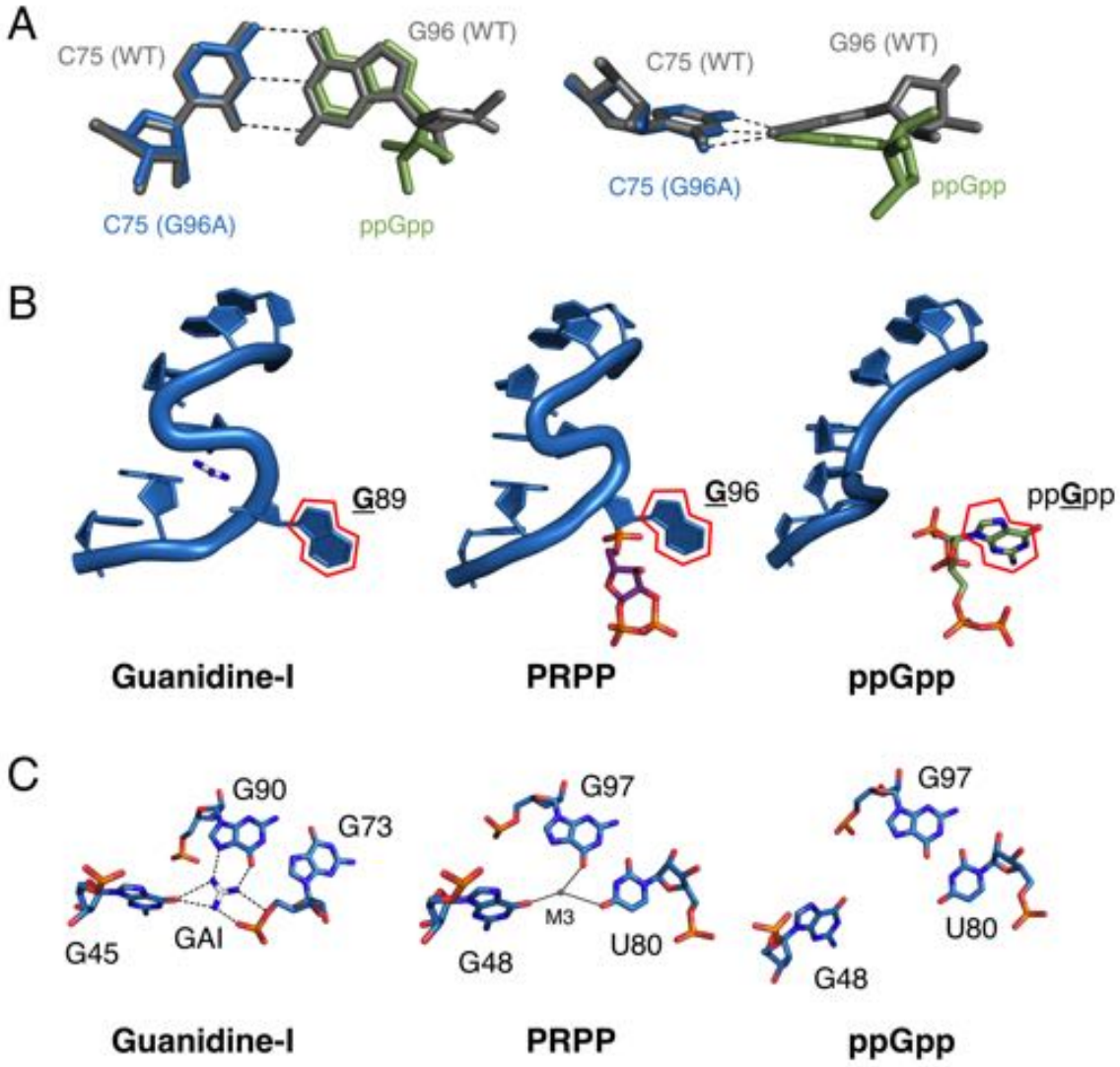


Figure 3.11: Structural comparison of the guanidine-I, PRPP, and ppGpp riboswitch aptamers. (A) Watson-Crick base pairs with C75 in the wild-type and G96A aptamers overlaid. Wild-type and G96A RNA is shown as gray and teal sticks, respectively. The base and ribose of ppGpp are shown as green sticks. Hydrogen bonds are shown as dashed lines. Left: face-on view of the preserved base pair. Right: edge-on view of the same interaction. (B) The S-turn motif in the guanidine-I and PRPP aptamers and the equivalent position in the ppGpp aptamer. The RNA is depicted as a teal cartoon. Guanidine, PRPP, and ppGpp are colored by element with white, purple, and green carbons, respectively. A red outline showcases the position of the conserved guanine base. (C) Nucleotides in the guanidine or M3 binding site, or the equivalent site in the ppGpp aptamer. The RNA is colored by element with teal carbons. Guanidine is colored by element with white carbons. Black dashes indicate hydrogen bonds. Solid black lines indicate coordination to a metal ion.

in the PRPP riboswitch and guanidine in the guanidine riboswitch is a possible case of molecular exaptation, suggesting that structured RNAs are functionally versatile and can readily adapt to new selection pressures. However, the natural history of the evolution of these two related aptamers remains uncertain.

The present structural data shed additional light on the potential mechanism of switching in tandem guanine-PRPP aptamers [45]. The PRPP riboswitch (an ON switch), is often found immediately downstream of a guanine riboswitch (an OFF switch), in an IMPLY two-input logic gate (Figure 3.13). In these tandem systems, transcription proceeds in all cases except when guanine is present and PRPP is not [45]. This suggests that PRPP binding disrupts formation of the guanine aptamer, allowing transcription to proceed when both ligands are present (Figure 3.13D). The *T. mathranii* PRPP aptamer studied in the present work is part of one of these tandem aptamer systems. Its predicted secondary structure shows that formation of the P0 stem of the PRPP aptamer and the P1 stem of the guanine aptamer are mutually exclusive. The present data reveal that the 5′ tail of the PRPP aptamer participates in P0 and plays a central role in PRPP recognition. In the proposed model, PRPP binding stabilizes P0 and disrupts the P1 helix of the guanine aptamer. The IMPLY character of this two-input gate may depend on the relative stabilities of the two helices, which in turn suggests that alternative logic gates could be constructed through mutation of P0 or P1. The ppGpp and T-box riboswitches are also often found in tandem. In contrast with the PRPP/guanine tandem system, the ppGpp and T-box riboswitches each maintain their own expression platform, suggesting that they fold independently. This is consistent with the AND behavior of this logic gate and the unimportance of the order of the two aptamers within the molecular circuit.

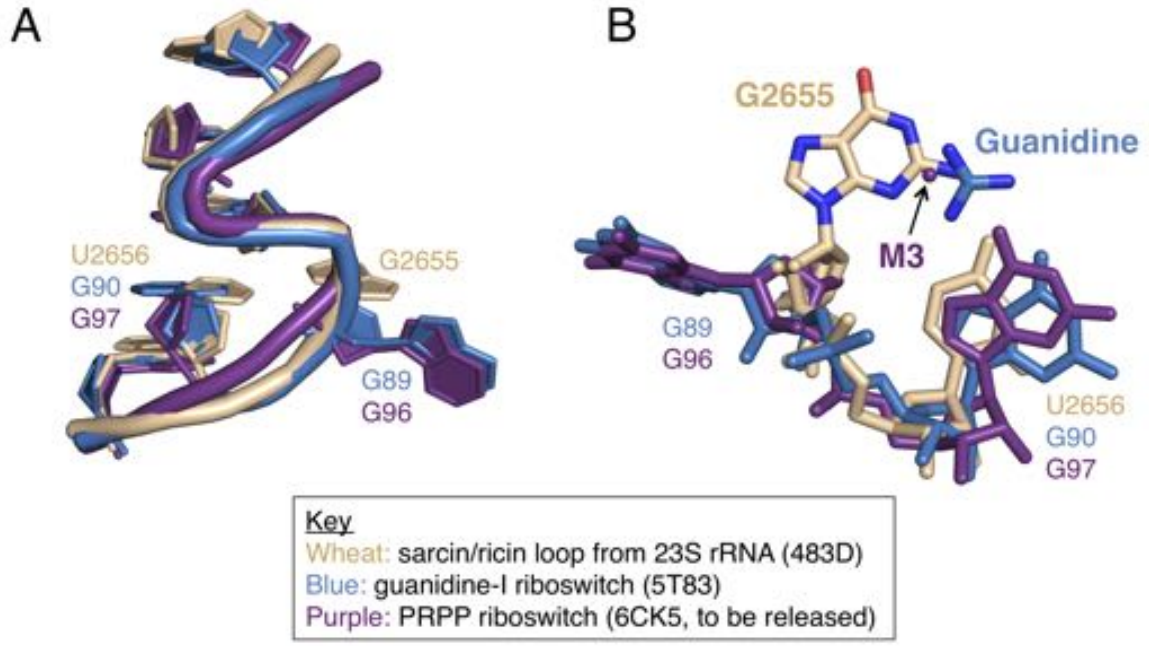


Figure 3.12: (A) Overlay of the S-turn found in the sarcin/ricin domain from *E. coli* 23S rRNA (PDB ID 483D, shown in wheat) with the S-turns in the guanidine-I riboswitch (5T83, shown in blue) and the PRPP riboswitch (6CK5, shown in purple) [150; 80]. (B) Bottom-up view of the overlay in part A showing overlap between the guanidino-moiety of G2655, guanidine, and M3.

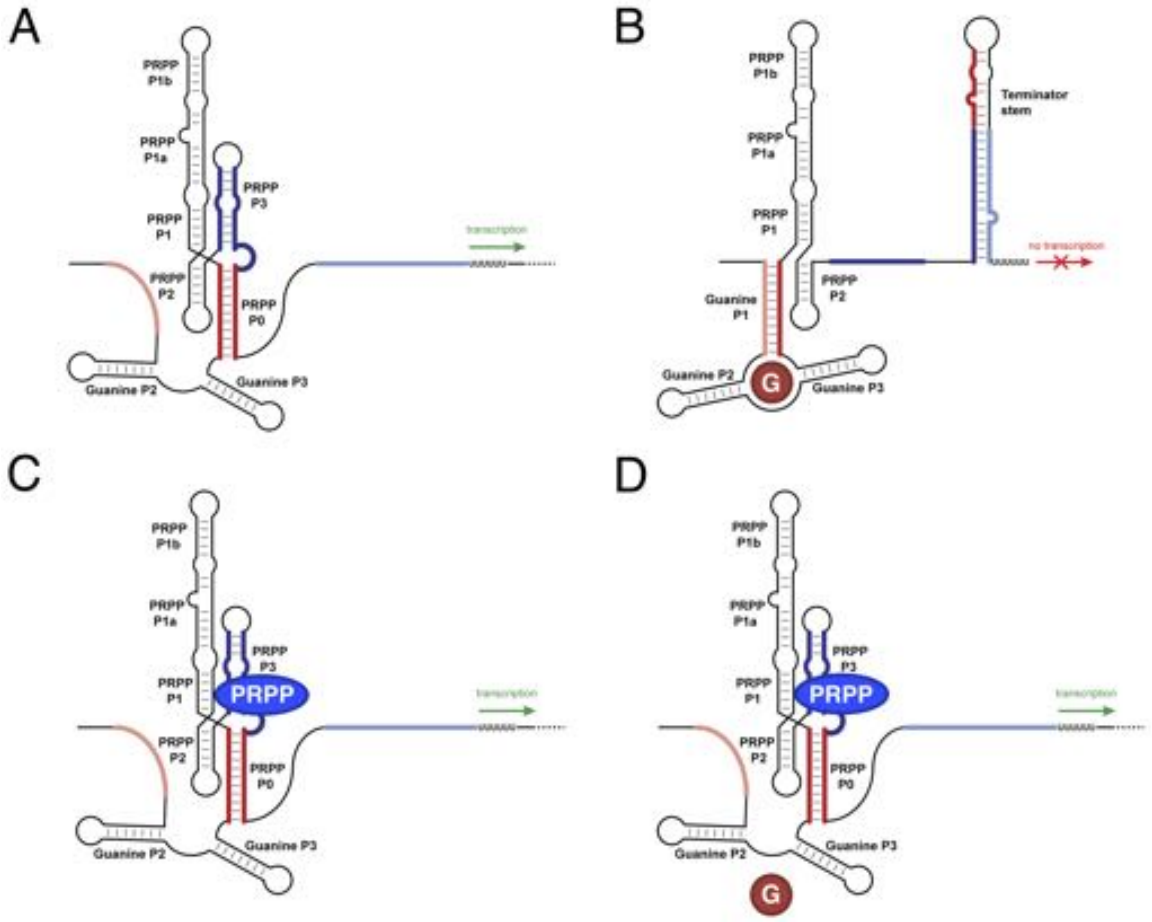


Figure 3.13: Predicted model of switching in the tandem guanine-PRPP riboswitch from *T. mathranii*. This model is informed by the crystal structures presented here as well as *in vitro* transcription termination data collected by Sherlock et al. on a tandem guanine-PRPP riboswitch from *Bacillus megaterium*. The secondary structure of the RNA is represented as black lines. PRPP aptamer P0 nucleotides are colored dark red. The competing nucleotides in the guanine aptamer is colored pink. PRPP P3 nucleotides are colored dark blue. The competing nucleotides in the terminator stem are colored light blue. PRPP is represented as a blue ellipse, and guanine is represented as a red circle. (A) Predicted structure in the absence of PRPP and guanine. (B) Predicted structure in the presence of guanine. (C) Predicted structure in the presence of PRPP. (D) Predicted structure in the presence of PRPP and guanine. In this model, the presence of PRPP stabilizes P0 and destabilizes the guanine aptamer domain.

The observed ability of the *ykkC* scaffold to adapt to new selection pressures via mutation of a few key residues is reminiscent of other accounts of functional plasticity in both proteins and RNAs. In a clinically relevant contemporary example, β-lactamases evolve to expand their catalytic repertoire through mutations in a flexible loop. These mutations preserve the overall architecture of the protein while enabling it to metabolize new antibiotics, contributing to the worldwide threat of antibiotic resistance [75; 76; 77; 78]. The ease of repurposing protein scaffolds for the development of new catalysts has also been exploited in the design of novel enzymes including a Diels-Alderase [151]. In a related RNA example, the *Tetrahymena* ribozyme scaffold supports catalytic activities including self-cleavage, RNA polymerization, and peptide bond hydrolysis, though this is likely due to the placing the substrates into physical proximity with each other [10; 152; 153; 154]. Natural riboswitch aptamers used as starting points in directed evolution experiments switch specificity, but maintain their overall fold [27]. The existence of riboswitch variants that use the same scaffold but bind slightly different ligands, including the adenine/guanidine and cyclic-di-GMP/cyclic-GMP-AMP riboswitch classes, has previously hinted at the adaptability of RNA elements [28; 34; 59; 37; 108; 60; 107; 106]. However, the present finding that a single conservative point mutation in the PRPP aptamer can dramatically alter both ligand specificity and tertiary structure reveals a striking example of RNA plasticity and speaks to the macromolecular evolvability of RNA.

A final key observation in this study is the direct visualization that an RNA element has evolved to specifically recognize PRPP. PRPP is a central metabolite, and likely has played that role since before the metabolic pathways of life's last universal common ancestor (LUCA) were fully developed [155]. It is possible that PRPP was used for the

synthesis of nucleotide precursors on the prebiotic Earth [156]. The finding that an extant RNA specifically recognizes PRPP lends credence to the hypothesis that RNA elements may have been capable of recognizing PRPP before the advent of coded protein synthesis.

# 4 Structure of the guanidine-II riboswitch aptamer

This section is adapted from Reiss et al., 2017, an article published in the journal *RNA* [157].

## 4.1 Background

The guanidine-II riboswitch, originally dubbed the mini-*ykkC* RNA, was first identified through bioinformatics in 2007 [91]. The RNA was named mini-*ykkC* because it controls a similar set of genes as the *ykkC* RNA motif, but is much smaller and simpler. Both riboswitches regulate the expression of genes that encode guanidine carboxylases and SugE-like efflux proteins that transport guanidine out of the cell. The *ykkC* and mini-*ykkC* riboswitches are likely employed by bacteria to de-toxify the cell in the presence of high guanidine concentrations.

After identifying that guanidine was the ligand for the *ykkC* RNA, the Breaker lab tested mini-*ykkC* for guanidine binding as well, due to the similarity in gene associations. The mini-*ykkC* RNA binds to guanidine with similar affinity to the guanidine-I riboswitch and it was renamed the guanidine-II riboswitch [40; 158]. The class II motif is made up of two short hairpins connected by a linker that is between 7 and 40 nucleotides in length. Each hairpin is capped by an identically conserved ACGR tetraloop and there are two conserved G-C base pairs at the base of each tetraloop, making for a total of 16 conserved nucleotides in each riboswitch (8 in each hairpin). In contrast, the guanidine-I aptamer is substantially larger and contains 35 nucleotides that are $\geq$97% conserved. Also unlike the guanidine-I riboswitch, guanidine-II was shown to bind guanidine in a

cooperative manner by in-line probing, suggesting multiple instances of binding [158]. Based on the consensus sequence, it was proposed that the two ACGR tetraloops each bind a molecule of guanidine and that the hairpins might interact with each other [158]. Consistent with this, in-line probing data showed guanidine-dependent modulation occurs in both tetraloops [158]. The dramatic differences between the guanidine-I and guanidine-II riboswitches suggest that the guanidine-II riboswitch recognizes its ligand in an entirely different way from the guanidine-I riboswitch. In addition to adopting distinct RNA folds in the aptamer domain, they also regulate gene expression by two different mechanisms. Based on bioinformatic data, the guanidine-II riboswitch is a translationally-controlled on-switch, while guanidine-I is a transcriptionally-controlled on-switch [40; 158; 91].

Guanidine riboswitches face the challenge of recognizing a small molecule in a specific manner despite the presence of other guanidino-containing metabolites in the cell that are present at micromolar concentrations, such as arginine and agmatine (Bennett et al., 2009; Caldara et al., 2008; Hamana, 1996). In addition, guanidine riboswitches must select against urea, which occurs at high concentrations and is similar in size and shape, differing only in having a carbonyl group where guanidine has an amino group. The guanidine-I riboswitch accomplishes this by recognizing every possible face of the guanidinium cation [80; 159]. In addition to hydrogen bonding, the class I riboswitch utilizes $\pi$-cation and ionic interactions.

The guanidine-II riboswitch was suspected to recognize two molecules of guanidine in a cooperative manner based on in-line probing data, suggesting that the tetraloops dimerize. To determine how the guanidine-II riboswitch recognizes guanidine and selects against similar compounds, and to determine the structural basis for tetraloop dimerization,

we set out to determine the crystal structure of the riboswitch in complex with its ligand. Here, we present the crystal structure of a dimerized P2 stemloop of the guanidine-II riboswitch from *Pseudomonas aeruginosa* determined at 1.57 Åresolution.

## 4.2 Results

### 4.2.1 Structure determination of the guanidine-II riboswitch aptamer

Biochemical analysis of full-length guanidine-II riboswitch constructs suggested that the RNA forms large, higher order structures, likely due to dimerization of the hairpins in trans (data not shown). Therefore, we used singlet hairpin constructs for crystallography. A previously characterized singlet stem loop construct shows guanidine-dependent modulation by in-line probing, likely due to the formation of homodimers [158]. Because the conserved tetraloops are identical in the P1 and P2 helix of most guanidine-II riboswitches, we expect that the conserved regions of the homodimer structure will adopt the same structure as a physiologically relevant heterodimer between P1 and P2. We targeted the 16-nucleotide P2 stem of the guanidine-II riboswitch aptamer from *P. aeruginosa* (Figure 4.1B), which controls a gene annotated as a SugE efflux protein. A member of the SugE family controlled by the guanidine-I riboswitch was previously shown to associate specifically with guanidine, suggesting that it may act to export guanidine out of the cell to avoid guanidine toxicity [40]. A non-conserved U in the sequence was modified to a 5-bromouridine and the structure was determined by SAD phasing. Guanidinium was clearly visible in the unbiased electron density maps. An Fo-Fc map contoured at 5$\sigma$ was calculated using phases from a model in which guanidinium was omitted, which reveals significant guanidinium-shaped density that supports our assignment of the ligand model (Figure 4.2A). X-ray statistics are shown in Table 4.1.

The crystal structure contains four P2 hairpins in the asymmetric unit (chains A, B, C, and D in the structure). The conserved nucleotides 5-12 are highly similar between

Table 4.1: X-ray statistics for the guanidine-II aptamer

| | **P2 hairpin with 5-bromoU (SAD)** | **P2 hairpin (native)** |
|---|---|---|
| **PDB** | **5VJB** | **5VJ9** |
| **Data collection:** | | |
| Beamline | 24-ID-C at APS | 24-ID-C at APS |
| Space group | P 21 21 21 | P 21 21 21 |
| **Unit cell:** | | |
| a, b, c (Å) | 50.2, 60.5, 71.9 | 50.3, 60.7, 72.3 |
| $\alpha$, $\beta$, $\gamma$ (°) | 90, 90, 90 | 90, 90, 90 |
| | | |
| Wavelength (Å) | 0.9193 | 0.9193 |
| Resolution (Å) | 40.00-2.10 (2.14-2.10) | 40.00-1.57 (1.60-1.57) |
| $R_{merge}$ | 0.068 (1.058) | 0.052 (1.052) |
| I/$\sigma$I | 31.3 (1.6) | 46.3 (1.7) |
| $CC_{1/2}$ in highest resolution shell | 0.552 | 0.565 |
| CC* in highest resolution shell | 0.843 | 0.850 |
| Completeness (%) | 99.9% (100.0%) | 99.8% (98.8%) |
| Redundancy | 6.8 (7.0) | 11.6 (6.9) |
| Total reflections | 172716 | 361260 |
| Unique reflections | 25445 | 31438 |
| **Refinement:** | | |
| Resolution (Å) | 40.00-2.10 | 40.00-1.57 |
| No. reflections | 13040 | 29787 |
| $R_{work}/R_{free}$ | 0.22/0.24 (0.32/0.39) | 0.19/0.23 (0.40/0.42) |
| **No. of atoms:** | | |
| Total | 1456 | 1727 |
| RNA | 1372 | 1368 |
| Ligand | 16 | 16 |
| Cations | 57 | 62 |
| Water | 11 | 281 |
| **B-factors** | | |
| Overall (Å$^2$) | 42.1 | 32.4 |
| RNA (Å$^2$) | 41.2 | 29.4 |
| Ligand (Å$^2$) | 37.3 | 26.1 |
| Cations (Å$^2$) | 66.4 | 49.8 |
| Water (Å$^2$) | 36.1 | 43.1 |
| **Root-mean-square deviations** | | |
| Bond lengths (Å) | 0.012 | 0.013 |
| Bond angles (°) | 1.995 | 1.841 |

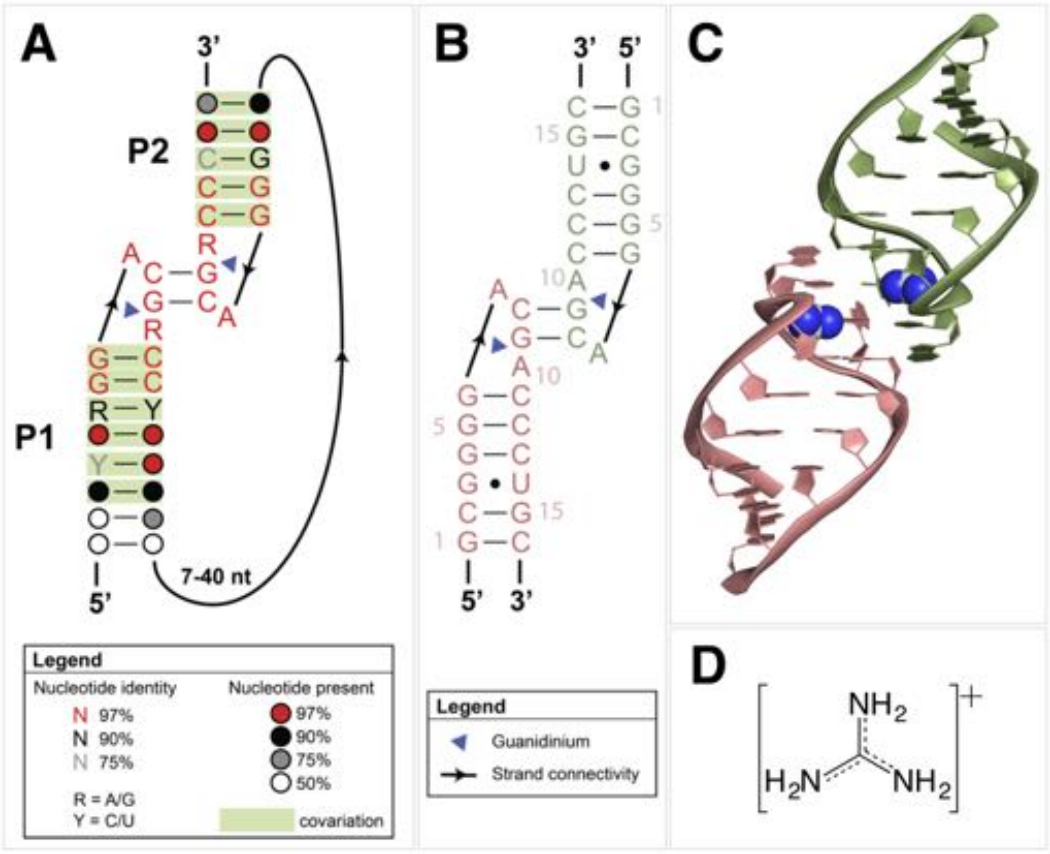


Figure 4.1: Overall structure of the guanidine-II riboswitch P2 hairpin from P. aeruginosa. (A) Updated consensus sequence to reflect secondary structure observed in the crystal structure. (B) Secondary structure of the crystal construct used. The pink and green hairpins represent chain A and chain B of the asymmetric unit and are equivalent in sequence. Blue triangles represent guanidinium. (C) Overall crystal structure of the guanidine II riboswitch P2 hairpin dimer from P. aeruginosa. All figures are generated using chains A and B and all distances are calculated using chain A. (D) Chemical structure of the guanidinium cation.

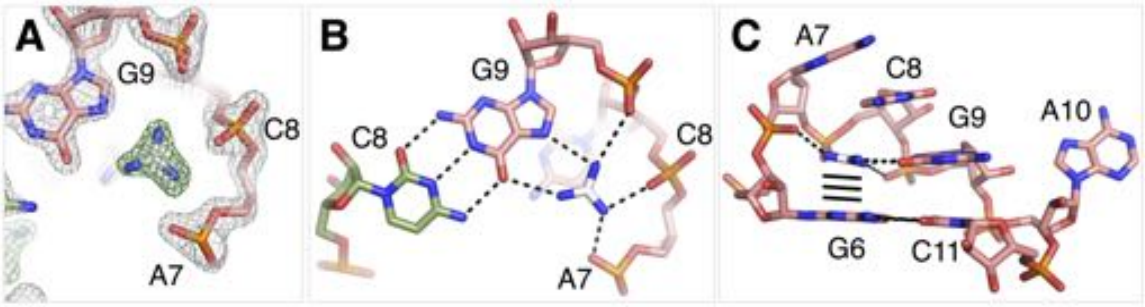


Figure 4.2: Guanidinium binds within the conserved ACGR tetraloop. (A) 2Fo-Fc map of binding pocket contoured at 2$\sigma$ (grey) and Fo-Fc map of binding pocket contoured at 5$\sigma$ (green), where the input model lacked guanidinium. Pink sticks represent chain A and white sticks represent guanidinium. (B) Hydrogen bonds donated from guanidinium to the riboswitch. Pink sticks represent chain A, green sticks represent chain B, and white sticks represent guanidinium. (C) The ACGA tetraloop and G-C base pair below. Hydrogen bonds from guanidinium to RNA are represented by grey dashes. Possible cation-pi interactions are represented by black dashes. Color scheme is the same as in Figure 2A.

the four molecules (average rmsd= 0.39 Å). Chains A+B and chains C+D dimerize through their tetraloops and this appears to be the physiologically relevant interface. The two dimers form crystal-packing interactions with each other.

### 4.2.2 Guanidinium recognition

Two stem loops form a dimer via a kissing-loop interaction between the two ACGA tetraloops. The two hairpins are rotated 180 apart from each other and the interaction occurs strictly through the tetraloops and the cytosine in the first G-C base pair below the tetraloops. The tetraloop of each hairpin also creates a binding pocket for the guanidinium cation (Figure 4.1C). Although the guanidine-I and guanidine-II riboswitches adopt completely different RNA folds, they use remarkably similar strategies to recognize the ligand. Guanidinium has a pKa of 13.6 [93], which means that under physiological conditions it is present almost exclusively as a cation and has six potential hydrogen bond donors (Figure 4.1D). In the guanidine-II riboswitch, five of the six positions make direct contact to the RNA. One edge of guanidinium donates two hydrogen bonds to the Hoogsteen face of G9, a highly-conserved nucleotide located in the tetraloop. One hydrogen bond is donated to the O6 carbonyl oxygen (3.1 Å) and one is donated to the N7 imino (3.1 Å) (Figure 4.2B). On the other two edges of guanidinium, three hydrogen bonds are donated to non-bridging phosphate oxygens, specifically the pro-$S_P$ of A7, pro-$R_P$ of C8, and pro-$R_P$ of G9 (Figure 4.2B). It is likely that the negatively charged phosphate groups in the binding pocket also stabilize guanidinium cation binding through ionic effects. The ionic interactions likely contribute to selection of guanidinium over urea, which is uncharged. The last hydrogen in guanidinium is not recognized by the riboswitch and appears to be solvent exposed. In one set

of the two dimers in the asymmetric unit (chains C and D), there is a solvent bridge comprised of two water molecules that hydrogen bond to the two guanidiniums and to each other. Ligand-dependent RNA structure modulation is observed with methylguanidine and aminoguanidine as well as a few other small, guanidino-containing compounds [158]. The guanidine-II riboswitch is somewhat more promiscuous than the guanidine-I riboswitch, but modulation by these other compounds is still weaker than the modulation caused by guanidine [158]. The solvent exposed region on each of the two binding pockets are oriented towards each other, suggesting that the riboswitch may be selecting against larger guanidine-containing molecules, like arginine and agmatine, by sterically preventing the two from binding simultaneously. The guanidinium is also sandwiched above and below by conserved nucleotides G6 and C8 (Figure 4.2C) and is potentially forming $\pi$-cation interactions, in which the positively charged ligand is interacting with the electron-rich $\pi$ orbitals of RNA bases [160; 94]. Based on a published analysis of $\pi$-cation interactions found in 282 crystal structures containing non-redundant protein-RNA interfaces, interactions between the guanidine moiety of arginine and guanine bases are very common [96]. Cytosine is sometimes observed forming $\pi$-cation interactions with guanidino moieties in arginine, but much less commonly compared to guanine. Empirical energy calculations between cytosine and the guanidino moiety of arginine suggest that this interaction is unfavorable on average [96]. Nucleotide C8 may make little or no direct contribution to ligand binding energetics. It is conserved for its role in the inter-hairpin base pairing at the dimerization interface. $\pi$-cation interactions are another way that the riboswitch can select for the guanidinium cation over urea, since urea does not have the positive charge needed to interact with the $\pi$ system of a base.

### 4.2.3 Structural aspects of the dimerization interface

The main structural motif that forms the dimerization interface is a kissing-loop interaction between the two tetraloops. C8 and G9 form Watson Crick base pairing with the equivalent C8 and G9 of the dimerization partner (Figure 4.3A). The G9 nucleotide also forms bifurcated hydrogen bonds to guanidinium via the Hoogsteen face (Figure 4.2B). This kissing-loop interaction positions the two guandiniums within 10 Åof each other. In the tetraloop, there are two adenines, one on each side of the kissing loop nucleotides C8 and G9 (Figure 4.2C). Highly conserved nucleotide A7 stacks on top of C8 and forms two hydrogen bonds to the sugar edge of C11 of the dimerization partner (Figure 4.3B). Nucleotide A10, which is highly conserved as a purine (A or G) flips out from the tetraloop and forms stacking interactions with A10 of the other hairpin (Figure 4.3C). Guanidinium binding may result in a rearrangement of the tetraloop such that A7 and A10 are more accessible to form dimerization interactions. In this case, guanidinium binding would shift the equilibrium toward the dimeric state. Consistent with this model, in-line probing of a P1 hairpin from the *Gloeobacter violaceus* guanidine-II riboswitch shows that the equivalents of nucleotides A7 and A10 in the crystal structure modulate in a guanidine-dependent manner [158].

## 4.3 Discussion

### 4.3.1 Guanidinium recognition in class-I and class-II guanidine riboswitches

The guanidine-I and guanidine-II riboswitches adopt completely different RNA folds, but use similar strategies to recognize guanidinium (Figure 4.4A). In the guanidine-I riboswitch,

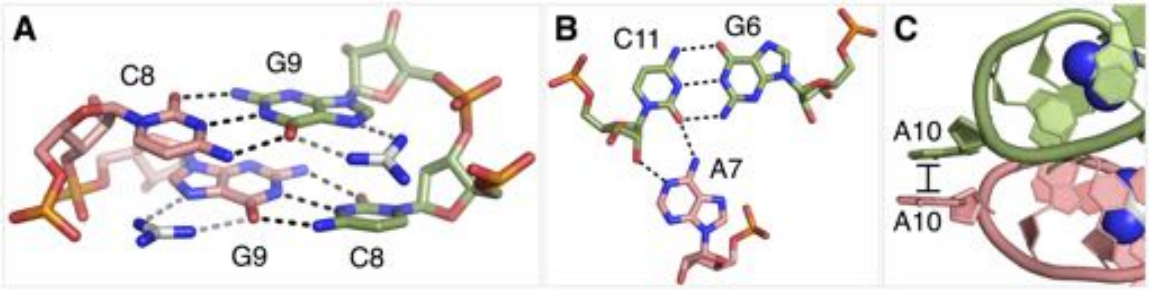


Figure 4.3: The two hairpins dimerize head-to-head through a kissing loop interaction. (A) The kissing loop interaction at the dimerization interface. Color scheme is the same as in Figure 2B. (B) Hydrogen bonds formed at the dimer interface between the Watson-Crick face of A7 in chain A and the sugar edge of C11 in chain B. (C) Stacking interactions formed by A10 in chain A and A10 in chain B.

one edge of guanidinium forms two hydrogen bonds with the Hoogsteen face of G90 in a bifurcated manner. The same hydrogen bonds are formed in the guanidine-II riboswitch, involving G9 (Figure 4.4B). Both riboswitches also use phosphate oxygens in the RNA backbone to recognize guanidinium. In the guanidine-I riboswitch, the guanidinium donates two hydrogen bonds to phosphate oxygens of G73. In the guanidine-II riboswitch, guanidinium forms hydrogen bonds to three non-bridging phosphate oxygens. In both classes of riboswitch, the proximity of the negatively charged phosphate groups stabilizes the binding of the guanidinium cation through ionic interactions (Figure 4.4B). Both riboswitches use guanine bases to form $\pi$-cation interactions with guanidinium.

The strategies used to recognize guanidinium by the two classes of riboswitch are the same, but the RNA context differs. In the class I guanidine riboswitch, the single binding site is created at the interface between two helical elements, P1a and P3. The class II guanidine riboswitch has two binding sites, one in each of the hairpins that dimerize upon guanidine binding. Each guanidinium ligand makes contacts to only a single hairpin, but the binding sites are located at the dimerization interface.

### 4.3.2 Proposed switching model in the full length aptamer

In a wild type guanidine-II aptamer, the P1 and P2 hairpins are separated by a linker that is 7 to 40 nucleotides in length (Figure 4.1A). In the crystal structure, the 3′-nucleotide of one hairpin and the 5′-nucleotide of the second hairpin are about 50 Åapart. This is the distance that must be spanned by the random linker between the two hairpins. Nucleotides in a splayed conformation extend a distance of approximately 7.2 Åper nucleotide, measuring from phosphate to phosphate. This provides enough length to cover the 50 Ådistance, even

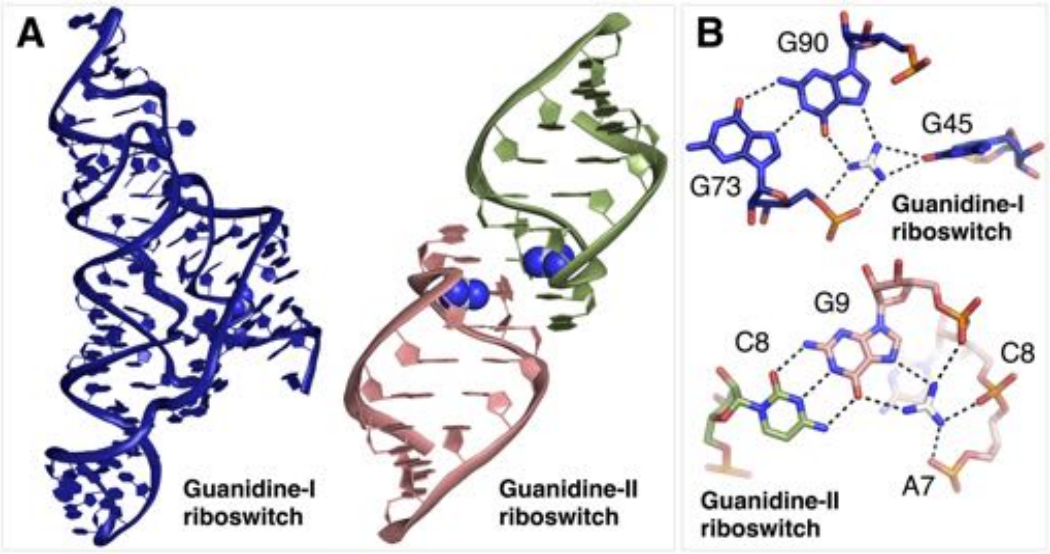


Figure 4.4: Side-by-side comparison of the guanidine-I riboswitch [80] and the guanidine-II riboswitch [157] crystal structures. (A) Ribbon diagrams of class I and class II guanidine riboswitch crystal structures. The guanidine-I riboswitch is in navy blue and the guanidine-II P2 hairpin dimer is shown in pink and green (one color for each hairpin monomer). Guanidinium is represented as spheres in white (carbon) and blue (nitrogen). (B) Stick models showing the guanidinium binding pockets in the class I and class II guanidine riboswitches. Overall colors are the same as Figure S1A with coloring by element.

for the shortest of linker regions.

The guanidine-II riboswitch is a translational on-switch. The structure of the two hairpins provides a model for how guanidine could regulate expression of the downstream message. In the case of this *P. aeruginosa* riboswitch, the Shine-Dalgarno sequence is sequestered by base-pairing with part of the linker region and nucleotides at the 5′ end of the P1 helix (Figure 4.5). In the guanidine-bound structure, head-to-head dimerization of P1 and P2 enforces the spatial separation of the Shine-Dalgarno sequence from the anti-Shine Dalgarno sequence. This would expose the Shine-Dalgarno sequence and promote translation in the presence of guanidine (Figure 4.5). The crystal structure does not allow us to distinguish between a conformational selection mechanism or an induced fit mechanism for switching. Gel shift data collected on the P1 hairpin from *P. aeruginosa* suggests that it is energetically plausible for guanidine binding to cause this sort of massive rearrangement in RNA helix formation. In this experiment, I showed that increasing concentrations of the *P. aeruginosa* P1 hairpin cause the RNA to eventually form duplex in the absence of ligand, due to the complementarity of the strands (Figure 4.6A,B). I am confident that, in the absence of ligand, the dimer observed in the gel shift experiment is an RNA duplex rather than a tetraloop-tetraloop dimer, since I solved the crystal structure of this RNA in the absence of guanidine to 2.75 Åresolution (Figure 4.6B). At 200 μM RNA, virtually all the RNA is in the duplex form (Figure 4.6A). Upon addition of guanidine, the resulting species appears as a smear in the native gel. This is consistent with the formation of a different dimeric form with a faster off-rate than the duplex. Importantly, the 10 mM guanidine added is far too low a concentration to cause significant denaturing effects. Also consistent

is the affinity of mini-*ykkC* RNAs for guanidine (mid to high micromolar), which is usually too weak an interaction to be observed cleanly by gel shift. I expect that the tetraloop-tetraloop dimer forms in the presence of guanidine, but the off-rate of ligand binding and/or the tetraloop-tetraloop interaction is fast enough that the complex falls apart as the gel is running (Figure 4.6A). If this interpretation is correct, guanidine is able to cause the complete melting of a 14 bp duplex, followed by rearrangement into the dimer of two hairpins, suggesting that guanidine could induce a comparably dramatic rearrangement of helices *in vivo*. Exposure of the Shine-Dalgarno sequence of a bacterial mRNA allows the ribosome to be recruited for translation of the downstream genes that control guanidine toxicity in the presence of ligand.

Bioinformatic analysis has revealed that operons controlled transcriptionally by class I guanidine riboswitches often contain a class II riboswitch that controls translation of the 125 kDa carboxylase gene [158]. The combination of these two riboswitches allows the cell to exercise guanidine-dependent gene control at both the transcriptional and translational level. The large difference in both RNA fold and the mechanism of gene control enables tight regulation of genes meant to relieve guanidinium toxicity. The dissociation constants measured by in-line probing for all three guanidine riboswitch classes are mid-micromolar and the guanidine-II class displays cooperative binding [40; 158; 161]. It makes sense that the dissociation constants would be similar for all three classes, since these RNAs are likely tuned to turn on at the concentration of guanidine that causes toxicity in bacteria.

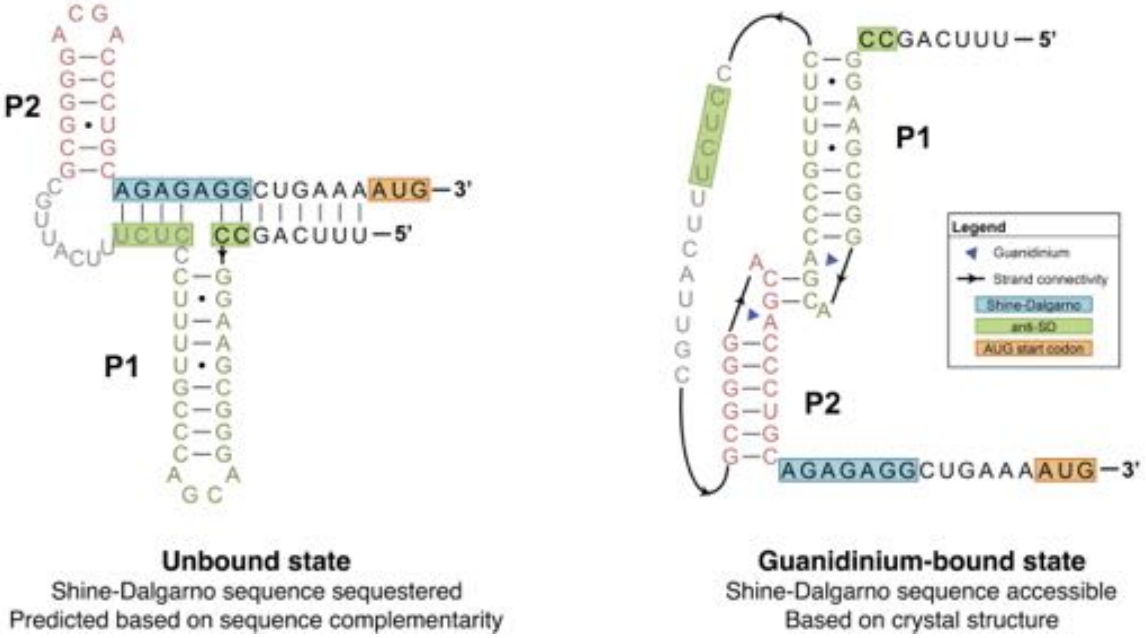


Figure 4.5: Predicted switching model for the P. aeruginosa guanidine II riboswitch. Upon guanidinium binding, the Shine-Dalgarno sequence is made accessible by spatial separation from the anti-Shine-Dalgarno (anti-SD) sequence. Green nucleotides represent hairpin P1 and pink represents hairpin P2.

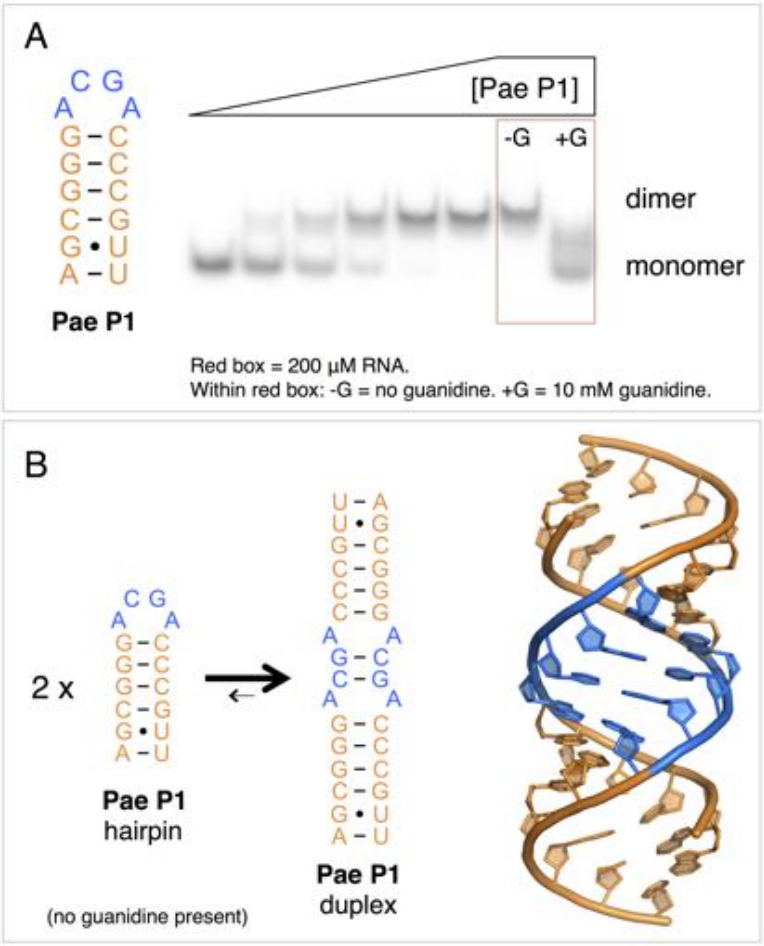


Figure 4.6: Gel shift data demonstrating a large conformational rearrangement of the *P. aeruginosa* P1 hairpin from the guanidine-II riboswitch in the absence and presence of guanidine. (A) Left: secondary structure of the *P. aeruginosa* P1 hairpin, with the basepairing region shown in orange and the tetraloop in blue. Right: gel shift with 5′-end labeled *P. aeruginosa* P1 hairpin RNA. Increasing RNA concentration leads to duplex formation (top band) in the absence of ligand. The flat region of the gradient is 200 µM RNA with and without ligand present. (B) Illustration of duplex formation in the absence of ligand, including a crystal structure on the right solved to 2.75 Åin the absence of guanidine. Color scheme of the secondary structure is the same as in A.

### 4.3.3 Guanidine riboswitches showcase convergent evolution in RNA

As of today, three structurally distinct classes of guanidine riboswitch have been identified in bacteria [40; 158; 161]. The structures of classes I and II are the most structurally distinct. Class I is large, heavily conserved with many non-canonical base pairs and complex tertiary interactions, and controls gene expression at the transcriptional level. Class II is small, exceptionally simple, and controls genes through cooperative binding of two guanidine ligands at the translational level. The crystal structure of class III was also recently solved [162]. Class III is smaller and simpler than class I, but larger and more complex than class II. Like class II, it contains a conserved ACGA motif (ACGR in class II) and it controls downstream genes at the translational level. Like class I, it binds a single equivalent of guanidine. The overall structure of guanidine-III is very different from the other two classes. It forms an H-type pseudoknot and a triple helix. In between these two structural motifs, a conserved core forms the binding pocket for guanidine. C6 and G7 of the conserved ACGA motif form the same interactions to guanidine in class III as they do in class II, but the RNA context is very different from class II [162]. It is unclear if the ACGA motif has arisen from divergent on convergent evolution.

The existence of multiple guanidine riboswitch classes demonstrates that RNA has found a way to bind to guanidine to mediate guanidine toxicity independently more than once. Additionally, bacteria that found themselves containing multiple guanidine riboswitch classes have taken advantage of the fact that they regulate gene expression at different levels (transcription vs. translation) and it has been beneficial for organisms that have them both located in the same operon for especially tight gene regulation [158]. The

evolutionary relationship between classes II and III is unclear. Sequence and structural dissimilarities between class I and classes II and III suggest that the recognition strategies employed by these RNAs, most notably using the Hoogsteen face of guanine and π-cation interactions, have arisen by convergent evolution. The three classes of guanidine riboswitch demonstrate the structural versatility of RNAs when performing a similar function.

# 5 Structure of the ZMP/ZTP riboswitch aptamer

This is the first structure I worked on in the Strobel lab. The goal was to unpack the structural features that impose both high affinity and specificity on a riboswitch aptamer. Like the guanidine riboswitch, the ZMP/ZTP riboswitch must reject molecules that are chemically similar to the native ligand that also exist at high concentrations in bacteria, while maintaining high affinity for its native ligand.

## 5.1 Background

In 2013, Ron Breaker's lab determined that the conserved *pfl* RNA was a riboswitch that responds to the purine biosynthetic intermediate 5-aminoimidazole 4-carboxamide monophosphate (abbr. ZMP or AICAR) or its triphosphorylated derivative (ZTP). This class of riboswitch was named because it is commonly found upstream of the pyruvate formate lyase (*pfl*) gene [39]. This enzyme is responsible for the production of formate and acetyl-CoA from pyruvate and coenzyme A (CoA). Formate is used in the folate cycle to make 10-formyl THF, which is needed for ring-forming reactions in *de novo* synthesis of purines and histidine. *pfl* RNAs recognize ZMP and ZTP (Figure 5.1) with nanomolar affinity and turn on transcription of downstream genes. Other genes commonly regulated by *pfl* RNAs encode enzymes in the folate cycle and *de novo* purine biosynthesis (Figure 5.2B). In *Clostridium bartlettii*, the riboswitch controls transcription of the purH gene, which encodes the enzyme phosphoribosylaminoimidazole carboxamide formyltransferase. purH is a bifunctional enzyme that converts 10-formyl THF and ZMP to THF and formyl-ZMP and then subsequently catalyzes the cyclization of the pyrimidine ring to produce inosine monophosphate (IMP). If

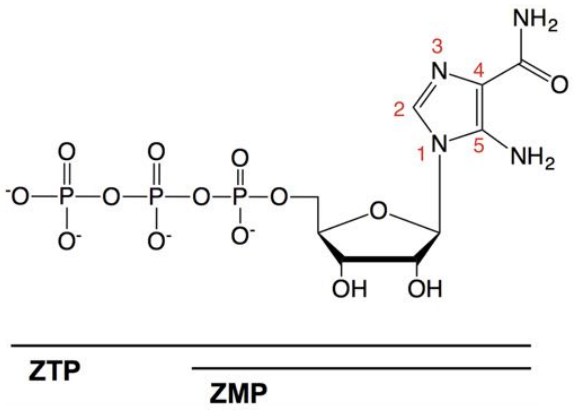


Figure 5.1: The chemical structure of the Z ribonucleotide

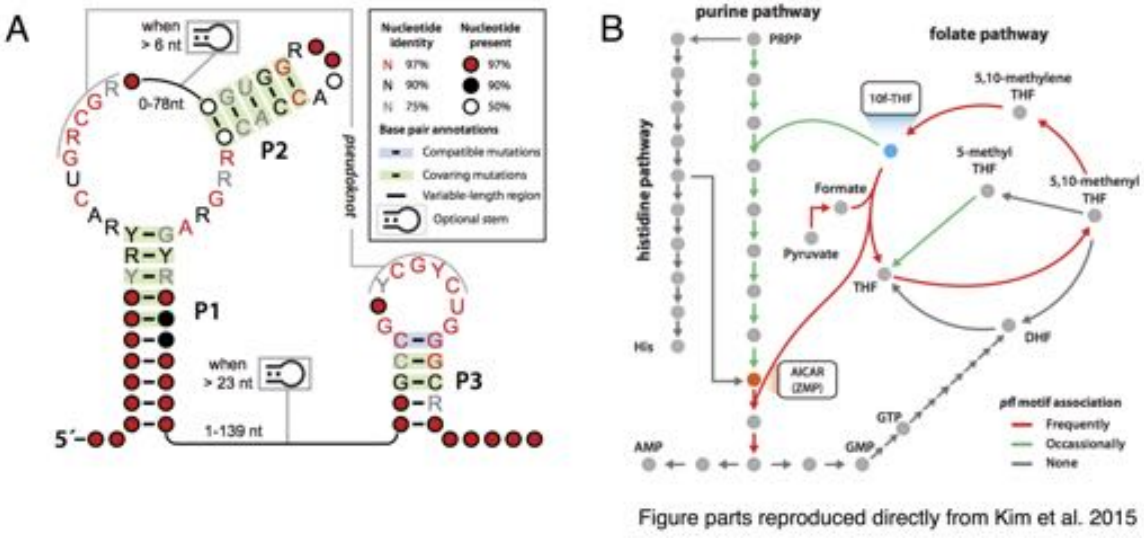


Figure 5.2: A. Consensus motif for the ZMP riboswitch. B. Genes controlled by the ZMP riboswitch. Figure parts A and B reproduced directly from Kim, PB et al. 2015 [39]

there is not enough enzyme to handle the existing pool of substrate, ZMP will build up and the riboswitch provides a feedback mechanism to increase enzyme production. Through the riboswitch, ZMP and ZTP act as signaling molecules to upregulate the expression of enzymes in the folate cycle and the *de novo* purine biosynthesis pathway.

The existence of the ZMP/ZTP riboswitch also sheds light on a 30-year-old controversy in the literature regarding the role of ZMP as an alarmone in bacteria. In 1982, while studying stress response in bacteria, Bruce Ames and Barry Bochner found that cellu-

lar ZTP levels increased drastically in response to folate stress. They also found that mutant strains incapable of making ZMP were extremely sensitive to folate antimetabolites [163]. Eight years later, Christopher Rohlman and Rowena Matthews reported that ZTP was not related to folate stress [164]. From this time up until the discovery of the riboswitch, ZMP and ZTP were largely unstudied in bacteria. The existence of the ZMP/ZTP riboswitch shows that ZMP and ZTP play a signaling role in the folate cycle.

## 5.2 Results

### 5.2.1 Structure determination of the ZMP/ZTP riboswitch aptamer

I set out to determine the crystal structure of the *C. bartlettii* ZMP/ZTP riboswitch aptamer, which the Breaker lab used in their biochemical studies. I grew native crystals of this aptamer in a condition containing 23% PEG 400, 50 mM MES, pH 5.2, 100 mM $MgSO_4$, and 400 mM KCl in the C2 space group. I attempted to phase the structure by soaking in or co-crystallizing with iridium (III) hexammine, but could never get enough anomalous signal. It is possible that this was due to the high concentration of sulfate precipitating out the iridium (III) hexammine, which is a documented issue in RNA crystallography [165]. Before I could solve the phase problem, three structures of the ZMP riboswitch were published independently [166; 107; 167]. I used the structure from Trausch et al. to solve the phase problem for my native dataset, resulting in an electron density map out to 2.95 Åresolution (X-ray statistics are located in Table 5.1. The secondary and tertiary structure of the crystal construct are displayed in Figure 5.3A and B.

### 5.2.2 Overall architecture

The RNA structure is dominated by a coaxial stack between P1 and P2. Conserved nucleotides 13-16 zip up into a helix with conserved nucleotides 53-56 as an extension of P1. Two of these base pairs are a pair of conserved sheared GA base pairs. A stretch of conserved nucleotides in between helix P1 and P2 form a pseudoknot with the conserved L3 loop, as predicted by bioinformatics. The short P3 helix lies parallel to the P1 helix, which it docks into. The non-conserved P5 helix is positioned in between the pseudoknot

Table 5.1: X-ray statistics for the ZMP/ZTP aptamer from *Clostridium bartlettii*

| | ***C. bartlettii* ZMP/ZTP aptamer (native)** |
|---|---|
| **PDB** | **not deposited** |
| **Data collection:** | |
| Beamline | 14-1 at SSRL |
| Space group | C2 |
| **Unit cell:** | |
| a, b, c (Å) | 105.2, 67.0, 46.9 |
| $\alpha$, $\beta$, $\gamma$ (°) | 90, 90.2, 90 |
| | |
| Wavelength (Å) | 1.1048 |
| Resolution (Å) | 50.00-2.95 (3.06 - 2.95) |
| $R_{merge}$ | 0.085 (0.550) |
| I/$\sigma$I | 18.1 (2.1) |
| $CC_{1/2}$ in highest resolution shell | 0.886 |
| CC* in highest resolution shell | 0.969 |
| Completeness (%) | 99.5% (97.2%) |
| Redundancy | 4.5 (3.6) |
| Total reflections | 13362 |
| Unique reflections | 6952 |
| **Refinement:** | |
| Resolution (Å) | 40.00-2.95 |
| No. reflections | 6613 |
| $R_{work}/R_{free}$ | 0.17/0.23 |
| **No. of atoms:** | |
| Total | 2018 |
| RNA | 1973 |
| Ligand | 22 |
| Cations | 3 |
| Water | 20 |
| **B-factors** | |
| Overall (Å$^2$) | 82.3 |
| RNA (Å$^2$) | 82.7 |
| Ligand (Å$^2$) | 63.4 |
| Cations (Å$^2$) | 68.7 |
| Water (Å$^2$) | 64.4 |
| **Root-mean-square deviations** | |
| Bond lengths (Å) | 0.0075 |
| Bond angles (°) | 1.969 |

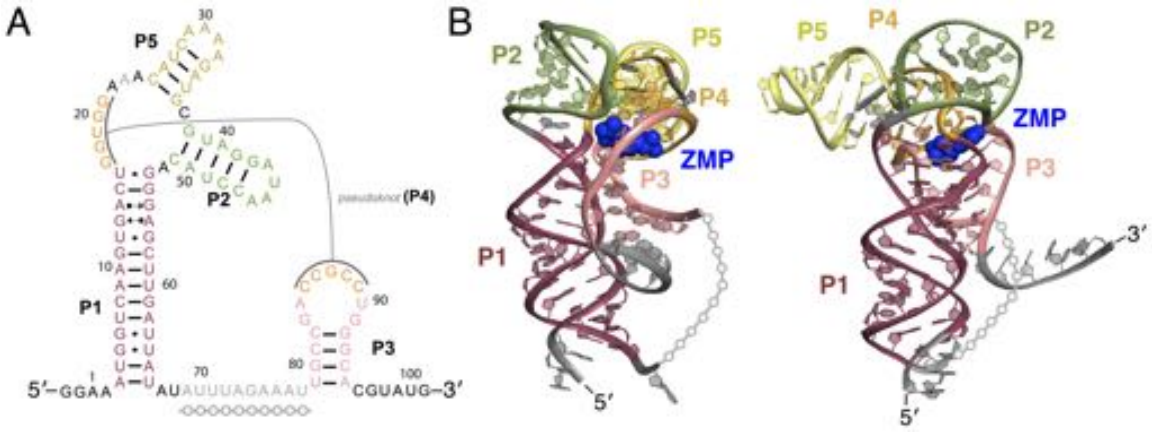


Figure 5.3: Structure of the ZMP/ZTP riboswitch from *C. bartlettii*. (A) Secondary structure of the *C. bartlettii* ZMP/ZTP riboswitch aptamer, colored by paired element. Base pairing notation is depicted as defined by Leontis and Westhof [168] White circles with grey outline under the linker region indicate nucleotides that are not seen in the electron density due to being unstructured. (B) Two views of the *C. bartlettii* ZMP/ZTP riboswitch aptamer crystal structure, rotated 90° apart. RNA is represented as a cartoon colored as in part A. ZMP is shown in blue spheres. The white circles with a grey outline show where the unstructured linker would connect the RNA.

and helix P2, but sticks straight out into solvent and does not form any interactions with the rest of the aptamer. The ZMP binding pocket is located in a conserved pocket where helices P1, P2, P3, and the pseudoknot (P4) come together. The binding pocket exists mainly at the border between helix P3 and the pseudoknot and extensive contacts are made to both of these helices.

### 5.2.3 ZMP recognition by the riboswitch

Like any nucleotide monophosphate, the ZMP ligand has three components: the base, sugar, and phosphate. The Z base is a major focal point of recognition. The Watson-Crick face of highly conserved U90 forms interactions to the "Hoogsteen" face of the Z base (Figure 5.4A). The N3 of ZMP (equivalent to the N7 position of purine bases) accepts a hydrogen bond from the N3 of U90. In solution, the carboxamide moiety is free to rotate, meaning that the nitrogen and oxygen can exchange positions. The proximity of the O2 of U90 to the carboxamide moiety suggests that the closest atom to it is a hydrogen bond donor. Therefore, the carboxamide nitrogen has been positioned facing U90. The carboxamide nitrogen also donates a hydrogen bond to a nonbridging phosphate oxygen of C47. This orientation of the carboxamide means that the "Hoogsteen" face of ZMP resembles that of AMP. On the other side of the carboxamide moiety, the oxygen makes an inner sphere coordination to a magnesium ion. The same magnesium ion is also coordinated by the pro-$S_P$ nonbridging phosphate oxygen of U16 and and the pro-$R_P$ nonbridging phosphate oxygen of C47 (Figure 5.4). N5 of ZMP donates two hydrogen bonds to the O5′ bridging phosphate oxygen of U16 and to the N7 of G83. Stacking interactions between the $\pi$-orbitals of the imidazole ring in ZMP and the $\pi$ systems of G17 and G91 are also

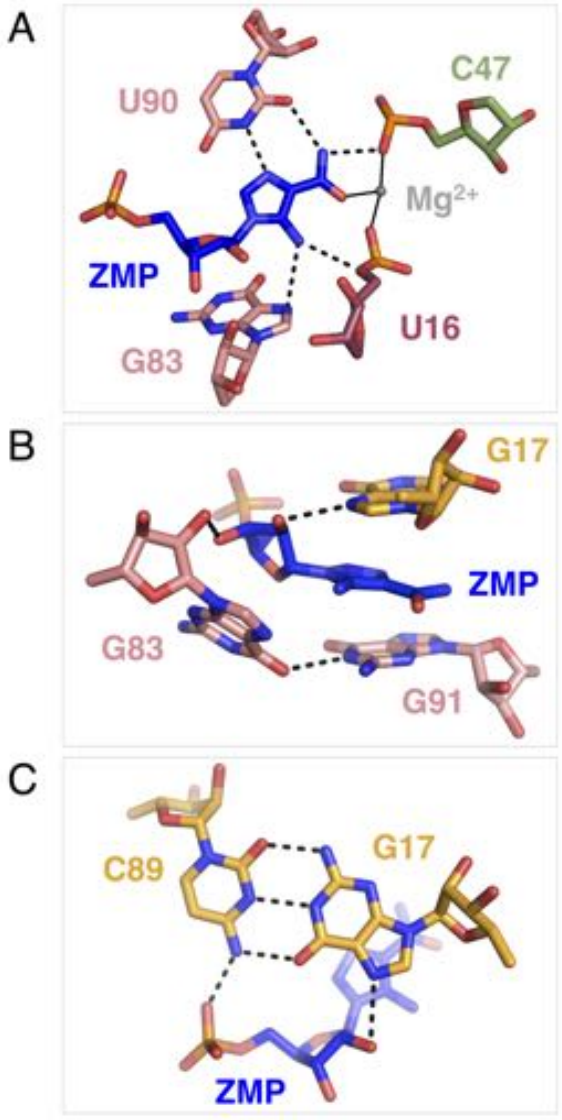


Figure 5.4: The binding pocket of the ZMP/ZTP riboswitch aptamer. RNA and ZMP are represented as sticks and colored by element. The carbon coloring is the same as in the previous figure. (A) Interactions between the Z base and the RNA. Dashed lines indicate hydrogen bonds and solid lines indicate metal coordination. (B) Contacts from the ribose of the ligand to the RNA. Dashed lines represent hydrogen bonds. (C) Single contact from the phosphate of ZMP to the RNA. The dashed lines represent hydrogen bonds.

observed (Figure 5.4B).

The two hydroxyls on the ribose of ZMP are each engaged in a hydrogen bond (Figure 5.4B). The 3′ hydroxyl hydrogen bonds to the 2′ hydroxyl of G83 and the 2′ hydroxyl hydrogen bonds to the N7 of G17. Additionally, the sugar is positioned over the base of G83, suggesting that there is a stabilizing electrostatic interaction between the two. Only one hydrogen bond is formed from the phosphate on ZMP to the RNA. One phosphate oxygen accepts a hydrogen bond from the N4 of C89 (Figure 5.4C). This is consistent data from the Breaker lab showing that the Z ribonucleoside (AICAr), which lacks the phosphate group, binds only about 10-fold weaker than ZMP [39].

### 5.2.4 Investigation of the magnesium ion in the ZMP/ZTP riboswitch binding pocket

A magnesium cation in the binding pocket of the ZMP riboswitch appears to make important inner sphere coordinations for ligand recognition, as mentioned earlier. It also coordinates nonbridging phosphate oxygens from U16 and C47, which brings them into close proximity of the binding pocket. I set out to investigate whether the magnesium ion contributed to the binding affinity of the aptamer. I also set out to investigate whether the magnesium ion is involved in rejecting other ligands that look similar to ZMP, which would make it an important contributor to specificity. An obvious look-a-like ligand is AMP, which has a complete six-membered ring rather than a carboxamide moiety. AMP could form the same interactions as ZMP via its Hoogsteen face, its sugar, and its phosphate. It differs at the N1, C2, and N3 positions. It is possible that the ZMP riboswitch rejects AMP by using the magnesium ion to coordinates U16 in a way that could sterically clash with a full purine ring.

To investigate the contribution to affinity and specificity made by the magnesium ion, I introduced a phosphorothioate substitution at the pro-$S_P$ nonbridging phosphate oxygen of U16. Phosphorothioate substitutions have been used extensively to investigate functional magnesium ions in ribozymes [169; 170; 171; 172; 173; 174]. Thousand or more fold defects in ribozyme rate have been observed for phosphorothioate substitutions at sites that coordinate an active site magnesium ion [172]. Oxygen is massively preferred over sulfur as a ligand for inner sphere magnesium coordination [175]. See Eckstein et al. for more review on the subject [176]. By introducing a sulfur atom at this position, I expected to destabilize the magnesium ion from its binding site either partially or fully. I performed a gel shift assay using radiolabeled ZMP and varied concentrations of either WT or U16 $R_P$ or $S_P$ phosphorothioate substituted RNA. The ZMP dissociation constant for the wild type and $R_P$ substituted RNAs is approximately 60 nM (Table 5.2). There is no defect in the $R_P$ substituted RNAs, suggesting that the sulfur substitution alone does not cause binding issues. The $S_P$ substituted RNA, however, has a $K_d$ of 390 nM. I have interpreted this to mean that the magnesium ion has been destabilized to some degree, resulting in a 7-fold decrease in affinity to ZMP through the loss of a Mg-O inner sphere coordination. I attempted to perform a manganese rescue (9 mM $MgCl_2$ and 1 mM $MnCl_2$, compared to my normal 10 mM $MgCl_2$), but this was unsuccessful (data not shown). The binding defect is minimal, which makes a lack of manganese rescue difficult to interpret, as manganese rescue is often imperfect [174]. From this data alone, I cannot rule out the possibility that the destabilization of magnesium has caused changes in RNA folding that have resulted in a binding defect, rather than specifically the loss of a single inner sphere coordination from an RNA-coordinated magnesium ion to the ZMP ligand.

In addition to studying ZMP affinity, I was interested in examining the contribution of the magnesium ion toward specificity of ZMP over AMP, which is present at millimolar concentrations in the cell. To do so, I used a competition gel shift assay, where the labeled ZMP is in trace, the RNA is at a constant concentration (set such that ZMP is mostly bound in the absence of competitor), and I varied the concentration of cold AMP. The wild type RNA is about 1400 fold more selective for ZMP over AMP ($K_{d,\ AMP}$ = 83 $\mu$M). Both the $R_P$ and the $S_P$ substituted RNAs have very similar or, if anything, slightly tighter affinities for AMP (Table 5.2, Figure 5.5). The $S_P$ substituted RNA is 8-fold less selective than wild type for ZMP over AMP (although it is still 180-fold more selective for ZMP), due mostly to a reduction in ZMP affinity. This suggests that the magnesium ion is not involved in AMP binding and contributes to an increase in ZMP affinity without increasing AMP affinity. In order to rule out effects from RNA folding, in the future, I will test binding of the wild type RNA to a ZMP analog that contains a sulfur-substitution at the carboxamide oxygen, which was recently synthesized by Tanya Berbasova in our lab.

### 5.2.5 Interactions within the RNA aptamer

Helix P1 and P3 are oriented parallel to each other and a large number of helical packing interactions are made between the two helices. The conserved, sheared GA pairs are critical for the interaction between P1 and P3. A14 makes a type II A-minor interaction with G93 and A56 forms an imperfect type I A-minor interaction with the G93-C81 base pair. A litany of other helical packing interactions also occur. I will describe these from the perspective of the P1 helix, going from 5′ to 3′ for each strand. In the first strand, G13 positions A56 to make the A-minor, but does not directly participate in packing interactions with P3.

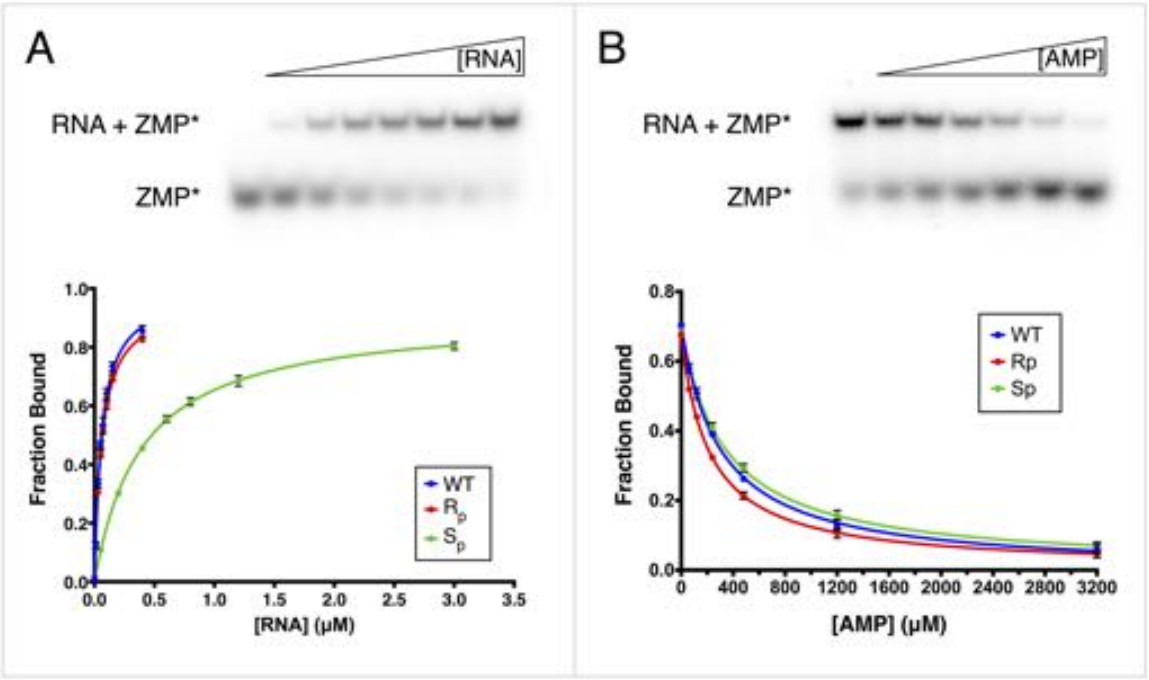


Figure 5.5: Gel shift assays for U16 phosphorothioate-substituted (PS) riboswitches. (A) Direct binding assay measuring ZMP affinity for WT, U16 $R_P$ PS, and U16 $S_P$ PS RNAs. Top: representative gel shift. Bottom: Binding data fit to a hyperbolic curve. (B) Competition gel shift assay measuring AMP affinity for WT, U16 $R_P$ PS, and U16 $S_P$ PS RNAs. Top: representative competition gel shift. Bottom: Binding data fit to a competition binding fit, defined in the materials and methods, equations 6 and 7.

Table 5.2: Dissociation constants for ZMP and AMP to phosphorothioate-substituted *C. bartlettii* ZMP riboswitch aptamers

| **RNA** | **$K_d$ for ZMP** | **$K_d$ for AMP** | **AMP $K_d$ / ZMP $K_d$** | **Fold less selective vs. WT** |
|---|---|---|---|---|
| WT | $58 \pm 2$ nM | $83 \pm 4$ μM | 1400 | 1 |
| U16 $R_P$ PS | $61 \pm 3$ nM | $64 \pm 2$ μM | 1000 | 1.4 |
| U16 $S_P$ PS | $390 \pm 10$ nM | $70 \pm 3$ μM | 180 | 8.0 |

A14 is involved in the previously mentioned A-minor. The 2′ hydroxyl of C15 donates a hydrogen bond to the N7 of G83 and accepts a hydrogen bond from the N2 of G92. In the second strand, the N2 of G53 donates a hydrogen bond to a nonbridging phosphate oxygen on A84. The 2′ hydroxyl and N2 of G54 donate hydrogen bonds to the O4′ of G83 and a nonbridging phosphate oxygen on A84, respectively. The 2′ hydroxyl of G55 hydrogen bonds to the 2′ hydroxyl of C82. G55 is also involved in the sheared GA pair with A14 and therefore plays a role in positioning A14 to form the A minor interaction. A56 is involved in the previously discussed A minor interaction. The 2′ hydroxyl of G57 donates a hydrogen bond to the O3′ nonbridging oxygen of C94. Together, these interactions stabilize the long-range, parallel packing between P1 and P3, which sets up the binding pocket for ZMP at the top of the P3 helix.

Helix P2 coaxially stacks on P1 and conserved nucleotides in P2 interact with the pseudoknot to form part of the ZMP binding pocket. There is a $\geq$97% conserved C47-G42 base pair at the end of the P2 helix. The N4 of C47 donates a hydrogen bond to a nonbridging phosphate oxygen of G18 in the pseudoknot. The conserved nucleotide A46 in L2 forms a type I A-minor interaction with the G17-C89 base pair in the pseudoknot. The C47/G18 hydrogen bond and the A-minor interaction position the pro-$R_P$ nonbridging phosphate oxygen of C47 to make the magnesium-mediated contacts to ZMP and U16.

## 5.3 Discussion

This resolution does not improve on the existing structures and in fact pales in comparison to the crystal structure solved by Trausch et al. to 1.80 Åresolution. However, it does offer a distinct advantages over the other three structures. My structure of is an entirely wild type sequence of the exact aptamer used by Ron Breaker's lab for biochemical studies [39]. This allows direct comparisons to be made between their data and my own structure. We characterized the solution behavior of this riboswitch by gel filtration and have shown that it exists almost exclusively as a monomer, making it ideal for future biochemical experiments. Many ZMP/ZTP riboswitch aptamers suffer from their tendency to form non-natural oligomers, with the most common form being a dimer. Examination of the structure and consensus motif (Figure 5.4A) explains why. Stems P1 and P3 are separated by a non-conserved linker region. Any interactions between P1 and P3 and the formation of the pseudoknot are long-range interactions encouraged by the physical tethering contributed by the linker region. The linker region must be long enough to allow this interaction and short enough to effectively increase the probability that the two aptamer halves will associate on a kinetic timescale, but is otherwise not important. In the context of an mRNA, the pseudoknot and ribose zipper will always occur in *cis*. However, in a system where the isolated aptamer exists in solution, these interactions can form in *trans* as well as in *cis*, resulting in a non-natural dimer. The Trausch et al. structure was solved as a dimer [166]. Solution studies I perfomed on the *C. bartlettii* aptamer show that it exists mostly as a monomer immediately after transcription and that native purification of the monomer results in a species that remains stably as a monomer in solution

(Figure 5.6A,B). The presence of some dimer and oligomer appears after a freeze-thaw cycle (Figure 5.6C). Gel filtration followed by multi-angle light scattering (MALS) shows that a native purification method maintains the monomer, while a denaturing purification followed by re-folding results in messy oligomerization, likely from interactions in *trans* (Figure 5.6D,E). This construct is well-behaved in solution relative to other constructs I tested and my crystal structure allows direct comparisons to be made between the structure and biochemical data.

Another difference between my structure and the other three structures is the presence of a non-conserved helix, P5. Consistent with its lack of conservation, P5 sticks out into solution and makes no contacts with the rest of the RNA. It may assist in properly orienting the pseudoknot and the P2 helix relative to each other, although I have not performed any experiments to test this hypothesis.

The U16 phosphorothioate substitutions implicate the binding site magnesium ion in both ligand affinity and specificity. The ZMP/ZTP riboswitch must both bind to ZMP/ZTP and reject AMP/ADP/ATP, which exist in millimolar concentrations in the cell. Kim et al. demonstrated that the *C. bartlettii* ZMP/ZTP riboswitch does not differentiate ligands that contain one phosphate from ligands with more than one phosphate, which is why I will be discussing metabolite concentrations in terms of nucleotide pools [39] Interestingly, the dissociation constant of AMP for the wild type aptamer is about 80 μM by competition gel shift, which is well below the concentration of adenosine nucleotides present in the cell. I expect that *in vivo*, where the riboswitch is likely under kinetic control, that the adenosine nucleotide pool would need to be at a much higher concentration to trigger the riboswitch. The U16 $S_P$ phosphorothioate substitution, which was designed

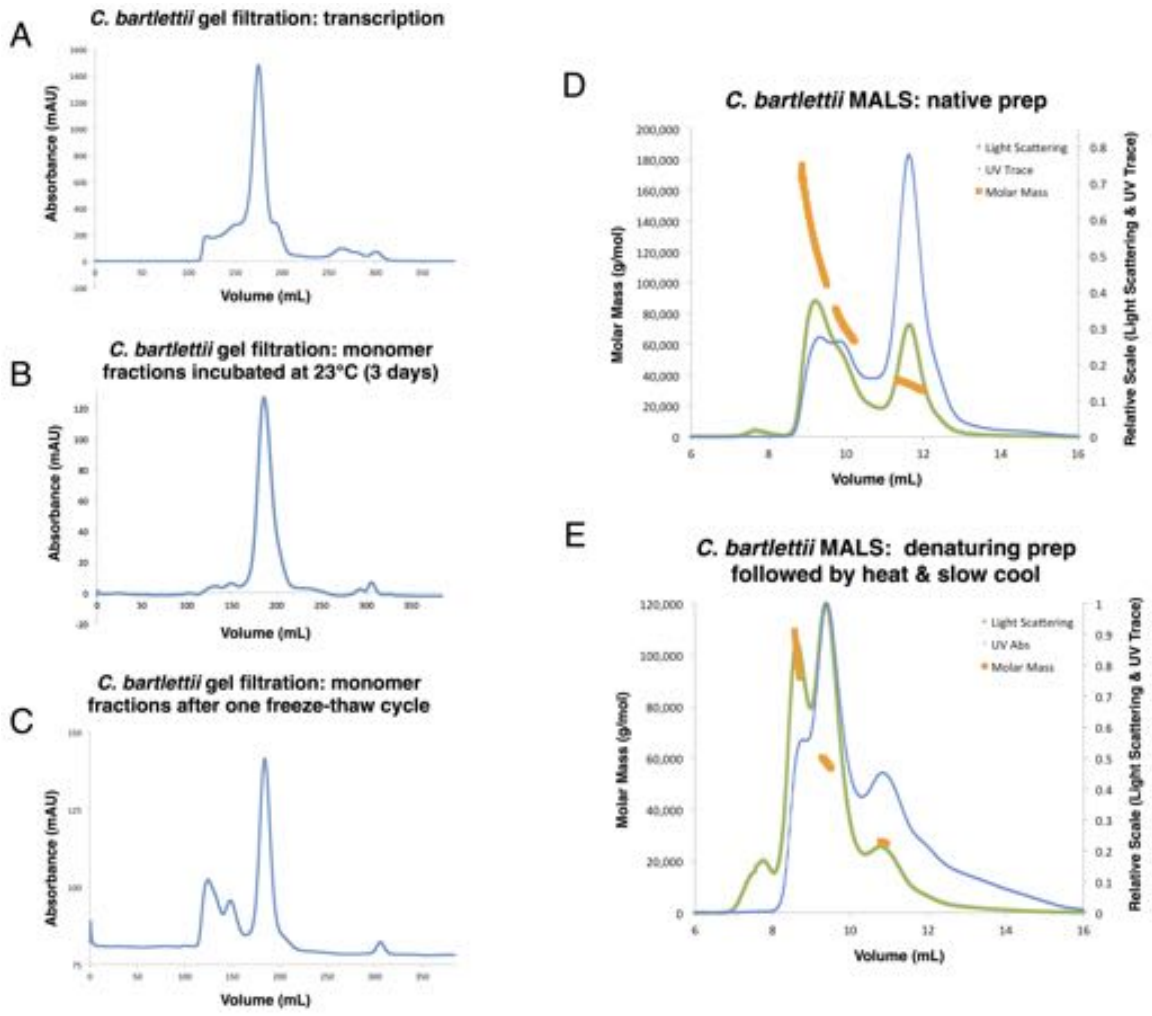


Figure 5.6: Solution behavior of the *C. bartlettii* ZMP/ZTP riboswitch aptamer examined by gel filtration. (A) $UV_{255\,nm}$ trace of a transcription reaction run on gel filtration, where the major peak is monomer. (B) Monomer fractions that have been incubated at room temperature for three days, showing that the RNA remains monomeric. (C) Monomer fractions after a freeze thaw cycle. The major species is still monomer, but there is some oligomerization. (D) Multi-angle light scattering data for a natively-purified RNA after a freeze thaw cycle. The $UV_{255\,nm}$ trace is depicted in blue. Light scattering is shown in green. Molar mass is shown in orange. (E) Multi-angle light scattering for aptamer purified by denaturing gel and re-folded. Significant oligomerization is observed. Coloring is the same as in part D.

to destabilize magnesium ion binding, is 8-fold less selective than the wild type for ZMP binding over AMP, due mainly to a decrease in ZMP affinity. The magnesium ion provides a way for the riboswitch to increase ZMP affinity without increasing AMP affinity. During times of folate stress, the ZMP and ZTP pool reaches low millimolar concentrations [163], similar to that of the AMP/ADP/ATP pool concentrations [177]. The wild type exhibits 1400-fold selectivity for ZMP over AMP, which decreases to 180-fold selectivity with the destabilization of the magnesium ion. This remaining selectivity could reasonably be explained by the two hydrogen bonds donated by N3 of ZMP to the RNA. In the case of an adenosine nucleotide, N3 would be a hydrogen bond acceptor, not a donor. Together, these data explain the selectivity of the riboswitch for ZMP over AMP.

The structure of the ZMP/ZTP riboswitch further demonstrates the versatility of RNA in ligand recognition and specificity. This RNA takes advantage of a magnesium ion in order to increase affinity for Z nucleotides, while simultaneously rejecting adenosine nucleotides.

# 6 Conclusions

Together, these data contribute to the field's knowledge of the versatility and evolvability of RNA structure and have implications for the RNA world theory. A significant amount of energy in the field has been used to demonstrate that RNAs can be catalytic, which bolsters the RNA world theory by demonstrating that RNA could be self-replicating and perform reactions necessary for early life. This data demonstrates that RNA is highly evolvable, which is another requirement of the RNA hypothesis. The *ykkC* RNA is a particularly compelling scaffold, since it is capable of specifically to PRPP, which is an activated form of ribose 5-phosphate that may have been used in the RNA world as a substrate for nucleotide synthesis. Additionally, *ykkC* RNA has an exceptionally rugged functional landscape. Without this versatility, RNAs of new functions would not arise readily from a hypothetical first self-replicating RNA.

# 7 Materials and methods

## 7.1 Materials and methods: guanidine-I riboswitch

### 7.1.1 RNA preparation and purification

Plasmids containing the *ykkC* RNA from *Sulfobacillus acidophilus* were obtained from GeneArt at ThermoFisher Scientific. RNA was transcribed using a T7 RNA polymerase and purified natively by gel filtration in the presence of 1 mM guanidine (Sigma Aldrich).

### 7.1.2 Crystallization and structure determination

RNA (130 μM) in a folding buffer (10 mM $MgCl_2$, 10 mM KCl, 10 mM HEPES-KOH, pH 7.5, and 10 mM guanidine) was mixed 1:1 with a solution of 32% MPD, 40 mM $SrCl_2$, 12 mM spermine, 40 mM Na-acetate, pH 4.6, and 800 μM iridium (III) hexamine. Crystals were grown at 23°C using the microbatch-under-oil method using 2:1 paraffin:silicon oil. Crystals appeared within 24 hours. The crystals were flash frozen in liquid nitrogen without any further preparation.

Unbiased phases were generated using a combination of SIRAS and MAD phasing. SAD data collected at the iridium hexamine peak wavelength and a native dataset were collected at beamline 19BM at the Advanced Photon Source (APS). Three-wavelength MAD data were collected at beamline 8.2.2 at the Advanced Light Source (Berkeley, CA). Data were processed using HKL2000. Phenix AutoSol was used to locate heavy atom sites and to generate initial phases with both the SIRAS data and the MAD data. Three iridium sites were located with AutoSol and verified using anomalous difference Fourier methods.

Model building was done in Coot. Refmac and Phenix Refine were used for refinement. Each strontium site was identified by first modeling a magnesium ion and then using difference Fourier methods to determine if density was still unaccounted for. Figures of the crystal structure were made in PyMol.

### 7.1.3 $K_d$ determination of RNAs by equilibrium dialysis

Natively purified RNAs were buffer exchanged into an RNA-folding solution containing 50 mM HEPES-KOH, pH 7.5, 200 mM KCl, and 20 mM $MgCl_2$. The RNA and trace amounts of [$^{14}$C]-guanidine-HCl (Moravek Biochemicals) were added to separate sides of a 10,000 MWCO Dispo Equilibrium Dialyzer from Harvard Apparatus and were equilibrated overnight at room temperature. The amount of [$^{14}$C]-guanidine on each side of the dialysis membrane was determined by scintillation counting in Ultima Gold on a PerkinElmer Tri-Carb 4910 TR scintillation counter. The fraction bound was determined by the following equation:

$$Fraction\ bound = \frac{cpm\ on\ RNA\ side - cpm\ on\ ligand\ side}{cpm\ on\ RNA\ side} \tag{1}$$

The data were fit in GraphPad Prism to the following equation:

$$F = \frac{B_{max} * X}{K_d + X} \tag{2}$$

where F= fraction bound, $B_{max}$= fraction bound at saturation, and X= riboswitch concentration. For mutants that did not reach saturation over the riboswitch concentrations tested,

$B_{max}$ was fixed at 1.0, in agreement with the $B_{max}$ of the wild type and crystallization constructs.

## 7.2 Materials and methods: PRPP and ppGpp riboswitches

### 7.2.1 RNA Transcription and Purification

RNA was prepared essentially as in Reiss et al.8 Plasmids containing *ykkC* PRPP riboswitch DNA from T. mathranii downstream of the T7 promoter were obtained from GeneArt at Thermo Fisher Scientific. The aptamer domain was extended at the 5′ end by one nucleotide to aid transcription by T7 polymerase [178]. Plasmid DNA was prepared using a QIAgen MaxiPrep kit and the accuracy of the sequence was verified using Sanger sequencing [179]. Template DNA for transcription was made using PCR with Phusion polymerase. Template was purified using the Zymo Research DNA Clean and Concentrator 500 kit.

RNA was transcribed from 20 ng/μL PCR template using T7 polymerase in the presence of 80 mM HEPES-Na pH 7.5, 5 mM DTT, 1 mM spermidine, 0.12 mg/mL bovine serum albumin, 6 mM NTPs, 44 mM $MgCl_2$, and 1 U/nL inorganic pyrophosphatase [180]. Transcription reactions proceeded for approximately 4 hours at 37°C. Monomeric RNA was exchanged into gel filtration buffer (50 mM MES pH 6.2-6.3, 100 mM KCl, 10 mM $MgCl_2$), filtered, and purified natively on a HiLoad 26/600 Superdex 75 pg gel filtration column in a cold room (6 $\pm$ 2°C). Monomers eluted at ca. 0.6 column volume, and were pooled and concentrated to greater than 100 μM.

### 7.2.2 Crystallization and structure determination of the wild-type aptamer

Crystals were grown using the microbatch-under-oil method with 2:1 paraffin:silicon oil. In all cases, crystals appeared within two days. Initial crystallization screening was performed using Hampton Research Natrix HT at 23°C and 30°C. To produce the wild type

crystals used for data collection, 2 μL of 150 μM RNA in 10 mM $MgCl_2$, 10 mM KCl, 10 mM HEPES-KOH pH 7.5, and 10 mM PRPP was mixed with 1 uL of a solution of 80 mM sodium chloride, 20 mM barium chloride dihydrate, 40 mM sodium cacodylate trihydrate pH 5.6, 45% v/v (+/-)-2-methyl-2,4-pentanediol (MPD), and 12 mM spermine tetrahydrochloride and incubated at 30°C. To produce the G96A crystals, 150 μM RNA in 10 mM $MgCl_2$, 10 mM KCl, 10 mM HEPES-KOH pH 7.5, and 1 mM ppGpp was mixed a solution of 80 mM sodium chloride, 40 mM sodium cacodylate, pH 7.0, 30% MPD, and 12 mM spermine (1 μL RNA solution plus 0.8 μL reagent) and incubated at 23°C.

Crystals were flash-frozen without further preparation. For the wild type aptamer, a solution was generated using molecular replacement with the *ykkC* guanidine riboswitch as an initial model (PDB ID: 5T83)8. For the G96A mutant, a solution was generated using molecular replacement with chain A of the PRPP riboswitch structure presented in this study as an initial model. Data were processed using HKL-2000 [181]. Model building was performed in Coot [182].

The wild type aptamer crystallized in space group P21 with two molecules present in the asymmetric unit. Discussion is for the most part limited to chain A as there is better structural information for this entity than for chain B. The first component modeled was the RNA. Further unaccounted-for electron density was assigned to metal ions and water molecules. This process yielded a structure in which one significant area of electron density in each chain was unaccounted for. One molecule of PRPP and its two associated metal ions fit well in this area of density.

The G96A aptamer crystallized in space group P1 with four molecules in the asymmetric unit. Discussion in the manuscript is limited to chains A-C, due to chain D

yielding generally poorer density. Overall, chain D is consistent with chains A-C, but more subject to error in individual atom positions. Regions disagreeing with the wild type (mainly in the S-turn) were deleted and re-modeled. Very large $\sigma$ peaks in the difference Fourier map, corresponding to the very electron dense pyrophosphate moieties of ppGpp, were used to identify the ppGpp binding pocket.

Refinement of the two structures was performed with Refmac and Phenix [183; 184]. Refinement was concluded when no more entities could be modeled into the electron density and computational refinement ceased to produce improvements in $R_{work}$ and $R_{free}$. Metal ions were identified by first modeling a magnesium ion and evaluating coordination geometry, B factors, and unaccounted-for density using difference Fourier methods, followed by reassignment where appropriate. The figures of the crystal structure were made in PyMOL [185]. The ligand interaction map was made in ChemDraw.

### 7.2.3 Synthesis of ($\beta$-$^{33}$P) PRPP and determination of the dissociation constant of the PRPP-aptamer complex

($\beta$-$^{33}$P) PRPP was synthesized using E. coli ribose-phosphate pyrophosphokinase (RPPK) obtained from Abbexa. 17.7 $\mu$g/mL RPPK was incubated at 37°C for two hours in the presence of 50 mM potassium phosphate dibasic pH 8, 10 mM ribose 5-phosphate, 5 mM $MgCl_2$, and trace quantities of ($\beta$-$^{33}$P) ATP [186]. PRPP and ATP were separated on a native 20% acrylamide gel at 4°C and PRPP was eluted overnight in 400 $\mu$L dH2O at 4°C.

### 7.2.4 Purification of SAS1 for synthesis of ppGpp

SAS1 enzyme was expressed and purified based on the protocol in Steinchen et al [187]. Briefly, the SAS1 protein from B. subtilis was amplified by colony PCR, cloned into a pET-28aM vector, and transformed into E. coli BL21(DE3) cells. A 15 mL starter culture was used to inoculate 1.5L Terrific Broth plus 50 μg/mL kanamycin and grown at 37°C. At OD600 about 0.8, expression was induced with 0.5 mM IPTG and the culture was shaken overnight at 18°C. Cells were then pelleted and lysed using a microfluidizer (lysis buffer: 50 mM Tris, pH 8.0, 300 mM NaCl, 20 mM imidazole, 20 mM $MgCl_2$, 20 mM KCl) and the lysate was run on a nickel column. The protein was eluted from the column with 400 mM imidazole (elution buffer: 50 mM Tris, pH 8.0, 400 mM NaCl, 400 mM imidazole, 0.2 mM TCEP). A band running between 25 and 30 kDa was seen on an SDS-PAGE gel, indicating SAS1 was successfully eluted. The eluted protein was diluted in 50 mM Tris, pH 8.0 and run on a Q column (HiTrap Q column, 5 mL) to remove contaminants. Finally, the Q column fractions were pooled and run on a gel filtration column (Superdex 200, running buffer: 20 mM HEPES-Na, pH 7.5, 200 mM NaCl, 20 mM KCl, 20 mM $MgCl_2$), and a peak eluted consistent with the relevant tetrameric assembly of the protein. The protein was concentrated and frozen at -80°C in aliquots for storage.

### 7.2.5 Synthesis and purification of ppGpp

The SAS1 protein accepts GDP (or GTP) and ATP as substrates and catalyzes the transfer of the γ and β phosphates from ATP onto the 3′ end of GDP or GTP to form ppGpp or pppGpp, respectively. To make unlabeled ppGpp for crystallography, a reaction setup based on the

protocol of Steinchen et al [187] was used. Briefly, 5 mM GDP, 5 mM ATP, and 5 μM SAS1 were combined in reaction buffer (100 mM HEPES-Na, pH 7.5, 200 mM NaCl, 20 mM $MgCl_2$, and 20 mM KCl) and incubated at 37°C for two hours. A chloroform extraction was performed to remove SAS1, followed by 10-fold dilution in $ddH_2O$ and purification by Q column (HiTrap Q HP, 5 mL column volume), where buffer A (10 mM HEPES-KOH, pH 7.5) was used to bind nucleotides to the column and a gradient of buffer B (2 M NaCl) was used to elute the nucleotides. Nucleotides eluted from the column such that the number of phosphates positively correlated with %B. ppGpp eluted last at about 15% buffer B (approximately 300 mM NaCl). ppGpp was then precipitated by lithium chloride (LiCl) precipitation. Eluate from the Q column was brought to 1 M LiCl, 4 volumes of ethanol were added, and the tubes were frozen at -20°C before centrifuging at 6000 rpm in an Eppendorf F-45-18-11 fixed-angle centrifuge rotor at 4°C for 10 min to pellet the precipitate. The supernatant was discarded and the pellet was washed twice with cold (-20°C) ethanol, repeating the freezing and pelleting steps between each wash step. After pouring off the ethanol of the final wash, pellets were completely dried. A dry, white powder resulted. Concentration was calculated by measuring UV absorbance at 252 nm ($\varepsilon_{252}$ = 13600 L $mol^{-1}$ $cm^{-1}$).

### 7.2.6 Synthesis and purification of [3′-β-$^{32}$P]-ppGpp

A reaction mixture resembling that in the previous section was made, substituting 5 mM ATP for 150 μCi [γ-$^{32}$P]-ATP (Perkin Elmer). [3′-β-$^{32}$P]-ppGpp was purified using a 20% denaturing polyacrylamide gel to separate it from [γ-$^{32}$P]-ATP. The band was soaked in 300 μL $ddH_2O$ overnight at 4°C. The gel slice was then filtered off and the solution containing

[3′-β-$^{32}$P]-ppGpp was frozen at -20°C for use in binding assays.

### 7.2.7 Determination of dissociation constants by equilibrium dialysis

The dissociation constant of the PRPP-RNA and ppGpp-RNA complexes were determined by equilibrium dialysis using cassettes with a 10 kDa cutoff obtained from Harvard Apparatus, essentially as in Reiss et al [80]. Trace quantities of radiolabeled ligand were dissolved in equilibrium dialysis buffer (50 mM HEPES-KOH, pH 7.5, 200 mM KCl, 20 mM $MgCl_2$) and were added to one side of the cassette, while varying concentrations of RNA dissolved in the same buffer were added to the other side of the cassette. The cassettes were incubated at room temperature overnight with gentle shaking and recovered by centrifugation. For ppGpp, which experiences negligible amounts of degradation overnight, 20 μL of the recovered material was directly subjected to scintillation counting. For PRPP, 10 μL of the recovered material was subjected to scintillation counting, while another 10 μL was electrophoresed on a denaturing 20% acrylamide gel containing 7.5 M urea. The latter step allowed determination of the amount of PRPP remaining in each sample after overnight incubation at room temperature in the presence of magnesium. The fraction of ligand bound in each cassette was determined using the following equation:

$$Fraction\ bound = \frac{cpm\ on\ RNA\ side - cpm\ on\ ligand\ side}{cpm\ on\ RNA\ side} \quad (3)$$

The data were fit in GraphPad Prism to the following equation:

$$F = \frac{B_{max} * X}{K_d + X} \tag{4}$$

where F= fraction bound, $B_{max}$= fraction bound at saturation, and X= riboswitch concentration. For constructs that did not reach saturation over the riboswitch concentrations tested, $B_{max}$ was fixed to 0.9564 for binding assays performed with [$^{33}$P]-PRPP or 0.9761 for for binding assays performed with [$^{32}$P]-ppGpp. These were the $B_{max}$ values obtained under conditions where saturation was reached with the same ligand.

### 7.2.8 Secondary structure prediction

Secondary structure predictions were obtained from RNAstructure using the default settings [188]. NCBI Reference Sequence NC_014209.1, location 657681 to 657861 was used for the sequence of the aptamer and expression platform. The secondary structure deemed most likely by the program was used for interpretation.

## 7.3 Materials and methods: guanidine-II riboswitch

### 7.3.1 RNA preparation

Unmodified RNA was purchased from Sigma Aldrich. RNA containing a single 5-bromouridine at position 14 was purchased from Dharmacon. All synthesized RNAs were purified by denaturing polyacrylamide gel electrophoresis.

### 7.3.2 Crystallization

Crystallization was performed using the microbatch-under-oil method with a 2:1 paraffin:silicon oil overlay. 200 μM *P. aeruginosa* P2 hairpin RNA in crystallization buffer (10 mM $MgCl_2$, 10 mM KCl, 10 mM HEPES-KOH, pH 7.5, and 40 mM guanidine) was mixed in a 1:1 ratio with crystallization reagent 1 for the 5-bromoU modified RNA (45% MPD, 50 mM MES, pH 5.6, 4 mM NaCl, 40 mM KCl, and 12 mM spermine) or crystallization reagent 2 for the unmodified RNA (40% MPD, 50 mM Na-acetate, pH 5.0, 12 mM NaCl, 80 mM KCl, and 12 mM spermine). Crystals appeared overnight at 23°C. Crystals were flash frozen in liquid nitrogen without further preparation, since 40-45% MPD is a cryoprotectant.

### 7.3.3 Structure determination

Unbiased phases were obtained by single wavelength anomalous diffraction (SAD) phasing. SAD data were collected at the peak wavelength for the bromine K edge at beamline 24-ID-C at the Advanced Photon Source (APS). Data were indexed, integrated, and scaled using HKL2000. The SHELXC/D/E suite was used to evaluate anomalous signal, locate

heavy atoms, and obtain the unbiased phases. Four bromine sites were identified, indicating four molecules in the asymmetric unit and allowing us to orient molecules within the density. Model building was performed in Coot. Refmac5 and phenix.refine were used for refinement. $R_{free}$ test sets were matched between the derivative and native datasets to prevent bias. Figures of the crystal structure were made using open source PyMol.

### 7.3.4 Gel shift of *P. aeruginosa* P1 hairpin

Deprotected *P. aeruginosa* P1 hairpins from Sigma Aldrich were 5′-end labeled using polynucleotide kinase and [$\gamma$-$^{32}$P]-ATP at 37°C for 1 hour. Labeled RNA was purified by 20% denaturing PAGE (7.5 M urea), followed by band excision and elution overnight into dd$H_2O$. Trace labeled RNA was used for the gel shifts. Unlabeled, deprotected *P. aeruginosa* P1 hairpin was titrated into trace labeled RNA at the following concentrations: 0 μM, 6.25 μM,12.5 μM, 25 μM, 50 μM, 100 μM, and 200 μM. The final gel shift mixtures consisted of trace [$^{32}$P]-P1 hairpin, varied cold P1 hairpin, 1X gel shift buffer (50 mM HEPES-KOH, pH 7.5, 200 mM KCl, 10 mM $MgCl_2$) and 1X sucrose loading dye. One additional condition was set up, identical to the 200 μM cold RNA condition, but with 10 mM guanidine added. These tubes were equilibrated overnight at room temperature. The next day, samples were electrophoresed on a 12% native polyacrylamide gel at 4°C for approximately 1 hour at 20 watts, with buffer recirculation every 20 minutes. The gel and running buffer both contained 10 mM $MgCl_2$. The gel was then dried under vacuum and exposed to a phosphor screen overnight. The screen was then phosphorimaged.

### 7.3.5 Accession number

Coordinates for the guanidine-II riboswitch have been deposited in the protein data bank (PDB) under accession codes 5VJ9 (native) and 5VJB (5-BrU). There are four molecules (two dimers) in the asymmetric unit.

## 7.4 Materials and methods: ZMP/ZTP riboswitch

### 7.4.1 RNA preparation and purification

Plasmids containing the *ykkC* RNA from *C. bartlettii* were created using primer extension to generate the insert, followed by restriction digestion and cloning into a pUC19 vector. Large-scale PCR was used to generate the template for transcription. RNA was transcribed using a T7 RNA polymerase and purified natively by gel filtration on a HiLoad 26/60 Superdex 75 pg column using an NGC chromatography system from Bio-Rad.

### 7.4.2 Crystallization and structure determination

RNA (100 μM) in a folding buffer (10 mM $MgCl_2$, 10 mM KCl, 10 mM HEPES-KOH, pH 7.5, and 250 μM ZMP) was mixed 1:1 with a solution of 23% PEG 400, 50 mM MES, pH 5.2, 100 mM $MgSO_4$, and 400 mM KCl. Crystals were grown at 30°C using the microbatch-under-oil method using 2:1 paraffin:silicon oil. The crystals were flash frozen in liquid nitrogen without any further preparation.

A native dataset was collected at beamline 14-1 at the Stanford Synchrotron Radiation Lightsource (SSRL). Data were processed using HKL2000. Refmac was used to perform molecular replacement using PDB ID 4XWF as a model [166]. Refmac and phenix.refine were used for refinement. To avoid significant bias from the original model, once I had built a complete model, I deleted and re-built the entire RNA section-by-section. Figures of the crystal structure were made in PyMol.

### 7.4.3 Synthesis and purification of [$^{32}$P]-ZMP

Adenosine kinase was purchased from R & D systems. 16 mM AICAr (the ribonucleoside version of Z) was incubated with 0.07 mg/mL adenosine kinase and 25 pmol (150 μCi) [γ-$^{32}$]-ATP in a buffer containing 25 mM HEPES, pH 7.0, 150 mM NaCl, 10 mM $MgCl_2$, and 10 mM $CaCl_2$ overnight at 37°C. Radiolabeled ZMP was separated from remaining radiolabeled ATP on a 20% polyacrylamide gel with 8 M urea, run at 30 watts for 1 hour at room temperature. The band was visualized by autoradiography, excised, and eluted into 300 μL of $ddH_2O$ overnight at 4°C.

### 7.4.4 Creation of phosphorothioate-substituted RNAs

An RNA oligonucleotide (nts -1 through 19 of the ZMP/ZTP riboswitch aptamer from *C. bartlettii*) was ordered from Dharmacon with a phosphorothioate substitution at position U16. The $R_P$ and $S_P$ diastereomers were separated by reversed-phase HPLC on a Phenomenex Luna 5μm C18(2) column (100 Åpore size, 250 x 4.6 mm). The mobile phase was 0.1 M ammonium acetate, pH 6.0 with an increasing gradient of acetonitrile to elute the oligonucleotides. The diastereomers eluted as separate peaks (first the $R_P$, then the $S_P$) between 6.0% and 8.0% acetonitrile. Fractions corresponding to the peaks were collected, lyophilized to dryness, and resuspended in $ddH_2O$. These RNAs were the 5′ components of the ligation.

The remaining, non-phosphorothioate-containing portion of the aptamer was made by *in vitro* transcription with 20 mM GMP added. T7 RNA polymerase can use GMP or GTP to initiate transcription [189]. Addition of 20 mM GMP resulted in a transcription

product with mostly monophosphate on the 5′ end rather than triphosphate. This RNA was purified by 8% denaturing PAGE (7.5 M urea). A band of appropriate size was excised, crushed and soaked overnight in 300 mM sodium acetate, pH 5.3, and gel slice was filtered off the next morning. An ethanol precipitation was performed (2.5 volumes 200 proof ethanol added, frozen at -80°C for 30 min, spun at 9250 rpm at 4°C for 30 minutes) and the pellet was dried and resuspended in dd$H_2O$. This RNA was the 3′ component of the ligation.

Annealing mixtures were made containing 9 μM 5′ component (oxy, U16 $R_P$ thiosubstituted, or U16 $S_P$ thiosubstituted, 9 μM DNA splint, 10 μM 3′ component, 1X T4 RNA ligase 2 buffer (50 mM Tris-HCl, 2 mM $MgCl_2$, 1 mM DTT, 400 μM ATP, pH 7.5), and 160 mM sodium chloride. These mixtures were heated to 90°C for 5 minutes and slow-cooled to room temperature. Once cooled, 10 U T4 RNA ligase 2 (New England BioLabs) was added and the reactions were incubated overnight at room temperature (approximately 23°). Full length ligation product was separated by 12% denaturing PAGE (7.5 M urea). This band was excised, crushed and soaked overnight in 300 mM sodium acetate, pH 5.3, and gel slice was filtered off the next morning. An ethanol precipitation was performed (2.5 volumes 200 proof ethanol added, frozen at -80°C for 30 min, spun at 9250 rpm at 4°C for 30 minutes) and the pellet was dried and resuspended in dd$H_2O$.

### 7.4.5 $K_d$ determination of RNAs for ZMP by gel shift

Natively purified RNAs were buffer exchanged into an RNA-folding solution containing 50 mM HEPES-KOH, pH 7.5, 200 mM KCl, and 10 mM $MgCl_2$. Varied concentrations of RNA were mixed with trace quantities of [$^{32}$P]-ZMP and a 1X sucrose loading dye contain-

ing xylene cyanol and bromophenol blue. These tubes were equilibrated overnight at room temperature. The next day, samples were electrophoresed on a 12% native polyacrylamide gel at 4°C for approximately one hour, with buffer recirculation every 20 minutes. The gel and running buffer both contained 10 mM $MgCl_2$. The gel was then dried under vacuum and exposed to a phosphor screen overnight. The screen was then phosphorimaged and ImageQuant was used to quantify the intensity of the upper and lower bands (RNA-ligand complex and free ligand, respectively). The fraction bound was determined by the fraction of intensity in the top band. The data were fit in GraphPad Prism to the following equation:

$$F = \frac{B_{max} * X}{K_d + X} \tag{5}$$

where F= fraction bound, $B_{max}$ = fraction bound at saturation, and X= riboswitch concentration. $B_{max}$ was always allowed to float.

### 7.4.6 $K_d$ determination of RNAs for AMP by competition gel shift

Samples were prepared essentially as the in the previous section, except for the following things: (1) RNA was kept at a constant concentration corresponding to approximately 70-80% bound (150 nM for WT and $R_P$, 1.5 μM for $S_P$) and (2) cold AMP was added in increasing concentrations. The fraction bound was determined by the fraction of intensity in the top band. The data were fit in GraphPad Prism to equation 7. b is defined in equation 6.

$$b = \frac{K_{d_Z} + K_{d_Z} * X}{K_{d_{comp}} + R + C_Z} \tag{6}$$

$$Y = F + (\frac{F_0}{2 * C_Z}) * \sqrt[2]{b - b^2 - 4 * R * C_Z} \tag{7}$$

where $K_{d_Z}$ = dissociation constant of RNA construct for ZMP, X = competitor concentration, $K_{d_{comp}}$ = dissociation constant of RNA construct for competitor, R = riboswitch concentration, $C_Z$ = radiolabeled ZMP concentration, F = fraction bound, and $F_0$ = fraction bound at zero competitor.

# 8 Appendix

## 8.1 RNA sequences

***Sulfobacillus acidophilus* native guanidine-I riboswitch aptamer**

5′ - GUCUAAAGUUUGCUAGGGUUCCGCUAGAUACUAGGUCUGGUCCAAGAG

CA AACGGCUUUCACAAAGCCACACGGAAGGAUAAAAGCCUGGGAGAU-3′

***Sulfobacillus acidophilus* guanidine-I riboswitch aptamer with engineered tandem GU iridium (III) hexammine binding site**

5′ - GUCUAAAGUUUGCUAGGGUUCCGCGUCAUAGGUGGUCUGGUCCAAGAG

CA AACGGCUUUCACAAAGCCACACGGAAGGAUAAAAGCCUGGGAGAU-3′

***Thermoanaerobacter mathranii* PRPP riboswitch aptamer**

5′ - GUGAAAGUGUACCUAGGGUUCCAGCCUAUUUGUAGGUGUUCGGACCGA

GCGGUACAGGUAUAUUUUUAUAUACCACACCUUAGGGACAAAAGCCCGGG

AGGAUAGGUUUCACUCGUA-3′

***Thermoanaerobacter mathranii* G96A ppGpp aptamer**

5′ - GUGAAAGUGUACCUAGGGUUCCAGCCUAUUUGUAGGUGUUCGGACCGA

GCGGUACAGGUAUAUUUUUAUAUACCACACCUUAGGGACAAAAGCCCGAG

AGGAUAGGUUUCACUCGUA-3′

***Pseudomonas aeruginosa* full length mini-*ykkC***

5′ - GGAAGCGGGACGACCCGUUUUCCCUCUUUCAUUGCGCGGGGACGACCCUGC - 3′

***Pseudomonas aeruginosa* mini-*ykkC* 6 bp P1 hairpin**

5′ - AGCGGGACGACCCGUU - 3′

***Pseudomonas aeruginosa* mini-*ykkC* P2 hairpin**

5′ - GCGGGGACGACCCUGC - 3′

***Pseudomonas aeruginosa* mini-*ykkC* P2 hairpin with 5-bromoU substitution**

5′ - GCGGGGACGACCC(5-BrU)GC - 3′

***Clostridium bartlettii* riboswitch aptamer**

5′-GGAAAUGGUCAAGUGACUGGUGGAAACAUCAAAAGAUGCGUAGGAUAAC CUACAGGGAGCUUGAUUAUAUAUUUAGAAAUUGCCGACCGCCUGGGCACG UAUG-3′

***Clostridium bartlettii* U16 phosphorothioate oligonucleotide, nt -1 through 19, 5′ side of ligation**

5′ - GGAAAUGGUCAAGUGAC(*U)GGU - 3′

***Clostridium bartlettii* oligonucleotide, nt 20 through 101, 3′ side of ligation**

5′-GGAAACAUCAAAAGAUGCGUAGGAUAACCUACAGGGAGCUUGAUUAUAU

AUUUAGAAAUUGCCGACCGCCUGGGCACGUAUG-3′

## 8.2 Protein sequences

***Bacillus subtilis*, SAS1 enzyme**

N terminus - MMDDKQWERFLVPYRQAVEELKVKLKGIRTLYEYEDDHSPIEFVTG

RVKPVASILEKARRKSIPLHEIETMQDIAGLRIMCQFVDDIQIVKEMLFARKDFTV

VDQRDYIAEHKESGYRSYHLVVLYPLQTVSGEKHVLVEIQIRTLAMNFWATIEHS

LNYKYSGNIPEKVKLRLQRASEAASRLDEEMSEIRGEVQEAQAAFSRKKKGSEQ

Q - C terminus

# References


[1] Crick FH (1958). On protein synthesis. Symposia of the Society for Experimental Biology *12*, 138–163.

[2] Hoagland MB, Zamecnik PC, and Stephenson ML (1957). Intermediate reactions in protein biosynthesis. Biochimica et Biophysica Acta *24*, 215–216. doi: 10.1016/0006-3002(57)90175-0.

[3] Hoagland MB, Stephenson ML, Scott JF, Hecht LI, and Zamecnik PC (1958). A soluble ribonucleic acid intermediate in protein synthesis. Journal of Biological Chemistry *231*, 241–257.

[4] Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, Penswick JR, and Zamir A (1965). Structure of a ribonucleic acid. Science *147*, 1462–1465. doi:10.1126/science.147.3664.1462.

[5] Madison JT, Everett GA, and Kung H (1966). Nucleotide Sequence of a Yeast Tyrosine Transfer RNA. Science *153*, 531–534. doi:10.1126/science.153.3735.531.

[6] Kim SH, Quigley GJ, Suddath FL, McPherson A, Sneden D, Kim JJ, Weinzierl J, and Rich A (1973). Three-Dimensional Structure of Yeast Phenylalanine Transfer RNA: Folding of the Polynucleotide Chain. Science *179*, 285–288. doi: 10.1126/science.179.4070.285.

[7] Kim SH, Suddath FL, Quigley GJ, McPherson A, Sussman JL, Wang AHJ, Seeman

NC, and Rich A (1974). Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. Science *185*, 435–440. doi:10.1126/science.185.4149.435.

[8] Ladner JE, Jack A, Robertus JD, Brown RS, Rhodes D, Clark BF, and Klug A (1975). Structure of yeast phenylalanine transfer RNA at 2.5 A resolution. PNAS *72*, 4414–4418.

[9] Cech TR, Zaug AJ, and Grabowski PJ (1981). In vitro splicing of the ribosomal RNA precursor of tetrahymena: Involvement of a guanosine nucleotide in the excision of the intervening sequence. Cell *27*, 487–496. doi:10.1016/0092-8674(81)90390-1.

[10] Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, and Cech TR (1982). Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. Cell *31*, 147–157. doi:10.1016/0092-8674(82)90414-7.

[11] Robertson HD, Altman S, and Smith JD (1972). Purification and Properties of a Specific Escherichia coli Ribonuclease which Cleaves a Tyrosine Transfer Ribonucleic Acid Precursor. Journal of Biological Chemistry *247*, 5243–5251.

[12] Guerrier-Takada C, Gardiner K, Marsh T, Pace N, and Altman S (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell *35*, 849–857. doi:10.1016/0092-8674(83)90117-4.

[13] Ban N, Nissen P, Hansen J, Moore PB, and Steitz TA (2000). The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 å Resolution. Science *289*, 905–920. doi:10.1126/science.289.5481.905.

[14] Woese CR, Dugre DH, Saxinger WC, and Dugre SA (1966). The molecular basis for the genetic code. PNAS *55*, 966–974.

[15] Crick FH (1968). The origin of the genetic code. Journal of Molecular Biology *38*, 367–379.

[16] Orgel LE (1968). Evolution of the genetic apparatus. Journal of Molecular Biology *38*, 381–393.

[17] Gilbert W (1986). Origin of life: The RNA world. Nature *319*, 618. doi: 10.1038/319618a0.

[18] Wagner A (2008). Robustness and evolvability: a paradox resolved. Proceedings of the Royal Society of London B: Biological Sciences *275*, 91–100. doi: 10.1098/rspb.2007.1137.

[19] Draghi JA, Parsons TL, Wagner GP, and Plotkin JB (2010). Mutational robustness can facilitate adaptation. Nature *463*, 353–355. doi:10.1038/nature08694.

[20] Wagner GP and Altenberg L (1996). Perspective: Complex Adaptations and the Evolution of Evolvability. Evolution *50*, 967–976. doi:10.1111/j.1558-5646.1996.tb02339.x.

[21] Burch CL and Chao L (2000). Evolvability of an RNA virus is determined by its mutational neighbourhood. Nature *406*, 625–628. doi:10.1038/35020564.

[22] Ekland EH and Bartel DP (1996). RNA-catalysed RNA polymerization using nucleoside triphosphates. Nature *382*, 373. doi:10.1038/382373a0.

[23] Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, and Bartel DP (2001). RNA-Catalyzed RNA Polymerization: Accurate and general RNA-templated primer extension. Science *292*, 1319–1325. doi:10.1126/science.1060786.

[24] Wochner A, Attwater J, Coulson A, and Holliger P (2011). Ribozyme-catalyzed transcription of an active ribozyme. Science *332*, 209–212. doi: 10.1126/science.1200752.

[25] Jenison RD, Gill SC, Pardi A, and Polisky B (1994). High-resolution molecular discrimination by RNA. Science *263*, 1425–1429. doi:10.1126/science.7510417.

[26] Tuerk C and Gold L (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505–510. doi:10.1126/science.2200121.

[27] Porter EB, Polaski JT, Morck MM, and Batey RT (2017). Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. Nature Chemical Biology *13*, 295–301. doi:10.1038/nchembio.2278.

[28] Mandal M, Boese B, Barrick JE, Winkler WC, and Breaker RR (2003). Riboswitches control fundamental biochemical pathways in *bacillus subtilis* and other bacteria. Cell *113*, 577–586. doi:10.1016/S0092-8674(03)00391-X.

[29] Barrick JE, Corbino KA, Winkler WC, Nahvi A, Mandal M, Collins J, Lee M, Roth A, Sudarsan N, Jona I, Wickiser JK, and Breaker RR (2004). New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. PNAS *101*, 6421–6426. doi:10.1073/pnas.0308014101.

[30] Winkler WC and Breaker RR (2003). Genetic control by metabolite-binding riboswitches. ChemBioChem *4*, 1024–1032. doi:10.1002/cbic.200300685.

[31] Mandal M and Breaker RR (2004). Gene regulation by riboswitches. Nature Reviews Molecular Cell Biology *5*, 451–463. doi:10.1038/nrm1403.

[32] Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, and Breaker RR (2002). Genetic control by a metabolite binding mRNA. Chemistry & Biology *9*, 1043–1049. doi:10.1016/S1074-5521(02)00224-7.

[33] Winkler WC, Cohen-Chalamish S, and Breaker RR (2002). An mRNA structure that controls gene expression by binding FMN. PNAS *99*, 15908–15913. doi:10.1073/pnas.212628899.

[34] Mandal M and Breaker RR (2004). Adenine riboswitches and gene activation by disruption of a transcription terminator. Nature Structural & Molecular Biology *11*, 29–35. doi:10.1038/nsmb710.

[35] Mandal M, Lee M, Barrick JE, Weinberg Z, Emilsson GM, Ruzzo WL, and Breaker RR (2004). A Glycine-Dependent Riboswitch That Uses Cooperative Binding to Control Gene Expression. Science *306*, 275–279. doi:10.1126/science.1100829.

[36] Corbino KA, Barrick JE, Lim J, Welz R, Tucker BJ, Puskarz I, Mandal M, Rudnick ND, and Breaker RR (2005). Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. Genome Biology *6*, R70. doi:10.1186/gb-2005-6-8-r70.

[37] Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH, and Breaker RR (2008). Riboswitches in eubacteria sense the second messenger cyclic di-GMP. Science *321*, 411–413. doi:10.1126/science.1159519.

[38] Baker JL, Sudarsan N, Weinberg Z, Roth A, Stockbridge RB, and Breaker RR (2012). Widespread genetic switches and toxicity resistance proteins for fluoride. Science *335*, 233–235. doi:10.1126/science.1215063.

[39] Kim PB, Nelson JW, and Breaker RR (2015). An Ancient Riboswitch Class in Bacteria Regulates Purine Biosynthesis and One-Carbon Metabolism. Molecular Cell *57*, 317–328. doi:10.1016/j.molcel.2015.01.001.

[40] Nelson JW, Atilho RM, Sherlock ME, Stockbridge RB, and Breaker RR (2017). Metabolism of free guanidine in bacteria is regulated by a widespread riboswitch class. Molecular Cell *65*, 220–230. doi:10.1016/j.molcel.2016.11.019.

[41] Stockbridge RB, Robertson JL, Kolmakova-Partensky L, and Miller C (2013). A family of fluoride-specific ion channels with dual-topology architecture. eLife *2*. doi:10.7554/eLife.01084.

[42] Ji C, Stockbridge RB, and Miller C (2014). Bacterial fluoride resistance, Fluc channels, and the weak acid accumulation effect. The Journal of General Physiology *144*, 257–261. doi:10.1085/jgp.201411243.

[43] Sudarsan N, Hammond MC, Block KF, Welz R, Barrick JE, Roth A, and Breaker RR (2006). Tandem Riboswitch Architectures Exhibit Complex Gene Control Functions. Science *314*, 300–304. doi:10.1126/science.1130716.

[44] Welz R and Breaker RR (2007). Ligand binding and gene control characteristics of tandem riboswitches in Bacillus anthracis. RNA *13*, 573–582. doi: 10.1261/rna.407707.

[45] Sherlock ME, Sudarsan N, Stav S, and Breaker RR (2018). Tandem riboswitches form a natural Boolean logic gate to control purine metabolism in bacteria. eLife *7*, e33908. doi:10.7554/eLife.33908.

[46] Winkler WC and Breaker RR (2005). Regulation of bacterial gene expression by riboswitches. Annual Review of Microbiology *59*, 487–517. doi: 10.1146/annurev.micro.59.030804.121336.

[47] Breaker RR (2012). Riboswitches and the RNA world. Cold Spring Harbor Perspectives in Biology *4*, a003566–a003566. doi:10.1101/cshperspect.a003566.

[48] Farnham PJ and Platt T (1981). Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription in vitro. Nucleic Acids Research *9*, 563–577.

[49] Wilson KS and Hippel PHv (1995). Transcription termination at intrinsic terminators: the role of the RNA hairpin. PNAS *92*, 8793–8797.

[50] Mondal S, Yakhnin AV, Sebastian A, Albert I, and Babitzke P (2016). NusA-dependent transcription termination prevents misregulation of global gene expression. Nature Microbiology *1*, 15007. doi:10.1038/nmicrobiol.2015.7.

[51] Toulokhonov I, Artsimovitch I, and Landick R (2001). Allosteric Control of RNA

Polymerase by a Site That Contacts Nascent RNA Hairpins. Science *292*, 730–733. doi:10.1126/science.1057738.

[52] Gusarov I and Nudler E (2001). Control of Intrinsic Transcription Termination by N and NusA: The Basic Mechanisms. Cell *107*, 437–449. doi:10.1016/S0092-8674(01)00582-7.

[53] Ha KS, Toulokhonov I, Vassylyev DG, and Landick R (2010). The NusA N-Terminal Domain Is Necessary and Sufficient for Enhancement of Transcriptional Pausing via Interaction with the RNA Exit Channel of RNA Polymerase. Journal of Molecular Biology *401*, 708–725. doi:10.1016/j.jmb.2010.06.036.

[54] Guo X, Myasnikov AG, Chen J, Crucifix C, Papai G, Takacs M, Schultz P, and Weixlbaumer A (2018). Structural Basis for NusA Stabilized Transcriptional Pausing. Molecular Cell *69*, 816–827.e4. doi:10.1016/j.molcel.2018.02.008.

[55] McDaniel BAM, Grundy FJ, Artsimovitch I, and Henkin TM (2003). Transcription termination control of the S box system: Direct measurement of S-adenosylmethionine by the leader RNA. PNAS *100*, 3083–3088. doi: 10.1073/pnas.0630422100.

[56] Shine J and Dalgarno L (1974). The 3′-terminal sequence of *escherichia* coli 16s ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. PNAS *71*, 1342–1346.

[57] Steitz JA and Jakes K (1975). How ribosomes select initiator regions in mRNA: base

pair formation between the 3' terminus of 16s rRNA and the mRNA during initiation of protein synthesis in Escherichia coli. PNAS *72*, 4734–4738.

[58] Weinberg Z, Nelson JW, Lünse CE, Sherlock ME, and Breaker RR (2017). Bioinformatic analysis of riboswitch structures uncovers variant classes with altered ligand specificity. PNAS *114*, E2077–E2085. doi:10.1073/pnas.1619581114.

[59] Serganov A, Yuan YR, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, and Patel DJ (2004). Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. Chemistry & Biology *11*, 1729–1741. doi:10.1016/j.chembiol.2004.11.018.

[60] Nelson JW, Sudarsan N, Phillips GE, Stav S, Lünse CE, McCown PJ, and Breaker RR (2015). Control of bacterial exoelectrogenesis by c-AMP-GMP. PNAS *112*, 5389–5394. doi:10.1073/pnas.1419264112.

[61] Smith KD, Lipchock SV, Livingston AL, Shanahan CA, and Strobel SA (2010). Structural and biochemical determinants of ligand binding by the c-di-GMP riboswitch,. Biochemistry *49*, 7351–7359. doi:10.1021/bi100671e.

[62] Ren A, Wang XC, Kellenberger CA, Rajashankar KR, Jones RA, Hammond MC, and Patel DJ (2015). Structural basis for molecular discrimination by a 3′,3′-cGAMP sensing riboswitch. Cell Reports *11*, 1–12. doi:10.1016/j.celrep.2015.03.004.

[63] Grundy FJ and Henkin TM (1998). The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in

Gram-positive bacteria. Molecular Microbiology *30*, 737–749. doi:10.1046/j.1365-2958.1998.01105.x.

[64] Epshtein V, Mironov AS, and Nudler E (2003). The riboswitch-mediated control of sulfur metabolism in bacteria. PNAS *100*, 5052–5056. doi: 10.1073/pnas.0531307100.

[65] Fuchs RT, Grundy FJ, and Henkin TM (2006). The $S_{mk}$ box is a new SAM-binding RNA for translational regulation of SAM synthetase. Nature Structural & Molecular Biology *13*, 226–233. doi:10.1038/nsmb1059.

[66] Weinberg Z, Regulski EE, Hammond MC, Barrick JE, Yao Z, Ruzzo WL, and Breaker RR (2008). The aptamer core of SAM-IV riboswitches mimics the ligand-binding site of SAM-I riboswitches. RNA *14*, 822–828. doi:10.1261/rna.988608.

[67] Poiata E, Meyer MM, Ames TD, and Breaker RR (2009). A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria. RNA *15*, 2046–2056. doi:10.1261/rna.1824209.

[68] Gayan MA, Sherlock ME, Weinberg Z, and Breaker RR (2017). SAM-VI RNAs selectively bind S-adenosylmethionine and exhibit similarities to SAM-III riboswitches. RNA Biology *0*, 1–8. doi:10.1080/15476286.2017.1399232.

[69] Loenen WaM (2006). S-adenosylmethionine: jack of all trades and master of everything? Biochemical Society Transactions *34*, 330–333. doi:10.1042/BST0340330.

[70] Montange RK and Batey RT (2006). Structure of the *S*-adenosylmethionine

riboswitch regulatory mRNA element. Nature *441*, 1172–1175. doi: 10.1038/nature04819.

[71] Gilbert SD, Rambo RP, Van Tyne D, and Batey RT (2008). Structure of the SAM-II riboswitch bound to S-adenosylmethionine. Nature Structural & Molecular Biology *15*, 177–182. doi:10.1038/nsmb.1371.

[72] Lu C, Smith AM, Fuchs RT, Ding F, Rajashankar K, Henkin TM, and Ke A (2008). Crystal structures of the SAM-III/$S_{mk}$ riboswitch reveal the SAM-dependent translation inhibition mechanism. Nature Structural & Molecular Biology *15*, 1076–1083. doi:10.1038/nsmb.1494.

[73] Trausch JJ, Xu Z, Edwards AL, Reyes FE, Ross PE, Knight R, and Batey RT (2014). Structural basis for diversity in the SAM clan of riboswitches. PNAS *111*, 6624–6629. doi:10.1073/pnas.1312918111.

[74] McCown PJ, Corbino KA, Stav S, Sherlock ME, and Breaker RR (2017). Riboswitch diversity and distribution. RNA *23*, 995–1011. doi:10.1261/rna.061234.117.

[75] Hujer AM, Hujer KM, and Bonomo RA (2001). Mutagenesis of amino acid residues in the SHV-1 -lactamase: the premier role of Gly238ser in penicillin and cephalosporin resistance. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology *1547*, 37–50. doi:10.1016/S0167-4838(01)00164-9.

[76] Kurokawa H, Yagi T, Shibata N, Shibayama K, Kamachi K, and Arakawa Y (2000). A New SHV-Derived Extended-Spectrum -Lactamase (SHV-24) That Hydrolyzes

Ceftazidime through a Single-Amino-Acid Substitution (D179g) in the -Loop. Antimicrobial Agents and Chemotherapy *44*, 1725–1727. doi:10.1128/AAC.44.6.1725-1727.2000.

[77] Wachino Ji, Doi Y, Yamane K, Shibata N, Yagi T, Kubota T, and Arakawa Y (2004). Molecular characterization of a cephamycin-hydrolyzing and inhibitor-resistant class A β-lactamase, GES-4, possessing a single G170s substitution in the ω-loop. Antimicrobial Agents and Chemotherapy *48*, 2905–2910. doi: 10.1128/AAC.48.8.2905-2910.2004.

[78] Yi H, Choi JM, Hwang J, Prati F, Cao TP, Lee SH, and Kim HS (2016). High adaptability of the omega loop underlies the substrate-spectrum-extension evolution of a class A -lactamase, PenL. Scientific Reports *6*, 36527. doi:10.1038/srep36527.

[79] Dodson G and Wlodawer A (1998). Catalytic triads and their relatives. Trends in Biochemical Sciences *23*, 347–352. doi:10.1016/S0968-0004(98)01254-7.

[80] Reiss CW, Xiong Y, and Strobel SA (2017). Structural basis for ligand binding to the guanidine-I riboswitch. Structure *25*, 195–202. doi:10.1016/j.str.2016.11.020.

[81] Meyer MM, Hammond MC, Salinas Y, Roth A, Sudarsan N, and Breaker RR (2011). Challenges of ligand identification for riboswitch candidates. RNA Biology *8*, 5–10. doi:10.4161/rna.8.1.13865.

[82] Ren A, Rajashankar KR, and Patel DJ (2012). Fluoride ion encapsulation by Mg2+ ions and phosphates in a fluoride riboswitch. Nature *486*, 85–89. doi: 10.1038/nature11152.

[83] Stockbridge RB, Koide A, Miller C, and Koide S (2014). Proof of dual-topology architecture of Fluc F channels with monobody blockers. Nature Communications *5*, 5120. doi:10.1038/ncomms6120.

[84] Stockbridge RB, Kolmakova-Partensky L, Shane T, Koide A, Koide S, Miller C, and Newstead S (2015). Crystal structures of a double-barrelled fluoride ion channel. Nature *525*, 548–551. doi:10.1038/nature14981.

[85] Kermani AA, Macdonald CB, Gundepudi R, and Stockbridge RB (2018). Guanidinium export is the primal function of SMR family transporters. PNAS 201719187. doi:10.1073/pnas.1719187115.

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

[87] Caldara M, Dupont G, Leroy F, Goldbeter A, Vuyst LD, and Cunin R (2008). Arginine biosynthesis in *escherichia coli* experimental perturbation and mathematical modeling. Journal of Biological Chemistry *283*, 6347–6358. doi: 10.1074/jbc.M705884200.

[88] Hamana K (1996). Distribution of diaminopropane and acetylspermidine in Enterobacteriaceae. Canadian Journal of Microbiology *42*, 107–114. doi:10.1139/m96-017.

[89] Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M,

Hirai K, Hoque A, Ho PY, Kakazu Y, Sugawara K, Igarashi S, Harada S, Masuda T, Sugiyama N, Togashi T, Hasegawa M, Takai Y, Yugi K, Arakawa K, Iwata N, Toya Y, Nakayama Y, Nishioka T, Shimizu K, Mori H, and Tomita M (2007). Multiple high-throughput analyses monitor the response of *e. coli* to perturbations. Science *316*, 593–597. doi:10.1126/science.1132067.

[90] Keel AY, Rambo RP, Batey RT, and Kieft JS (2007). A general strategy to solve the phase problem in RNA crystallography. Structure *15*, 761–772. doi: 10.1016/j.str.2007.06.003.

[91] Weinberg Z, Barrick JE, Yao Z, Roth A, Kim JN, Gore J, Wang JX, Lee ER, Block KF, Sudarsan N, Neph S, Tompa M, Ruzzo WL, and Breaker RR (2007). Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. Nucleic Acids Research *35*, 4809–4819. doi:10.1093/nar/gkm487.

[92] Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH, and Breaker RR (2010). Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. Genome Biology *11*, R31. doi:10.1186/gb-2010-11-3-r31.

[93] Perrin DD (1965). *Dissociation Constants of Organic Bases in Aqueous Solution*. Butterworths, London.

[94] Dougherty DA (2013). The cation-$\pi$ interaction. Accounts of chemical research *46*, 885–893. doi:10.1021/ar300265y.

[95] Gallivan JP and Dougherty DA (1999). Cation-$\pi$ interactions in structural biology. PNAS *96*, 9459–9464. doi:10.1073/pnas.96.17.9459.

[96] Zhang H, Li C, Yang F, Su J, Tan J, Zhang X, and Wang C (2014). Cation-pi interactions at non-redundant protein–RNA interfaces. Biochemistry. Biokhimiia *79*, 643–652. doi:10.1134/S0006297914070062.

[97] Szewczak AA, Moore PB, Chang YL, and Wool IG (1993). The conformation of the sarcin/ricin loop from 28s ribosomal RNA. PNAS *90*, 9581–9585.

[98] Correll CC, Munishkin A, Chan YL, Ren Z, Wool IG, and Steitz TA (1998). Crystal structure of the ribosomal RNA domain essential for binding elongation factors. PNAS *95*, 13436–13441. doi:10.1073/pnas.95.23.13436.

[99] Hendrix DK, Brenner SE, and Holbrook SR (2005). RNA structural motifs: building blocks of a modular biomolecule. Quarterly Reviews of Biophysics *38*, 221–243. doi:10.1017/S0033583506004215.

[100] Lustig B, Arora S, and Jernigan RL (1997). RNA base-amino acid interaction strengths derived from structures and sequences. Nucleic Acids Research *25*, 2562–2565. doi:10.1093/nar/25.13.2562.

[101] Leslie E O (2004). Prebiotic Chemistry and the Origin of the RNA World. Critical Reviews in Biochemistry and Molecular Biology *39*, 99–123. doi: 10.1080/10409230490460765.

[102] Strobel SA (2001). Biological catalysis: Repopulating the RNA world. Nature *411*, 1003. doi:10.1038/35082661.

[103] Kirschner M and Gerhart J (1998). Evolvability. PNAS *95*, 8420–8427.

[104] Ancel LW and Fontana W (2000). Plasticity, evolvability, and modularity in RNA. Journal of Experimental Zoology *288*, 242–283. doi:10.1002/1097-010X(20001015)288:3¡242::AID-JEZ5¿3.0.CO;2-O.

[105] Banerjee S, Pieper U, Kapadia G, Pannell LK, and Herzberg O (1998). Role of the ω-loop in the activity, substrate specificity, and structure of class A β-lactamase,. Biochemistry *37*, 3286–3296. doi:10.1021/bi972127f.

[106] Kellenberger CA, Wilson SC, Hickey SF, Gonzalez TL, Su Y, Hallberg ZF, Brewer TF, Iavarone AT, Carlson HK, Hsieh YF, and Hammond MC (2015). GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP. PNAS *112*, 5383–5388. doi:10.1073/pnas.1419328112.

[107] Ren A, Rajashankar KR, and Patel DJ (2015). Global RNA fold and molecular recognition for a *pfl* riboswitch bound to ZMP, a master regulator of one-carbon metabolism. Structure *23*, 1375–1381. doi:10.1016/j.str.2015.05.016.

[108] Smith KD, Lipchock SV, Ames TD, Wang J, Breaker RR, and Strobel SA (2009). Structural basis of ligand binding by a c-di-GMP riboswitch. Nature Structural & Molecular Biology *16*, 1218–1223. doi:10.1038/nsmb.1702.

[109] Ebbole DJ and Zalkin H (1987). Cloning and characterization of a 12-gene cluster from Bacillus subtilis encoding nine enzymes for de novo purine nucleotide synthesis. Journal of Biological Chemistry *262*, 8274–8287.

[110] Ebbole DJ and Zalkin H (1989). Bacillus subtilis pur operon expression and regulation. Journal of Bacteriology *171*, 2136–2141. doi:10.1128/jb.171.4.2136-2141.1989.

[111] Cashel M and Gallant J (1969). Two compounds implicated in the function of the RC gene of *escherichia coli*. Nature *221*, 838. doi:10.1038/221838a0.

[112] Dalebroux ZD and Swanson MS (2012). ppGpp: magic beyond RNA polymerase. Nature Reviews Microbiology *10*, 203–212. doi:10.1038/nrmicro2720.

[113] Gaca AO, Colomer-Winter C, and Lemos JA (2015). Many means to a common end: the intricacies of (p)ppGpp metabolism and its control of bacterial homeostasis. Journal of Bacteriology *197*, 1146–1156. doi:10.1128/JB.02577-14.

[114] Hove-Jensen B, Andersen KR, Kilstrup M, Martinussen J, Switzer RL, and Willemoës M (2017). Phosphoribosyl diphosphate (PRPP): biosynthesis, enzymology, utilization, and metabolic significance. Microbiology and Molecular Biology Reviews *81*, e00040–16. doi:10.1128/MMBR.00040-16.

[115] Hove-Jensen B (1988). Mutation in the phosphoribosylpyrophosphate synthetase gene (prs) that results in simultaneous requirements for purine and pyrimidine nucleosides, nicotinamide nucleotide, histidine, and tryptophan in Escherichia coli. Journal of Bacteriology *170*, 1148–1152. doi:10.1128/jb.170.3.1148-1152.1988.

[116] Jiménez A, Santos MA, and Revuelta JL (2008). Phosphoribosyl pyrophosphate synthetase activity affects growth and riboflavin production in Ashbya gossypii. BMC Biotechnology *8*, 67. doi:10.1186/1472-6750-8-67.

[117] White RH (1996). Biosynthesis of Methanopterin. Biochemistry *35*, 3447–3456. doi:10.1021/bi952308m.

[118] Cashel M (1970). Inhibition of RNA polymerase by ppGpp, a nucleotide accumulated during the stringent response to amino acid starvation in *e. coli*. Cold Spring Harbor Symposia on Quantitative Biology *35*, 407–413. doi: 10.1101/SQB.1970.035.01.052.

[119] O'Farrell PH (1978). The suppression of defective translation by ppGpp and its role in the stringent response. Cell *14*, 545–557. doi:10.1016/0092-8674(78)90241-6.

[120] Potrykus K and Cashel M (2008). (p)ppGpp: Still Magical? Annual Review of Microbiology *62*, 35–51. doi:10.1146/annurev.micro.62.081307.162903.

[121] Paul BJ, Berkmen MB, and Gourse RL (2005). DksA potentiates direct activation of amino acid promoters by ppGpp. PNAS *102*, 7823–7828. doi: 10.1073/pnas.0501170102.

[122] Ryals J, Little R, and Bremer H (1982). Control of rRNA and tRNA syntheses in Escherichia coli by guanosine tetraphosphate. Journal of Bacteriology *151*, 1261–1268.

[123] van Ooyen AJJ, Gruber M, and Jørgensen P (1976). The mechanism of action of ppGpp on rRNA synthesis in vitro. Cell *8*, 123–128. doi:10.1016/0092-8674(76)90193-8.

[124] Meola M, Yamen B, Weaver K, and Sandwick RK (2003). The catalytic effect of Mg2+ and imidazole on the decomposition of 5-phosphoribosyl--1-pyrophosphate

in aqueous solution. Journal of Inorganic Biochemistry *93*, 235–242. doi: 10.1016/S0162-0134(02)00578-0.

[125] Dennis AL, Puskas M, Stasaitis S, and Sandwick RK (2000). The formation of a 1–5 phosphodiester linkage in the spontaneous breakdown of 5-phosphoribosyl--1-pyrophosphate. Journal of Inorganic Biochemistry *81*, 73–80. doi:10.1016/S0162-0134(00)00117-3.

[126] Kornberg A, Lieberman I, and Simms ES (1955). Enzymatic synthesis and properties of 5-phosphoribosylpyrophosphate. J Biol Chem *215*, 389–402.

[127] Khorana H, Fernandes J, and Kornberg A (1958). Pyrophosphorylation of ribose 5-phosphate in the enzymatic synthesis of 5-phosphorylribose 1-pyrophosphate. Journal of Biological Chemistry *230*, 941–948.

[128] Remy CN, Remy WT, and Buchanan JM (1955). Biosynthesis of the purines VIII. Enzymatic synthesis and utilization of phosphoribosylpyrophosphate. Journal of Biological Chemistry *217*, 885–896.

[129] Schindelin H, Zhang M, Bald R, Fürste JP, Erdmann VA, and Heinemann U (1995). Crystal Structure of an RNA Dodecamer Containing theEscherichia coliShine-Dalgarno Sequence. Journal of Molecular Biology *249*, 595–603. doi: 10.1006/jmbi.1995.0321.

[130] Thompson RE, Li ELF, Spivey HO, Chandler JP, Katz AJ, and Appleman JR (1978). Apparent stability constants of H+ and Mg2+ complexes of 5-phosphoribosyl -1-pyrophosphate. Bioinorganic chemistry *9*, 35–45.

[131] Evans GL, Gamage SA, Bulloch EMM, Baker EN, Denny WA, and Lott JS (2014). Repurposing the Chemical Scaffold of the Anti-Arthritic Drug Lobenzarit to Target Tryptophan Biosynthesis in *Mycobacterium tuberculosis*. ChemBioChem *15*, 852–864. doi:10.1002/cbic.201300628.

[132] Héroux A, White E, Ross LJ, Kuzin AP, and Borhani DW (2000). Substrate Deformation in a Hypoxanthine-Guanine Phosphoribosyltransferase Ternary Complex. Structure *8*, 1309–1318. doi:10.1016/S0969-2126(00)00546-3.

[133] González-Segura L, Witte JF, McClard RW, and Hurley TD (2007). Ternary Complex Formation and Induced Asymmetry in Orotate Phosphoribosyltransferase ‡. Biochemistry *46*, 14075–14086. doi:10.1021/bi701023z.

[134] Hancock RD, Siddons CJ, Oscarson KA, and Reibenspies JM (2004). The structure of the 11-coordinate barium complex of the pendant-donor macrocycle 1,4,7,10-tetrakis(carbamoylmethyl)-1,4,7,10-tetraazacyclododecane: an analysis of the coordination numbers of barium(II) in its complexes. Inorganica Chimica Acta *357*, 723–727. doi:10.1016/j.ica.2003.06.016.

[135] Egli M and Sarkhel S (2007). Lone pair aromatic interactions: To stabilize or not to stabilize. Accounts of chemical research *40*, 197–205.

[136] Sarkhel S and Desiraju GR (2003). NH...O, OH...O, and CH...O hydrogen bonds in protein-ligand complexes: Strong and weak interactions in molecular recognition. Proteins: Structure, Function, and Bioinformatics *54*, 247–259. doi:10.1002/prot.10567.

[137] Ran J and Hobza P (2009). On the nature of bonding in lone pair -electron complexes: CCSD (T)/complete basis set limit calculations. Journal of chemical theory and computation *5*, 1180–1185.

[138] Chawla M, Chermak E, Zhang Q, Bujnicki JM, Oliva R, and Cavallo L (2017). Occurrence and stability of lone pair– stacking interactions between ribose and nucleobases in functional RNAs. Nucleic Acids Research doi:10.1093/nar/gkx757.

[139] Singh SK and Das A (2015). The n * interaction: a rapidly emerging non-covalent interaction. Phys. Chem. Chem. Phys. *17*, 9596–9612. doi:10.1039/C4CP05536E.

[140] Schneider DA and Gourse RL (2004). Relationship between Growth Rate and ATP Concentration in *Escherichia coli*: A BIOASSAY FOR AVAILABLE CELLULAR ATP. Journal of Biological Chemistry *279*, 8262–8268. doi: 10.1074/jbc.M311996200.

[141] Yaginuma H, Kawai S, Tabata KV, Tomiyama K, Kakizuka A, Komatsuzaki T, Noji H, and Imamura H (2015). Diversity in ATP concentrations in a single bacterial cell population revealed by quantitative single-cell imaging. Scientific Reports *4*. doi:10.1038/srep06522.

[142] Saxild HH and Nygaard P (1991). Regulation of levels of purine biosynthetic enzymes in Bacillus subtilis: effects of changing purine nucleotide pools. Journal of General Microbiology *137*, 2387–2394. doi:10.1099/00221287-137-10-2387.

[143] Jendresen CB, Kilstrup M, and Martinussen J (2011). A simplified method for rapid quantification of intracellular nucleoside triphosphates by one-dimensional

thin-layer chromatography. Analytical Biochemistry *409*, 249–259. doi: 10.1016/j.ab.2010.10.029.

[144] Nygaard P and Smith JM (1993). Evidence for a novel glycinamide ribonucleotide transformylase in Escherichia coli. Journal of Bacteriology *175*, 3591–3597. doi: 10.1128/jb.175.11.3591-3597.1993.

[145] Jensen KF, Houlberg U, and Nygaard P (1979). Thin-layer chromatographic methods to isolate 32p-labeled 5-phosphoribosyl--1-pyrophosphate (PRPP): Determination of cellular PRPP pools and assay of PRPP synthetase activity. Analytical Biochemistry *98*, 254–263. doi:10.1016/0003-2697(79)90138-6.

[146] Jørgensen CM, Fields CJ, Chander P, Watt D, Burgner JW, Smith JL, and Switzer RL (2008). pyr RNA binding to the Bacillus caldolyticus PyrR attenuation protein - characterization and regulation by uridine and guanosine nucleotides: RNA binding to PyrR. FEBS Journal *275*, 655–670. doi:10.1111/j.1742-4658.2007.06227.x.

[147] Bera AK, Zhu J, Zalkin H, and Smith JL (2003). Functional Dissection of the Bacillus subtilis pur Operator Site. Journal of Bacteriology *185*, 4099–4109. doi: 10.1128/JB.185.14.4099-4109.2003.

[148] Dawson RMC (1959). *Data for Biochemical Research*. Oxford: Clarendon Press.

[149] Kanjee U, Gutsche I, Alexopoulos E, Zhao B, Bakkouri ME, Thibault G, Liu K, Ramachandran S, Snider J, Pai EF, and Houry WA (2011). Linkage between the bacterial acid stress and stringent responses: the structure of the inducible lysine decarboxylase. The EMBO Journal *30*, 931–944. doi:10.1038/emboj.2011.5.

[150] Correll CC, Wool IG, and Munishkin A (1999). The two faces of the Escherichia coli 23 S rRNA sarcin/ricin domain: the structure at 1.11 å resolution11edited by D. E. Draper. Journal of Molecular Biology *292*, 275–287. doi:10.1006/jmbi.1999.3072.

[151] Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, StClair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, Houk KN, Michael FE, and Baker D (2010). Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science *329*, 309–313. doi:10.1126/science.1190239.

[152] Zaug AJ and Cech TR (1986). The intervening sequence RNA of Tetrahymena is an enzyme. Science *231*, 470–475. doi:10.1126/science.3941911.

[153] Piccirilli JA, McConnell TS, Zaug AJ, Noller HF, and Cech TR (1992). Aminoacyl esterase activity of the Tetrahymena ribozyme. Science *256*, 1420–1424. doi: 10.1126/science.1604316.

[154] Lau MWL and Ferré-D'Amaré AR (2016). Many Activities, One Structure: Functional Plasticity of Ribozyme Folds. Molecules *21*, 1570. doi: 10.3390/molecules21111570.

[155] Glansdorff N, Xu Y, and Labedan B (2008). The Last Universal Common Ancestor: emergence, constitution and genetic legacy of an elusive forerunner. Biology Direct *3*, 29. doi:10.1186/1745-6150-3-29.

[156] Akouche M, Jaber M, Maurel MC, Lambert JF, and Georgelin T (2017). Phosphoribosyl pyrophosphate: a molecular vestige of the origin of life on minerals. Angewandte Chemie *129*, 8028–8031. doi:10.1002/ange.201702633.

[157] Reiss CW and Strobel SA (2017). Structural basis for ligand binding to the guanidine-II riboswitch. RNA *23*, 1338–1343. doi:10.1261/rna.061804.117.

[158] Sherlock ME, Malkowski SN, and Breaker RR (2017). Biochemical validation of a second guanidine riboswitch class in bacteria. Biochemistry *56*, 352–358. doi: 10.1021/acs.biochem.6b01270.

[159] Battaglia RA, Price IR, and Ke A (2017). Structural basis for guanidine sensing by the ykkC family of riboswitches. RNA *23*, 578–585. doi:10.1261/rna.060186.116.

[160] Blanco F, Kelly B, Sánchez-Sanz G, Trujillo C, Alkorta I, Elguero J, and Rozas I (2013). Non-covalent interactions: complexes of guanidinium with DNA and RNA nucleobases. The Journal of Physical Chemistry B *117*, 11608–11616. doi: 10.1021/jp407339v.

[161] Sherlock ME and Breaker RR (2017). Biochemical balidation of a third guanidine riboswitch class in bacteria. Biochemistry *56*, 359–363. doi: 10.1021/acs.biochem.6b01271.

[162] Huang L, Wang J, Wilson TJ, and Lilley DMJ (2017). Structure of the Guanidine III Riboswitch. Cell Chemical Biology *24*, 1407–1415.e2. doi: 10.1016/j.chembiol.2017.08.021.

[163] Bochner BR and Ames BN (1982). ZTP (5-amino 4-imidazole carboxamide riboside 5'-triphosphate): a proposed alarmone for 10-formyl-tetrahydrofolate deficiency. Cell *29*, 929–937.

[164] Rohlman CE and Matthews RG (1990). Role of purine biosynthetic intermediates in response to folate stress in Escherichia coli. Journal of bacteriology *172*, 7200–7210.

[165] Batey RT and Kieft JS (2016). Soaking hexammine cations into RNA crystals to obtain derivatives for phasing diffraction data. In *Nucleic Acid Crystallography*, Methods in Molecular Biology, 219–232. Humana Press, New York, NY. DOI: 10.1007/978-1-4939-2763-0_14.

[166] Trausch JJ, Marcano-Velázquez JG, Matyjasik MM, and Batey RT (2015). Metal ion-mediated nucleobase recognition by the ZTP riboswitch. Chemistry & Biology *22*, 829–837. doi:10.1016/j.chembiol.2015.06.007.

[167] Jones CP and Ferré-D'Amaré AR (2015). Recognition of the bacterial alarmone ZMP through long-distance association of two RNA subdomains. Nature Structural & Molecular Biology *22*, 679–685. doi:10.1038/nsmb.3073.

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

[169] McSwiggen JA and Cech TR (1989). Stereochemistry of RNA cleavage by the Tetrahymena ribozyme and evidence that the chemical step is not rate-limiting. Science *244*, 679–683. doi:10.1126/science.2470150.

[170] Ruffner DE and Uhlenbeck OC (1990). Thiophosphate interference experiments locate phosphates important for the hammerhead RNA self-cleavage reaction. Nucleic Acids Research *18*, 6025–6029.

[171] Strobel SA and Shetty K (1997). Defining the chemical groups essential for Tetrahymena group I intron function by nucleotide analog interference mapping. PNAS *94*, 2903–2908.

[172] Wang S, Karbstein K, Peracchi A, Beigelman L, and Herschlag D (1999). Identification of the hammerhead ribozyme metal ion binding site responsible for rescue of the deleterious effect of a cleavage site phosphorothioate. Biochemistry *38*, 14363–14378. doi:10.1021/bi9913202.

[173] Basu S and Strobel SA (1999). Thiophilic metal ion rescue of phosphorothioate interference within the Tetrahymena ribozyme P4-P6 domain. RNA *5*, 1399–1407.

[174] Szewczak AA, Kosek AB, Piccirilli JA, and Strobel SA (2002). Identification of an active site ligand for a group I ribozyme catalytic metal ion. Biochemistry *41*, 2516–2525. doi:10.1021/bi011973u.

[175] Pecoraro VL, Hermes JD, and Cleland WW (1984). Stability constants of $Mg^{2+}$ and $Cd^{2+}$ complexes of adenine nucleotides and thionucleotides and rate constants for formation and dissociation of MgATP and MgADP. Biochemistry *23*, 5262–5271.

[176] Eckstein F (1985). Nucleoside Phosphorothioates. Annual Review of Biochemistry *54*, 367–402. doi:10.1146/annurev.bi.54.070185.002055.

[177] Albe KR, Butler MH, and Wright BE (1990). Cellular concentrations of enzymes and their substrates. Journal of Theoretical Biology *143*, 163–195. doi: 10.1016/S0022-5193(05)80266-8.

[178] Salvail-Lacoste A, Di Tomasso G, Piette BL, and Legault P (2013). Affinity purification of T7 RNA transcripts with homogeneous ends using ARiBo and CRISPR tags. RNA *19*, 1003–1014. doi:10.1261/rna.037432.112.

[179] Sanger F, Nicklen S, and Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. PNAS *74*, 5463–5467.

[180] Hartmann RK, editor (2009). *Handbook of RNA Biochemistry*. Wiley-VCH, Weinheim, 1st student ed edition.

[181] Otwinowski Z and Minor W (1997). [20] Processing of X-ray diffraction data collected in oscillation mode. Methods in enzymology *276*, 307–326.

[182] Emsley P and Cowtan K (2004). *Coot* : model-building tools for molecular graphics. Acta Crystallographica Section D Biological Crystallography *60*, 2126–2132. doi: 10.1107/S0907444904019158.

[183] Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, and Wilson KS (2011). Overview of the *CCP* 4 suite and current developments. Acta Crystallographica Section D Biological Crystallography *67*, 235–242. doi:10.1107/S0907444910045749.

[184] Adams PD, Afonine PV, Bunkczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, and Zwart PH (2010).

PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D: Biological Crystallography *66*, 213–221. doi:10.1107/S0907444909052925.

[185] Schrödinger L (2015). The PyMOL Molecular Graphics System, Version 1.8, 2015 .

[186] Switzer RL and Gibson KJ (1978). [1] Phosphoribosylprophosphate synthetase (ribose-5-phosphate pyrophosphokinase) from Salmonella typhimurium. Methods Enzymol *51*, 3–11.

[187] Steinchen W, Schuhmacher JS, Altegoer F, Fage CD, Srinivasan V, Linne U, Marahiel MA, and Bange G (2015). Catalytic mechanism and allosteric regulation of an oligomeric (p)ppGpp synthetase by an alarmone. PNAS *112*, 13348–13353. doi:10.1073/pnas.1505271112.

[188] Reuter JS and Mathews DH (2010). RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics *11*, 129. doi:10.1186/1471-2105-11-129.

[189] Martin CT and Coleman JE (1989). T7 RNA polymerase does not interact with the 5'-phosphate of the initiating nucleotide. Biochemistry *28*, 2760–2762. doi: 10.1021/bi00433a002.