# Abstract

## Investigating the Sequence Requirements for Specific Ligand Recognition in Variant Riboswitches

Kathryn Marie Barth

2024

Recognizing and responding to environmental stimuli is essential for organisms across all domains of life. Riboswitches sense environmental stimuli in the form of a small molecule ligand and then structurally rearrange to modulate gene expression in response. Specific ligand recognition is, therefore, a requirement of a functional riboswitch. This dissertation focuses on relating the primary sequence of a riboswitch to a role in ligand specificity. Unlike structural studies, which are used to identify nucleotides in direct ligand contact, I have taken a mutational approach to determine which nucleotides, when mutated, alter riboswitch function or ligand specificity. I focus on variant riboswitches where a small number of sequence changes can change the ligand specificity. I seek to understand how specificity is achieved in closely related variant systems. Peripheral nucleotides and the expression platform both impact ligand specificity in variant riboswitches as well as nucleotides that directly interact with the ligand. In addition to informing on the sequence requirements for specific ligand interaction, this investigation has led to the discovery of a new *ykkC* variant class as well as the development of potential synthetic pyrimidine-containing cyclic dinucleotide riboswitches.

# Investigating the Sequence Requirements for Specific Ligand Recognition in Variant Riboswitches

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

by
Kathryn Marie Barth

Dissertation Director: Dr. Scott A. Strobel
May 2024

## Table of Contents

# Acknowledgments

My laboratory independent research career began late in college in the organic synthesis lab of David Haines. Although I quickly learned that I am not a synthetic organic chemist, I want to acknowledge David for helping me persevere and keep investigating when I produce brown sludge. This inquisitive spirit has proven incredibly useful in my experiences in lab. John Goss was my second undergraduate mentor, and his support as I developed new methods and tested new equipment was invaluable. He taught me how to build my own experience and immerse myself in interesting research.

From my time at Yale, I would first like to acknowledge my advisor, Scott Strobel. Despite obligations that keep him occupied elsewhere, Scott has always made time for our research in the lab. He encourages us all to have a work-life balance and ensures that we have the tools we need to succeed, both of which have kept my graduate experience incredibly positive.

Thank you as well to my committee members, Ron Breaker and Matt Simon. Ron always has incredible insight into riboswitches and has pushed me to dig into investigations. Additionally, the Breaker Lab has been a valuable resource for their experience and support with reporter assay construction and execution. Matt and his lab have been particularly helpful for data analysis problems. Their willingness to help a graduate student in need is deeply appreciated.

I would also like to acknowledge the Strobel lab. We have worked through a lot of data analysis together, and we always manage to have a bit of fun. In particular, I want to thank my mentors, Chad Torgerson and Caroline Focht (South Caroline), who helped develop the high-throughput assays and analysis that I used throughout my time at Yale. Lori, Alysha, and Sabrina have always been supportive and encouraging. Special thanks to Dave Hiller and Sunitha Nallur, who have been instrumental in the data production and processing pipelines. Dave provided day-to-day support, and Sunitha helped with all bench work for the last three months of this work as I was writing, I could not have done this without them.

Finally, a personal acknowledgment to friends and family who have supported my research journey. From undergraduate friends I made while rowing to colleagues I worked with at Seattle Preparatory School to friends I met here at Yale, I could not have completed graduate school without your support, encouragement, and conversation. My parents encouraged investigation and independence from a young age, which has served me well. My older brother, John, led the way and always had graduate school advice to share. My younger brother, Chris, was a rational voice, keeping me sane along the way. Finally, Bryan, Rowan, and Luna made it through the last few crazy years with me.

I look forward to all the adventures we have next together!

# 1. Introduction

## 1.1 Sensing and responding to the environment

All organisms inherently need to respond to the world around them. This requires a means of sensing the environment and a mechanism to convert an environmental signal into a response. On the cellular level, the environment consists of a collection of small molecules ranging in size from a single atom, such as a fluoride ion, to a neighboring cell. Specifically recognizing a signal, and differentiating it from chemically or structurally similar signals, is necessary to generate a proper cellular response. Specific recognition can be achieved using the chemical properties of the molecular signal, such as polarity or charge, or the size of the signal molecule.

Most often, the cellular response to an environmental signal consists of turning genes on or off [1–3]. Being able to activate or deactivate cellular functions when necessary allows complex cellular systems to operate. Across all kingdoms of life, all organisms have mechanisms in place to increase or decrease gene expression. These systems can act on any part of the gene expression pathway from DNA, to RNA, to proteins [4]. Additionally, these systems utilize both protein and RNA as functional molecules.

Different biological processes require different responses in both stringency and magnitude. Some processes, including cell division or death, are highly regulated and require a specific signal for activation, while other processes, such as purine biosynthesis, control essential metabolites for growth and are, therefore, nearly always active [5,6]. Additionally, some processes, including stress response, activate multiple genes and thus require a broad-acting mechanism, such as the activation of a general transcription factor [1,7]. Other processes are more targeted and activate a single gene in a pathway. Sensing a signal and generating an appropriately tuned response is a key component of life.

## 1.2 Riboswitches

RNA can be used to sense signals and regulate a cellular process in response. This includes regulating transcription, translation, and transcript stability or longevity [8–10]. RNA is a fundamental molecule of cells composed of nucleic acid building blocks. Similar to DNA, RNA bases can pair with their counterparts – adenine (A) to uridine (U) and cytosine (C) to guanine (G) – to form stable structures. These structures are used in the function of these RNAs, including small molecule recognition and gene regulation.

Many regulatory RNAs work in trans; they affect any non-self RNA that is targeted. Riboswitches are regulatory RNAs that most often work in cis, impacting only the transcript that they are a part of [11]. They are predominantly found in the 5′ untranslated region (UTR) of mRNA transcripts from bacteria. Riboswitches specifically bind to a small molecule ligand, which results in a helical rearrangement that is used to regulate genes. There are over 55 distinct riboswitch classes that have been reported that respond to a diverse array of ligands and regulate an equally diverse set of genes [12]. Riboswitch-driven gene regulation relies on the formation of one of two mutually exclusive structures that will promote or inhibit gene expression. Increasing concentrations of the small molecule ligand will stabilize one of the alternate RNA conformations, resulting in modulation of the downstream gene. In this way, the RNA transcript is capable of self-regulation upon specific interaction with a small molecule signal.

Frequently, a single riboswitch is used to regulate a single gene. However, there are some instances where two riboswitches in sequence are used to regulate the same gene. These are called tandem riboswitches and function as Boolean logic gates [13–15]. Tandem riboswitches can be regulated by the same ligand or by two different ligands, and the relationship between the aptamers and expression platforms modulates the expression of the downstream gene.

### 1.2.1 Transcription Regulation by Riboswitches

The majority of riboswitches regulate the transcription or translation of genes, although some have been found that effect transcript stability or mRNA splicing. In bacteria, translation regulation is typically managed through sequestration of the Shine-Dalgarno sequence, where the ribosome binds to the transcript to begin translation (figure 1.1A). Transcription regulation is most often achieved through Rho-independent transcription termination (figure 1.1B) [11].

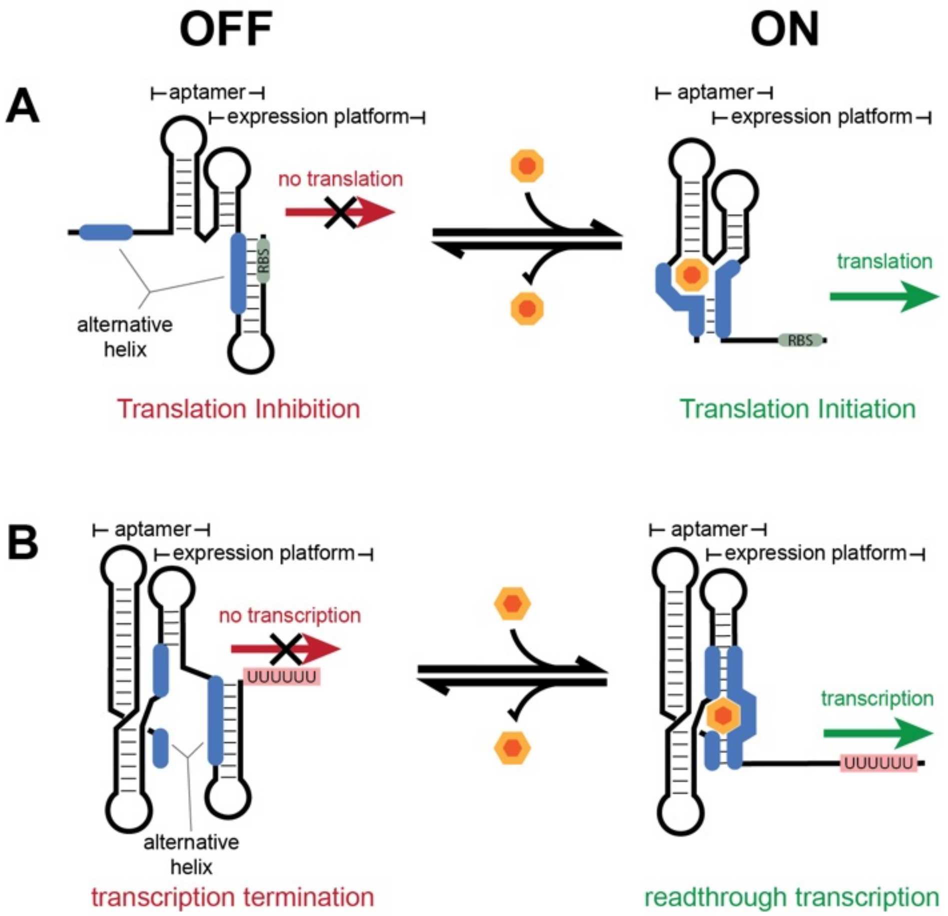

Figure 1.1: Riboswitch factor independent regulation of translation and transcription. (A) An ON switch that regulates translation and restricts access to the ribosome binding site (RBS) in the absence of ligand. (B) An ON switch that, in the absence of ligand, forms a terminator helix followed by a poly-uridine tract that terminates transcription.

Transcription termination requires two key structures: the terminator hairpin and the poly-uridine pause site. Specifically, a poly-uridine tract that contains at least seven uridine nucleotides downstream of a GC-rich helix induces RNA polymerase pausing [16,17]. The adenine-uridine Watson-Crick base pair is less stable than the cytosine-guanine base pair, and as such, RNA polymerase progression is slower. This slowed progression provides sufficient time for the terminator hairpin to fold. Due to the proximity of the hairpin and the RNA polymerase pause site, the base pairing of the hairpin extends into the RNA polymerase exit channel [18]. This, in turn, destabilizes the polymerase interactions with the DNA, causing the polymerase to fall off. As a consequence, transcription terminates after approximately the seventh uridine.

Riboswitch regulation of transcription termination can either stabilize or destabilize the terminator hairpin. For OFF switches, increased ligand concentrations will stabilize the terminator hairpin, resulting in increased rates of transcription termination. For ON switches, increased ligand concentrations stabilize an alternate helix that will occlude terminator formation, resulting in transcription of the full-length gene.

Regardless of regulation mechanism, all riboswitches are composed of two overlapping domains. The aptamer domain binds to the small molecule ligand and the expression platform structurally rearranges in response to this binding event to manipulate gene expression [19]. For transcriptional riboswitches, the expression platform includes the terminator helix as well as the alternate helix. Because of its role in ligand recognition, the aptamer domain is highly conserved across riboswitches of the same class. In contrast, the expression platform is highly dynamic and tailored to each riboswitch and thus has lower conservation. As a consequence, riboswitch studies tend to focus on the aptamer domain.

### 1.2.2 Riboswitch Identification

Riboswitch identification begins with the discovery of RNA aptamers. These highly structured ligand binding domains can be identified using bioinformatic approaches. Specifically, using homology search software to find predicted structured motifs in the 5′ untranslated region (UTR) of bacterial genes [20–23]. Upon discovery of a potential riboswitch, searches are expanded to identify structurally similar sequences that might regulate genes of similar function. Once these sequences are compiled, a consensus model of the predicted riboswitch can be generated [24]. This information is used to identify aptamer candidates.

Following bioinformatic identification, biochemical verification is required. To do this, the corresponding ligand needs to be identified. Frequently, the ligand is closely linked to the gene associated with the riboswitch. For instance, the fluoride riboswitches respond to the fluoride ion and regulate genes that mitigate fluoride toxicity [25]. Verification of the ligand, however, requires more than bioinformatic data; it is necessary to show a direct interaction between the predicted aptamer and the ligand. This is frequently done using binding assays or in-line probing where the aptamer is incubated with the ligand and the interactions are monitored [26]. Neither of these experiments can confirm that the aptamer is a functional riboswitch. To confirm function, the full riboswitch needs to be assessed for modulation of gene expression.

Typical functional experiments generally include reporter assays where a cell is expressing the riboswitch of interest upstream of a reporter gene [27,28]. Reporter assays are valuable because they function independently of riboswitch mechanism. Transcription, translation, and unknown mechanisms are all equally assessed using reporter assays.

## 1.3 One to One: Achieving Specificity

Specific recognition of a single ligand is essential to proper riboswitch function. A riboswitch that cannot discriminate between chemically and structurally similar ligands would fail to regulate gene expression properly. Understanding how a riboswitch can be specific for a ligand is a focus of research in this field.

### 1.3.1 Riboswitch structures highlight key nucleotides

Frequently, riboswitch structures contain all the necessary information to explain riboswitch specificity, although mutational assays are also helpful in determining which nucleotides contribute to specificity. As with other macromolecules, RNAs achieve specificity for a ligand through chemical interactions. Unlike proteins, which are composed of over 20 amino acids with different chemical properties to drive specificity, RNA works with four bases: A, T, G, and C. Additionally, the RNA phosphate-sugar backbone and indirect metal-mediated contacts are frequently involved in ligand recognition.

Conservation diagrams, partnered with mutational assays, aid in identifying sequences essential to riboswitch function, as the sequence elements that show the highest level of conservation are often involved in ligand recognition. However, determining which nucleotides are interacting with the ligand is most often achieved using X-ray crystallography. Crystallography can confirm the importance of specific nucleotides in contact with the ligand or those involved in essential structural motifs. Crystal structures have shown that for most riboswitches, nearly every face of the ligand is specifically recognized by the RNA. This explains some of the selectivity against chemically or structurally similar ligands. If every component is recognized, any modification to the ligand disrupts binding to the riboswitch.

### 1.3.2 Nucleobase recognition: purine riboswitches

The most common ligands for riboswitches are RNA-based molecules. This includes signaling molecules, RNA precursors, RNA derivatives, and enzyme cofactors. All of these molecules are thought to have been components of the early environment in which the first riboswitches evolved. As such, RNA-based molecules are key signals that riboswitches evolved to recognize. Over 35 classes of riboswitches recognize RNA or RNA-derived

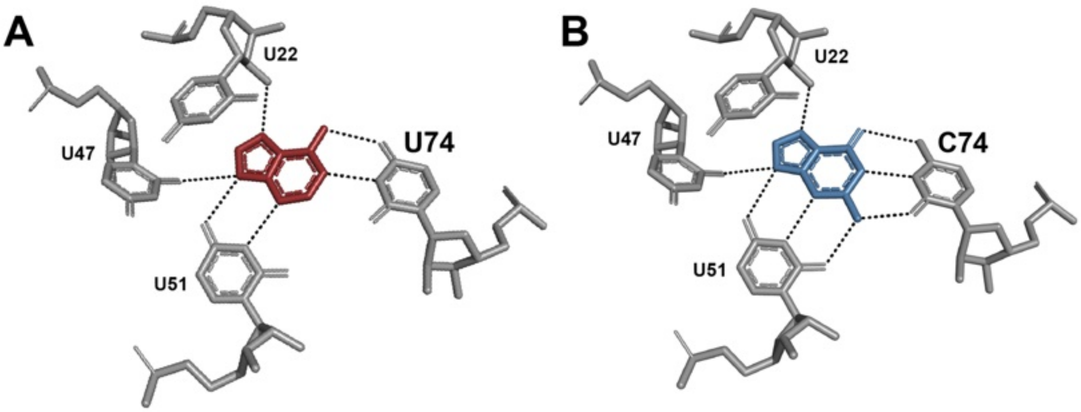

Figure 1.2: Purine riboswitch recognition of all ligand faces. (A) The ligand determining U74 Watson-Crick pairs with adenine. (B) The ligand determining C74 Watson-Crick pairs with guanine. PDB: adenine – 1Y26 and guanine - 6UBU

compounds, including the purine riboswitches [12,29]. The most common riboswitches in this family recognize either adenine or guanine. In this case, the nucleoside is recognized on all faces using RNA-RNA interactions. Specifically, both the minor groove face and the hoogsteen face are recognized through hydrogen bonds. The Watson-Crick face is also recognized using canonical base pairing: cytidine-guanine or uridine-adenine (figure 1.2) [30]. The riboswitch is specific for a single ligand through base pairing despite the structural and chemical similarity between guanine and adenine.

### 1.3.3 Recognizing positive charge: $Mn^{2+}$ and guanidine riboswitches

While it is not particularly common for riboswitches to recognize charged ions, cations lend themselves well to RNA interactions. RNA is composed of a negatively charged phosphate sugar backbone. This backbone frequently coordinates with positive metal ions, specifically magnesium ($Mg^{2+}$) and sodium ($Na^{+}$), to overcome the negative charge repulsion in RNA structures. Cations can also be recognized using ionic interactions with the negative phosphate backbone, as with the manganese ($Mn^{2+}$) riboswitch (figure 1.3A). In this case, the $Mn^{2+}$ ion is selected for over $Mg^{2+}$ despite an increased abundance of $Mg^{2+}$ in the cell. Crystal structures show that the ligand binding site is coordinated using a nitrogen, a soft ligand compared to oxygen, a hard ligand [31]. $Mg^{2+}$ has a strong preference for coordinating with hard ligands, and thus, specificity for $Mn^{2+}$ is achieved through the soft ligand interaction.

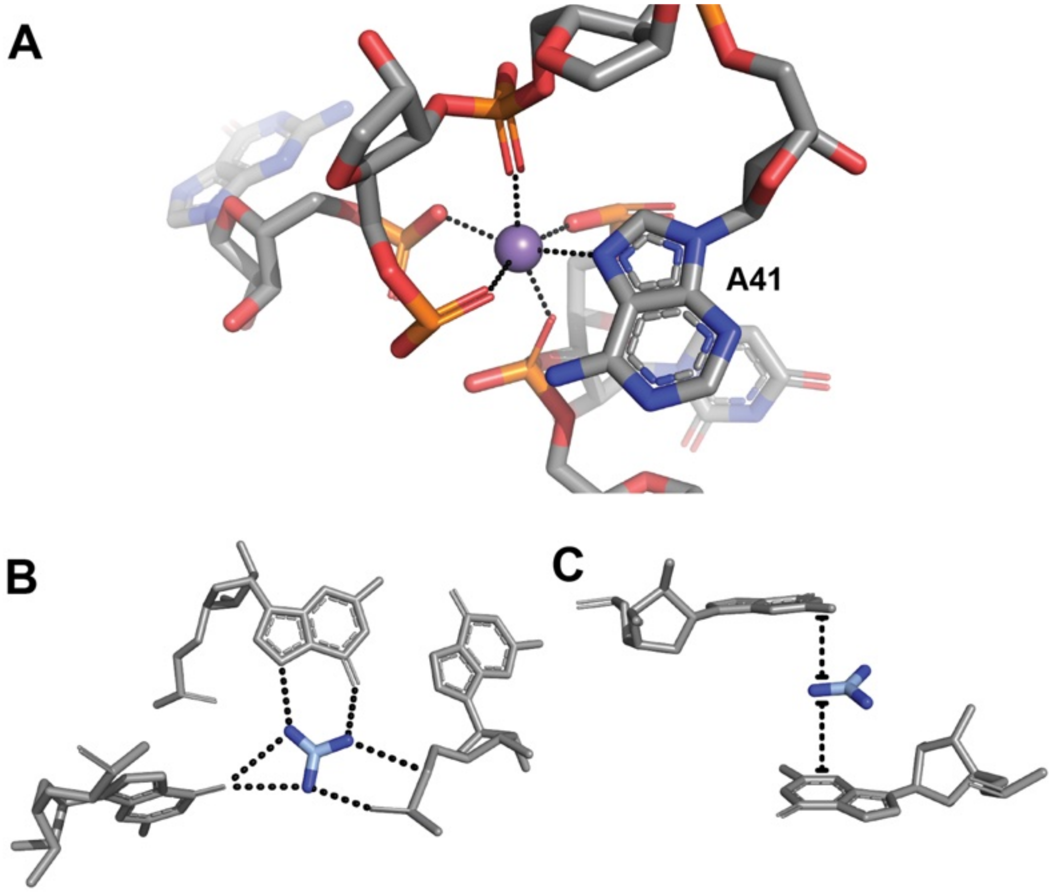

Figure 1.3: Cation recognition by riboswitches. (A) $Mn^{2+}$ is coordinated to five phosphate oxygens and a single adenosine nitrogen. A41 N7 coordination selects for $Mn^{2+}$ over $Mg^{2+}$. (B) All possible hydrogen bonds are made between guanidinium and the riboswitch. (C) cation-$\pi$ interactions from both above and below stabilize the guanidinium. PDB: Mn - 4Y1I and guanidine-I - 5T83

Additionally, the nucleotide bases of RNA can interact with cations. Specifically, the electron-rich region above and below the aromatic rings of the four bases is slightly negatively charged. This, in turn, can interact with a positively charged cation through a cation-π interaction. There are four known guanidine riboswitches that recognize the small cation guanidine [32–36]. All three of the classes that have been crystallized recognize guanidine using cation-π interactions rather than ionic interactions with the phosphate backbone [37–40]. To ensure specificity, all possible faces of the ligand are fully recognized via hydrogen bond interactions with the riboswitch (figure 1.3B & C).

### 1.3.4 Recognizing negative charge: fluoride riboswitches

As previously stated, RNA contains a negatively charged backbone. As such, specificity for anions is achieved through indirect contacts, most often coordination with metal ions. Frequently, anion recognition is only one of a number of RNA-ligand interactions, and thus, specificity can be determined in conjunction with hydrogen bonding patterns and cation-π interactions. However, the fluoride riboswitch recognizes the elemental anion and uses metal coordination alone to specifically respond to fluoride [25,41,42]. The fluoride riboswitch coordinates three $Mg^{2+}$ ions that each contact the $F^-$ ligand (figure 1.4). In this way, the riboswitch is charge-selective. However, $Cl^-$ also contains a negative charge and yet does not activate the riboswitch. Molecular dynamic simulations have concluded that $F^-$, which is highly water-soluble and

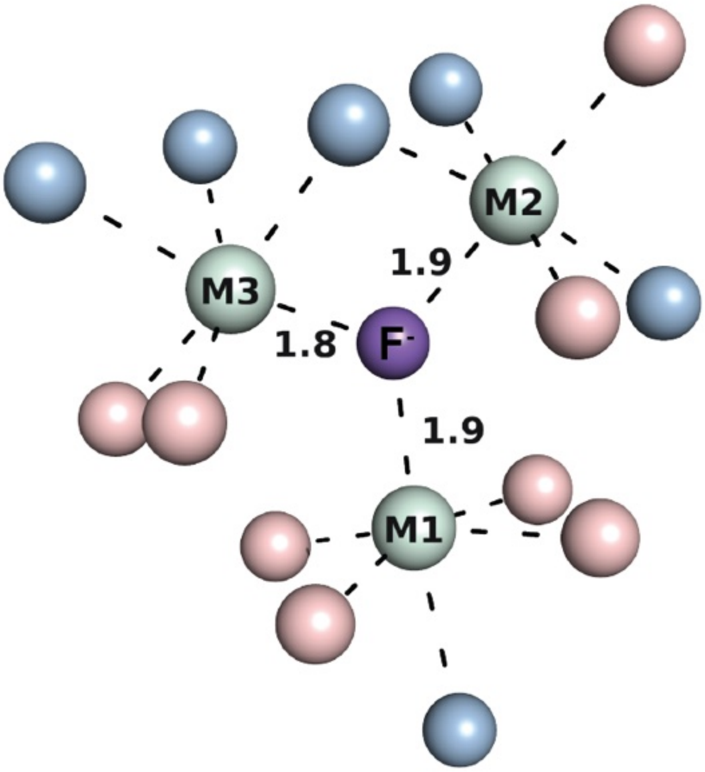

Figure 1.4: Fluoride is coordinated by three Mg2+ ions. The ions are size and charge-selective. The metal coordination geometry is completed by water (blue) and riboswitch oxygens (red). PDB: fluoride – 4ENC

thus more likely to be found in solution than in the core of a riboswitch, results in a shorter bond length than $Cl^-$ [42]. This produces a more stable metal-ligand cluster at the core of the riboswitch. As a consequence, the fluoride riboswitch uses a combination of charge selectivity and size selectivity to respond specifically to fluoride over other halides.

### 1.3.5 The importance of recognizing every face: TPP riboswitches

The most abundant riboswitch is the thiamine pyrophosphate (TPP) riboswitch [12,29]. This is also the most widespread riboswitch known to exist, with examples in plants and fungi as well as bacteria [43]. TPP molecules have three components: a 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) ring, a thiazole group, and a pyrophosphate group. HMP. The HMP and pyrophosphate groups are structurally and chemically related to RNA molecules and are recognized on every face using traditional RNA-RNA interaction motifs [44]. Specifically, HMP is related to cytosine and is recognized through hydrogen bonds with its Watson-Crick face, and the phosphate group is recognized via coordinated $Mg^{2+}$ ions (figure 1.5A). Unlike other riboswitches, which interact with every face of the ligand, the TPP riboswitch does not interact with the thiazole group of the ligand [44]. As a consequence, substitutions to the thiazole group are not discriminated against. This has been tested with pyrithiamine pyrophosphate (PTPP), which differs from TPP only at the thiazole group (figure 1.5B). In contrast to thiazole, which has a five-membered aromatic ring with a sulfur, PTPP consists of a six

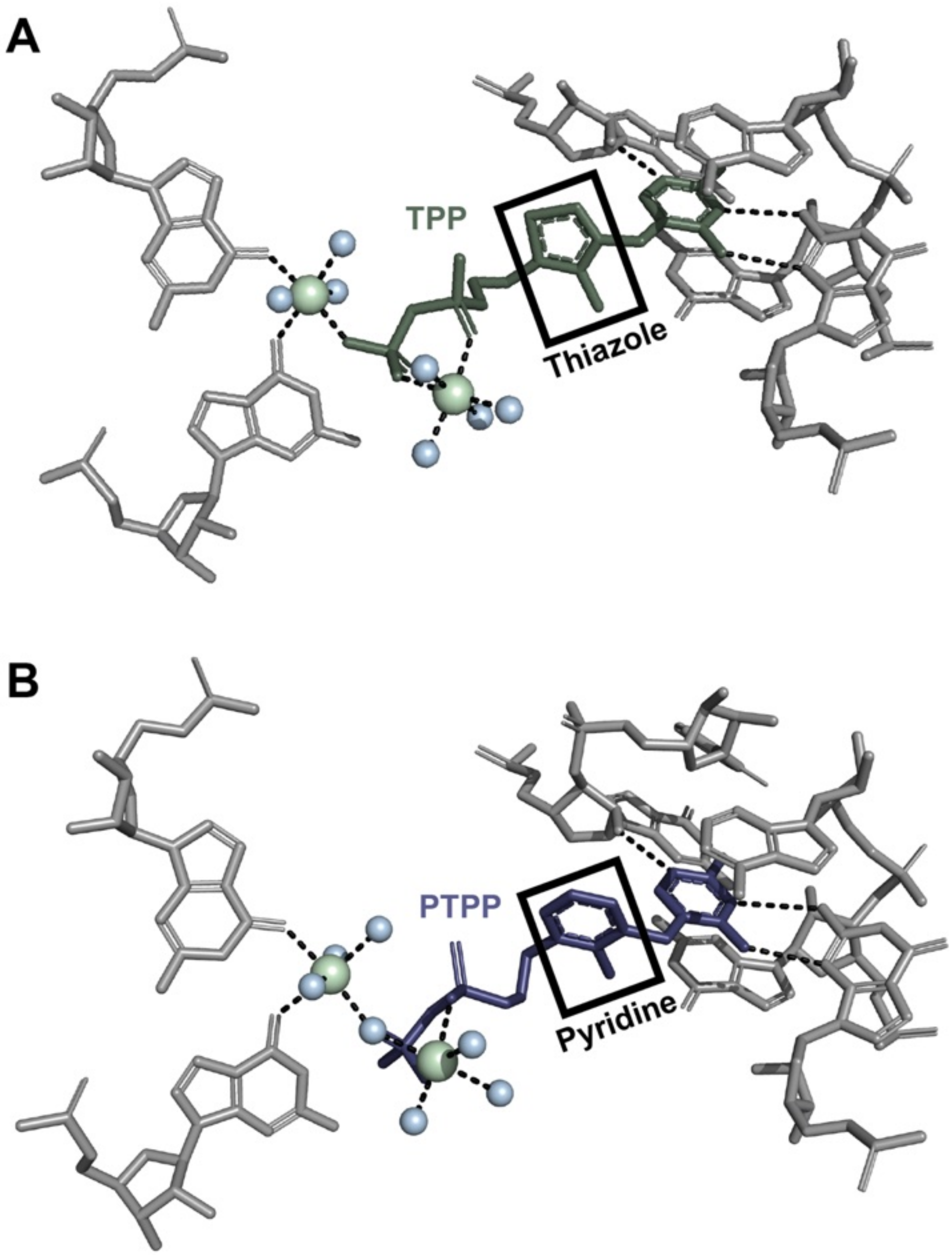

Figure 1.5: TPP riboswitches do not recognize the central thiazole group specifically. (A) A TPP riboswitch with TPP ligand. The pyrophosphate group is coordinated by $Mg^{2+}$ ions (green), and the HMP group is fully encased by nucleotides, including a Watson-Crick base pair and $\pi$ stacking. (B) A TPP riboswitch with PTPP ligand. The pyrophosphate and HMP groups are recognized the same, and the pyridine group is not contacted by the riboswitch. PDB: TPP - 2GDI and PTPP - 3D2V

membered aromatic ring with no sulfur. Despite the chemical differences between TPP and PTPP, the TPP riboswitch binds both ligands with similar affinity [45].

## 1.4 Variant Riboswitches

Riboswitches use a variety of RNA structures and folds to specifically recognize a single ligand. Riboswitches that share the same structure and ligand are in the same class. There are instances where multiple structures have evolved to recognize the same ligand, as is the case with the guanidine riboswitches. In other cases, a single RNA structure has evolved to recognize different ligands, like the purine riboswitches. These classes are called variant riboswitches. Variants have similar conservation diagrams but diverge at key sequence positions. These small nucleotide changes allow for specific recognition of different ligands.

Due to the high sequence and structure similarity between variants, it is often difficult to identify variant subclasses. In most cases, variants are grouped together in a single class during riboswitch discovery and only separated after a ligand has been identified [46]. Often, differentiation between variant classes only occurs after high-resolution structures have been published.

The search for variant riboswitches relies on three factors:

1) sequence conservation
2) aptamer structures
3) gene association

Together, these factors have aided in the discovery of over a dozen variant classes, including the purine riboswitches, the cyclic dinucleotide riboswitches, and the $Na^+/Li^+$ riboswitches [27,28,47–51].

### 1.4.1 Purine Riboswitches: Using Sequence to Identify Variants

The purine switches are an example where small changes in riboswitch primary sequence have the ability to dramatically alter ligand affinity and identity. Generally, nucleotides in direct contact with a ligand are often the most highly conserved nucleotides in a riboswitch because of their role in ligand specificity. Mutations in these core nucleotides can indicate the existence of variant classes. Primary sequence analysis has driven the discovery of several variants in the guanine class of riboswitches, one of the first variant classes uncovered.

Guanine riboswitches were first identified in 2003 in *Bacillus subtilis* upstream of purine salvage genes [48]. Structurally, these riboswitches form a 3-stem junction with highly conserved stem lengths in P2 and P3, which form a pseudoknot [30,52]. Following the identification of guanine as a ligand and the further characterization of guanine riboswitches, examples were discovered with sequence changes at highly conserved nucleotide positions. In the absence of a high-resolution 3° structure of the guanine riboswitch, the significance of these mutations was unknown. However, binding assays using these variant sequences confirmed that, instead of binding guanine, they bound adenine [47]. All of the variant sequences with altered ligand specificity contained a single mutation in the conserved core of the riboswitch, specifically, cytosine to uridine. High-resolution structures of the purine riboswitches found that the critical nucleotide identified in these variant switches Watson-Crick pairs with the purine ligand. Structurally, it follows that the single C to U mutation in the guanine riboswitch sequence fully switches the ligand specificity from guanine to adenine. In addition to the single nucleotide substitution, the adenine riboswitch is distinct from the guanine riboswitch in gene context. It is located upstream of purine efflux pumps and adenine deaminase enzymes, providing further evidence that adenine is the native ligand.

More recently, other guanine variants have been bioinformatically identified and biochemically classified, including the 2′-deoxyguanosine (2′-dG) variant and the xanthine variant [53,54]. The 2′-dG variant was singled out based on core sequence and structural distinctions from the two known purine riboswitches. Specifically, there were minor changes in the lengths of P2 and P3 as well as a U to C mutation in the core of the riboswitch, which distinguish the 2′-dG variant from the guanine parent construct and allow for recognition of the larger ligand [55]. Due to the sequence diversity and plasticity of the purine riboswitch scaffold, further variants are postulated to exist, although they have yet to be identified and characterized.

### 1.4.2 Cyclic Dinucleotides: Structure Informs Variant Discovery

Variant riboswitches often use conserved binding pockets to recognize different ligands. Low conservation of the nucleotides in the ligand binding pocket frequently indicates the existence of variant riboswitches. In some cases, these changes can be inferred from primary sequence conservation alone; however, this is not always the case. High-resolution crystal structures of a riboswitch with the ligand-bound provide essential information on the nucleotides involved in ligand recognition. Using crystal structures in combination with sequence conservation to identify variants is a common practice. The cyclic dinucleotide riboswitches are an example where structure aided in variant identification.

The cyclic dinucleotide riboswitches were first identified in 2007 as an orphan riboswitch family with no known ligand [20]. The primary ligand, cyclic-di-GMP, was identified shortly after the initial motif discovery, and the riboswitch structure was determined immediately after that [51,56]. Crystal structures showed that, similar to the purine riboswitches, the cyclic-di-GMP riboswitch recognizes its ligand using canonical nucleotide interactions – a Watson-Crick base pair at position 92 and a hoogsteen face base pair at position 20. Based on their role in ligand recognition, these two nucleotides were expected to show high conservation. Instead, there were a number of sequences identified that had mutations in one or both of these positions.

These potential variants were investigated by the Hammond and Breaker labs and found to bind cyclic-AMP-GMP instead of cyclic-di-GMP [49,50]. Unlike the purine riboswitches, the nucleotide change in the cyclic dinucleotide variant was not the Watson-Crick nucleotide. Instead, the Hoogsteen face base pair was mutated with a guanine to adenine change at position 20. There are sequences that contain a mutation to position 92, the Watson-Crick nucleotide, but none of these sequences have been shown to be functional yet [49,57].

While binding assays confirmed that the ligand for this variant class was cyclic-AMP-GMP, gene association was used to bioinformatically verify the result. The cyclic-di-GMP riboswitches exclusively reside upstream of genes implicated in the transition to and from sessile to motile life and biofilm formation [20,51]. In contrast, the cyclic-AMP-GMP riboswitches are generally found upstream of genes related to exoelectrogenesis and pili formation, although some are upstream of genes linked to cyclic-di-GMP riboswitches [49,50]. This unique gene context is an additional indicator of variant riboswitches.

### 1.4.3 *ykkC*: Genes Differentiate Variant Classes

Normally, riboswitches within a class regulate downstream genes with similar functions. Generally, the function of downstream genes is related to the ligand that the riboswitch binds, as is the case with the purine riboswitches. In some cases, gene annotations have been used independently of primary sequence to identify variants of a riboswitch class. One such example is the *ykkC* family of riboswitches. The *ykkC* family was first identified in *B. subtilis* in 2004, although the ligand remained unknown for over a decade [58]. This family was grouped together because of similarities in sequence and structure. Initially, all the *ykkC* RNAs were thought to regulate the same class of genes. However, with time, more gene functions were identified, and it became clear that the ykkC family had a particularly diverse genetic context, suggesting that it contained variant classes. Following the identification of guanidinium ions as the native ligand for some *ykkC* riboswitches, the remainder of the family was divided into variant subclasses based on the types of genes that they regulated [32].

The first class identified, the guanidine riboswitches, regulate genes for guanidine carboxylases and guanidine transporters [32]. The remaining subclasses, 2a, 2b, 2c, and 2d, regulate amino acid synthesis and transport, *de novo* purine biosynthesis, nucleotide hydrolases, and an unknown transporter, respectively [14,59–61]. Following the separation of *ykkC* subclasses based on genetic context, analysis of sequence and structure confirmed the delineation between classes 2a, 2b, 2c, and 2d. Once the variants had been grouped, the ligands were identified for most of the variant classes based on genetic context and confirmed with binding and functional assays. Class 2a is responsive to guanosine tetra/pentaphosphate ((p)ppGpp), which is an alarmone that signals starvation adaptation [59,62,63]. This agrees with the genetic context of 2a, which regulates branched chain amino acid synthesis and transport. Class 2b recognizes PRPP, a precursor in purine biosynthesis, which agrees with the gene context of that variant class, *de novo* purine biosynthesis [14]. Finally, 2c binds nucleoside diphosphates congruent with its position upstream of nucleotide hydrolases [60]. The native ligand for 2d remains unknown.

When comparing the variant sequences to the parent guanidine construct, nucleotide changes within the core structural motif can be identified. These primary sequence changes are localized near the ligand binding site, similar to those found in purine and cyclic dinucleotide riboswitch variants [38,64–66]. Together, the *ykkC* RNAs share a similar structural scaffold but incorporate minor changes in primary sequence that alter the architecture of the binding pocket. This, in conjunction with the diversity of gene associations, confirmed the presence of variants prior to ligand identification and structure determination.

## 1.5 Synthetic Riboswitches

While natural riboswitches have been found that recognize over 35 distinct ligands, there are a host of molecules that are not yet recognized by riboswitches. Moreover, riboswitches are a valuable tool for synthetic biology, and evolving them to recognize new ligands is a growing field. There are four features that make riboswitches such powerful synthetic biology tools: simplicity, evolvability, reversibility, and tunability [67,68]. Riboswitches use simple mechanisms that require only the RNA of interest and are thus easier to insert into cells than more complex systems that might require protein factors. Riboswitches are also highly evolvable, as seen in the variant families. They can theoretically be designed to recognize any small molecule ligand, either native to the cell or introduced by the investigator. Additionally, unlike most current mechanisms of synthetic gene regulation, riboswitches rapidly respond to stimuli in a reversible manner. Finally, riboswitches can be tuned based on the ligand concentration. This can be done by manipulating the cellular ligand concentration itself or by manipulating the expression platform of the riboswitch to change its sensitivity to ligand.

In addition to functioning as gene regulatory elements, riboswitches have also been developed as small molecule sensors. Riboswitches combined with a reporter have served to inform on the biological levels of a given small molecule ligand [69–71]. This reporter is occasionally a protein such as β-galactosidase or a fluorescent protein. In other cases, the riboswitch is placed in tandem with an RNA aptamer that fluoresces upon ligand binding to both aptamers.

### 1.5.1 Evolving Synthetic Riboswitches: SELEX

Nearly a decade before the first natural riboswitch was characterized, tools for synthetic aptamer evolution were being developed. An *in vitro* evolution method called SELEX, Systemic Enrichment of Ligands By EXponential enrichment, was designed to select RNA or DNA aptamers that bound to a ligand of interest [72]. This produces highly selective aptamers with tight ligand affinities. However, a functional riboswitch requires more than a strong aptamer. The aptamer must be appended to an expression platform and capable of ligand-dependent structural rearrangements [73,74]. For this reason, many modifications to SELEX have been developed in the last five years to optimize the selection of riboswitch aptamers. One of the new methods developed applies the key features of variants to novel aptamer development. Specifically, the SELEX-graftamer method evolves a new function from an existent riboswitch rather than starting *de novo* from random sequence [75]. Using this approach, it is possible to keep the same expression platform while developing a new ligand preference. While this method is a strong step in the right direction for synthetic riboswitch development, there is a lot of work that goes into fine-tuning the sensitivity of the evolved switch and the *in vivo* activity that cannot be streamlined with this method yet.

## 1.6 Investigating riboswitches using next-generation sequencing

Advances in sequencing technology have made in-depth analysis of DNA and RNA far more accessible over the last three decades. Intracellularly, this applies to whole genome sequencing as well as capture-based sequencing that can be used to identify nucleic acids that are interacting with a specific target. Outside of the cell, this has made analysis of synthetic DNA libraries possible. This includes processing the evolved sequences from SELEX or other aptamer evolution methods.

Next-generation sequencing (NGS) has been applied to the question of sequence-function relationships as well. Specifically, processing libraries that have undergone a functional assay *in vitro* provides information on which sequences impact the function of the RNA in question. NGS has been used to assess ribozymes - catalytic RNAs – in the Greenleaf and Yokobayashi labs. In both labs, fluorescent molecules were used to monitor ribozyme self-cleavage rates [76–80]. Additionally, previous work in the Strobel lab has investigated riboswitch transcription (Sequencing-based Mutational Analysis of RNA Transcription Termination, SMARTT) and translation (Sequencing-based Mutational Analysis of RNA Translation Initiation, SMARTI) [81–83]. These studies assessed how mutations to the core RNA sequence impacted the function of the riboswitch but did not assess ligand specificity. Unlike SELEX, which removes non-functional sequences, these functional assays retain all library sequences, including those that inhibit function. These studies generate quantitative data that inform on essential RNA structures and sequences that impact function. This information provides an understanding of the requirements for a functional RNA.

In this thesis, I use an NGS-based functional assay to explore the relationship between riboswitch sequence and ligand specificity in variant riboswitches. I am especially focused on nucleotides in the second-shell of ligand contact as determined by crystal structures. Due to the overlap between the aptamer domain and the expression platform, I have also generated data that investigates the effect of nucleotide changes in the expression platform on ligand specificity in variant riboswitches. I am hoping to build an understanding of ligand specificity in closely related systems under kinetic conditions. This will help inform on how variants are related and distinct within a class. It will also expand the criteria available for variant riboswitch discovery.

# 2. High-throughput analysis of the *ykkC* variant family

I have adapted this chapter from a manuscript currently in progress with Dave Hiller and Scott Strobel. Dave Hiller was instrumental in discussing data analysis and understanding the results.

## 2.1 Background

Riboswitches are composed of two overlapping domains. However, despite the functional consequences of the expression platform, the bulk of riboswitch studies have been performed on aptamers in isolation [14,38,59,64–66]. Based on conformational heterogeneity and the dynamic nature of the full-length riboswitch, it is frequently easier to study the aptamer alone. However, these studies present a static view of the aptamer and do not necessarily inform on the function of the full-length construct. Moreover, aptamer studies are frequently performed on systems at thermodynamic equilibrium, and recent studies suggest that both transcriptional and translational riboswitches are kinetically regulated [84]. Riboswitch publications from the last five years have started to address the gap between aptamer studies and functional riboswitches using high-throughput or single-molecule methods [81–83,85,86]. These studies have highlighted the role of the expression platform in ligand recognition in addition to its canonical role in signal transmission.

Variant riboswitches are RNAs with high sequence and structure similarity that recognize different ligands [87]. A wide variety of variant classes have been identified that recognize a commensurately diverse array of ligands and regulate a diverse group of genes [46,88]. The majority of variant families recognize structurally similar ligands. For example, the purine riboswitches recognize either adenine, guanine, 2’-deoxyguanine, or xanthine [47,48,53,54]. In contrast, the *ykkC* family is capable of recognizing structurally and biochemically diverse ligands [14,32,58–60].

The *ykkC* family of riboswitches currently encompasses five known variants. The first variant identified, labeled class I, recognizes guanidine and regulates genes for guanidine metabolism and transport [32]. Class 2a binds the alarmone ppGpp and regulates branched chain amino acid metabolism and glutamine biosynthesis (figure 2.1) [59]. Class 2b binds the purine precursor PRPP and correspondingly regulates purine metabolism (figure 2.1) [14]. Class 2c regulates nucleotide diphosphate-associated genes, NUDIX, and binds (d)CDP/ADP [60]. Last, class 2d has no known ligand but regulates genes for an unknown transporter related to the phosphonate utilization system, suggesting the ligand may be a small toxic metabolite like guanidine [60,61].

Structural studies shed light on how similar scaffolds can recognize such different ligands. The guanidine aptamer uses a two helix scaffold and a single added helix to form a short ligand binding pocket [38,65]. In contrast, crystal structures revealed that both the ppGpp and PRPP aptamers contain an additional P0 helix, shifting the binding pocket down the helical stem [64,66]. As ppGpp and PRPP are much larger than guanidine, this structural extension

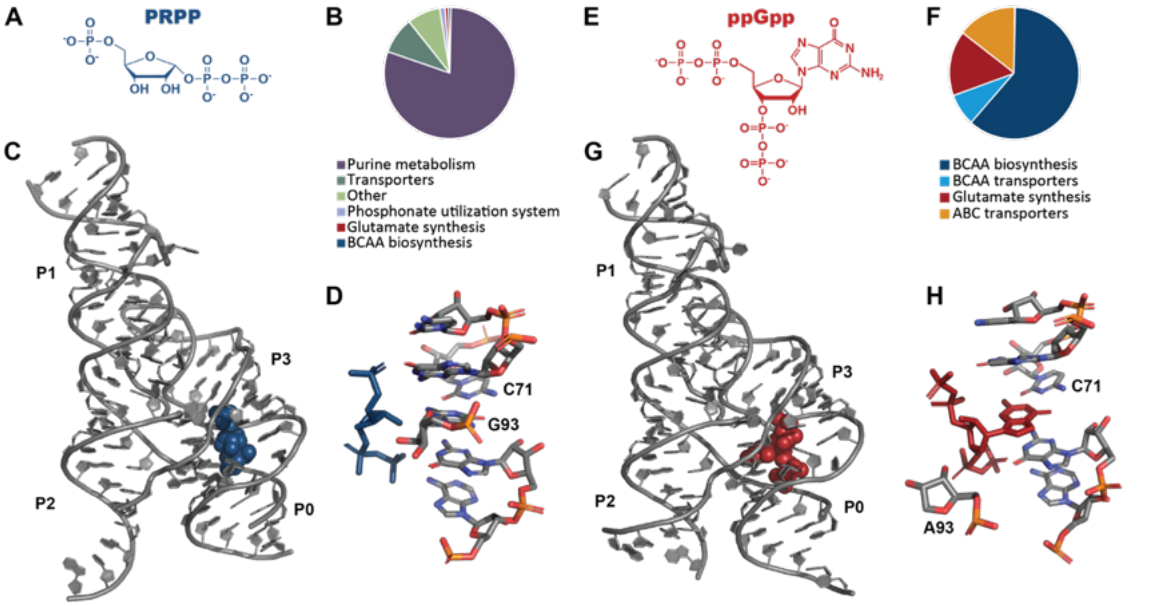

Figure 2.1: Structural features and genetic associations of PRPP and ppGpp riboswitches. (A) PRPP is a polyanionic precursor in purine biosynthesis pathways. (B) the PRPP riboswitch predominantly regulates genes related to purine metabolism and transport. (C) PRPP is recognized using nucleotides in P3. P0 extends below the ligand binding pocket, and P1 and P2 serve as a scaffold to support ligand binding. (D) In PRPP riboswitches, G93 base pairs with C71. (E) The alarmone ppGpp is a polyanionic ligand with a guanine base. (F) ppGpp riboswitches regulate genes for branched-chain amino acid biosynthesis and transport. (G) the structural scaffold of the ppGpp riboswitch matches that of the PRPP riboswitch with the ligand bound in the same location. (H) In ppGpp riboswitches, C71 base pairs with the guanine base, and A93 is oriented away from the ligand. PDB: 6CK4 and 6CK5 for PRPP and ppGpp, respectively.

of the ligand binding pocket allows for the specific recognition of larger polyanionic ligands. A mutational study on a PRPP aptamer showed that removing the P0 helix in combination with a few additional mutations to the aptamer is sufficient to change the binding specificity to guanidine [65].

The ppGpp and PRPP aptamers are much more similar to each other than to the guanidine aptamer. Specifically, the identity of a single nucleotide at position 93 has been used to differentiate between ppGpp and PRPP aptamers [14,59,64,66]. Crystal structures show that position 93 is in the ligand binding pocket, and biochemical studies show it is largely responsible for ligand identity and specificity [64,66] (figure 2.1). Changing the identity of the nucleotide at position 93 from G to A or U can change the binding specificity of a PRPP aptamer to ppGpp [64,66]. These studies have shown that mutations to the *ykkC* scaffold can change ligand specificity between more chemically diverse ligands, such as PRPP and guanidine, as well as the chemically related ppGpp and PRPP.

All mutational studies on the *ykkC* riboswitch have used the aptamer domain in isolation. The role of the expression platform in ligand specificity remains unexplored. Moreover, the sequence context requirements for riboswitch specificity remain uncertain. For instance, is it essential to have a specific surrounding sequence, in which case evolving different ligand specificities from a single aptamer requires a specific starting point, or can any PRPP aptamer switch to a ppGpp aptamer upon the introduction of a single mutation? And, is this relationship bi-directional - can a ppGpp riboswitch be mutated to respond to PRPP?

Here, we use a massively parallel functional assay to measure thousands of mutants of a ppGpp riboswitch. SMARTT (sequencing-based mutational analysis of RNA transcription termination) is a high-throughput mutational assay that generates quantitative ligand-based transcription termination data for thousands of mutants simultaneously [82]. We demonstrate that in this context, the previously reported A93G mutation, differentiating ppGpp from PRPP aptamers, is neither necessary nor sufficient to change the specificity of this riboswitch. Moreover, bioinformatic analysis finds that the ppGpp and PRPP *ykkC* variants have different expression platform requirements.

## 2.2 Results

### 2.2.1 A single mutation is insufficient to switch the functional specificity of a PRPP or ppGpp riboswitch

Reiss et al. and Peselis et al. have previously reported that a single mutation to the ligand binding pocket of a PRPP aptamer changed the binding specificity from the native PRPP ligand to ppGpp [64,66]. However, it is unclear if this result is representative of the *ykkC* class as a whole. To investigate the role of this binding site mutation in gene regulation, we assessed whether a single mutation to a PRPP or a ppGpp riboswitch could change ligand specificity in a functional transcription termination assay. This assay leverages the length difference between terminated transcripts and full-length

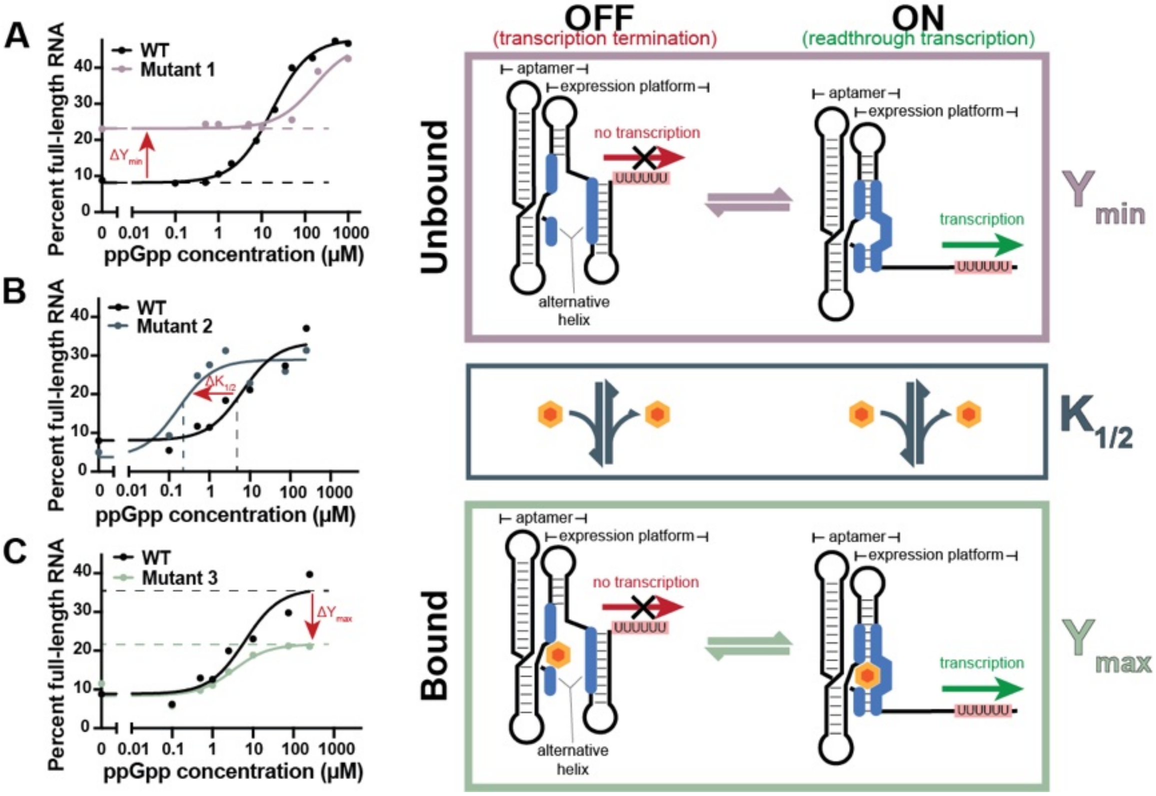

Figure 2.2: ykkC ON switch transcription energetics. (A) $Y_{min}$ describes the relationship between the termination and readthrough conformations in the absence of ligand. (B)The $K_{1/2}$ describes the relationship between the ligand present and ligand absent states of the riboswitch. (C) $Y_{max}$ describes the relationship between the termination and readthrough conformations in the presence of ligand.

transcription to assess the impact of ligand-driven conformational changes to the RNA. Three key fit parameters are identified from ligand-response curves: $Y_{min}$, $Y_{max}$, and $K_{1/2}$. The $Y_{min}$ describes the relationship between the termination state and the readthrough states of the riboswitch in the absence of ligand (figure 2.2 A, purple arrow). The $Y_{max}$ describes the relationship between the termination state and the readthrough state extrapolated to infinite ligand concentration (figure 2.2 C, green arrow). The relationships between these states can be manipulated by changing the stability of either the readthrough or the terminator conformation. The $K_{1/2}$ describes the relationship between the bound and unbound conformations (figure 2.2 B, gray arrow). Specifically, the $K_{1/2}$ measures the riboswitch sensitivity to ligand. Increased sensitivity to ligand, a lower $K_{1/2}$, indicates a shift in the equilibrium toward the ligand-bound states. With these three key parameters, the full riboswitch response to ligand can be described and analyzed.

Neither of the published PRPP specificity-swap constructs has a clear termination site, and thus, they are not amenable to transcription termination assays. Instead, a well-behaved PRPP riboswitch from *Parvimonus micra* (*P. micra*) was identified for transcription termination that responds to PRPP with a half-maximal termination ($K_{1/2}$) at $503 \pm 5$ μM, a $Y_{min}$ of 25%, and a $Y_{max}$ of 91% (figure 2.3 B). The *P. micra* riboswitch regulates a Xanthene/uracil/vitamin C permease gene, which is involved in nucleotide transport and metabolism. This gene association agrees with its annotation as a PRPP riboswitch. To see if a

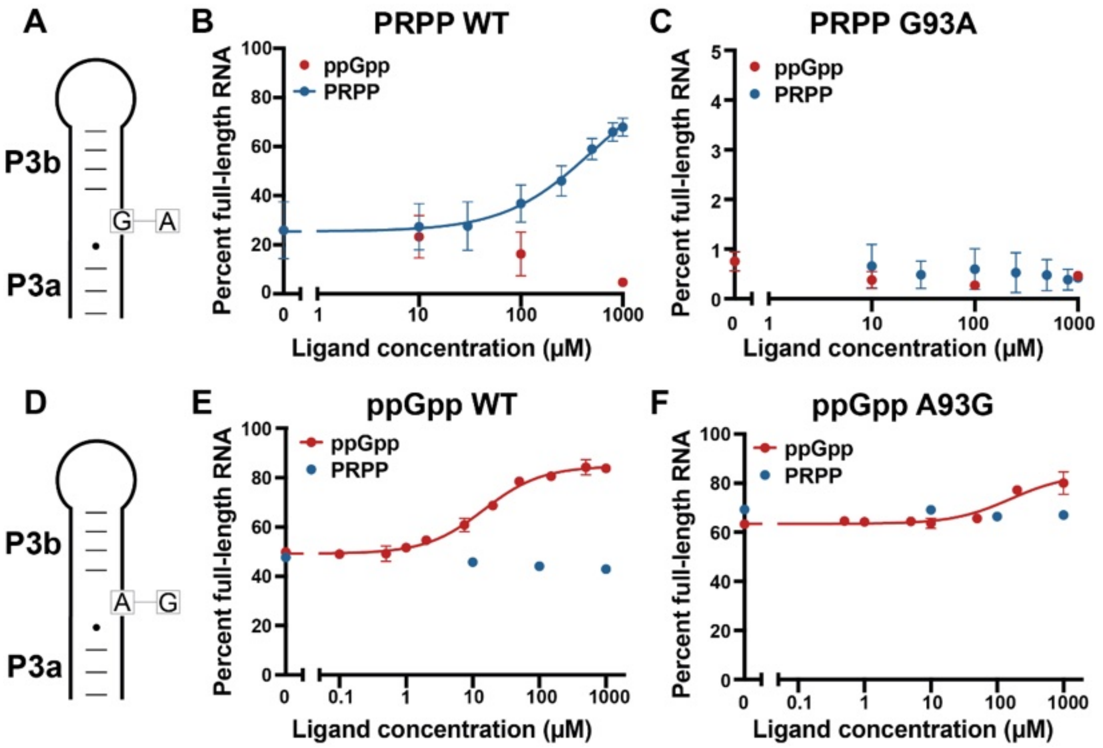

Figure 2.3: Transcription termination with *ykkC* constructs. (A) The *P. micra* riboswitch has a G in the ligand binding position. (B) Gel-based transcription termination of the WT *P. micra* riboswitch with ppGpp and PRPP. (C) Gel-based transcription termination of the binding site mutation, G93A, riboswitch with ppGpp and PRPP. (A) The *T. oceani* riboswitch has an A in the ligand binding position. (B) Gel-based transcription termination of the WT *T. oceani* riboswitch with ppGpp and PRPP. (C) Gel-based transcription termination of the binding site mutation, A93G, riboswitch with ppGpp and PRPP.

single mutation was sufficient to change the ligand specificity in a functional assay, a construct with the binding site G93A mutation was generated. The single G to A mutation in the binding pocket was not sufficient to generate a response to any concentration of ppGpp tested (figure 2.3C). Moreover, this binding site mutation resulted in a significant decrease in the $Y_{min}$. This decrease suggests that the off state has been over-stabilized, and neither PRPP nor ppGpp supplies sufficient energy to the ON state to prevent terminator formation.

To test the reverse direction, a native ppGpp switch from *Thermosediminibacter oceani* (*T. oceani*) that regulates ilvE, a branched chain amino acid aminotransferase, was selected. This ppGpp riboswitch showed a robust ligand response in transcription termination with a $K_{1/2}$ = 15±3 μM, a $Y_{min}$ of 49±2%, and a $Y_{max}$ of 85±2% (figure 2.3 E). I hypothesized that a single A to G mutation in the ligand binding pocket could switch the ligand specificity from ppGpp to PRPP. Using transcription termination, the A93G mutation decreased the ppGpp sensitivity over 10-fold ($K_{1/2}$ = 180±40 μM) but had no observable effect on PRPP-driven function (figure 2.3 F), resulting in a riboswitch that is still specific for ppGpp.

### 2.2.2 A high-throughput assay identifies sequences of a ppGpp riboswitch with functional activity

The initial tests on the *P. micra* and *T. oceani* riboswitches showed that a single mutation is insufficient to switch the ligand specificity in a functional

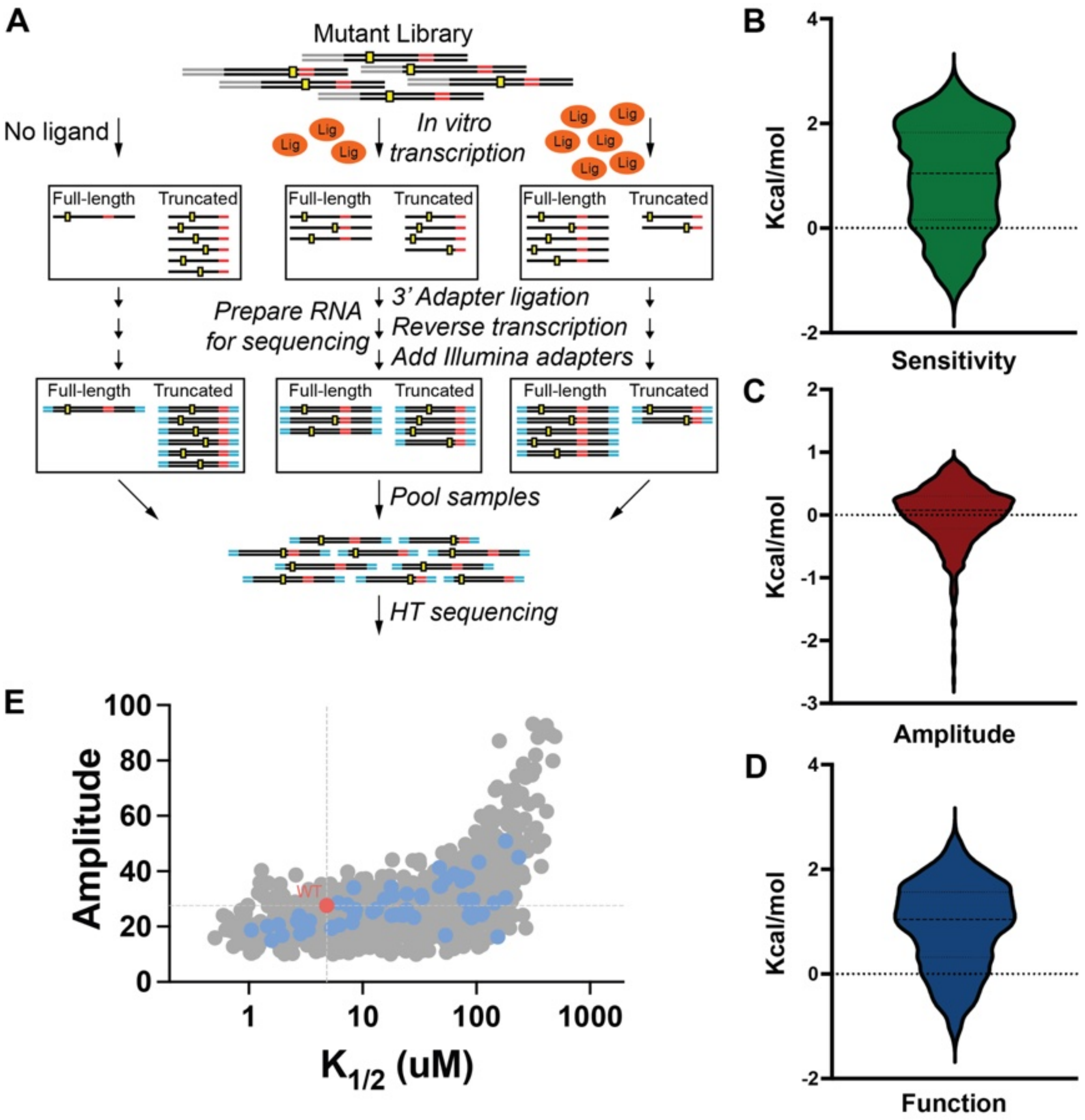

Figure 2.4: SMARTT accesses a wide range of possible $K_{1/2}$ and amplitude values. The SMARTT workflow begins with a mutant library followed by *in vitro* transcription and preparation for sequencing (A). The sequences are parsed and separated computationally and sensitivity ($K_{1/2}$) (B), amplitude (C), and function (D) are determined for each sequence. A wide range of functional riboswitches are identified with SMARTT (E). Sequences with too much error or $K_{1/2}$ outside the measurable window are shown in light gray, double mutants are shown in dark gray, single mutants are light blue, and WT is red.

Assay, so I set out to investigate the positions responsible for ligand recognition and riboswitch function. The decreased $Y_{min}$ observed in the *P. micra* mutant construct was not amenable to a high-throughput assay, so I proceeded with the *T. oceani* ppGpp riboswitch. I generated a mutant library and used a massively parallel approach, SMARTT, to assess the contribution of every position in the ppGpp riboswitch, including those structurally distal to the ligand binding pocket, to function. The mutational region was focused around the ligand binding pocket identified in crystal structures and the terminator helix as these regions show the most variability when comparing PRPP and ppGpp riboswitch consensus sequences. Full transcription termination ligand-response curves were generated for over 24,000 sequences encompassing all single mutants, most double mutants, and most triple mutants that contain the previously reported binding-site (A93G) mutation.

The SMARTT-generated transcription termination curves for the wild-type (WT) sequence are comparable to curves generated using the traditional gel-based transcription termination method, with a $K_{1/2} = 7 \pm 5 \mu M$ and an amplitude of 25% ($Y_{min} = 7.3\%$, figure 2.4). Additionally, the A93G mutant matched the gel-based transcription termination results with a $K_{1/2} = 26 \pm 10 \mu M$ and an amplitude of 10%. Together, these results provide confidence in the remainder of the generated data.

We extracted functional sequences that respond to ppGpp based on an amplitude > 10% and a $K_{1/2}$ within the ligand concentrations tested. Then, to

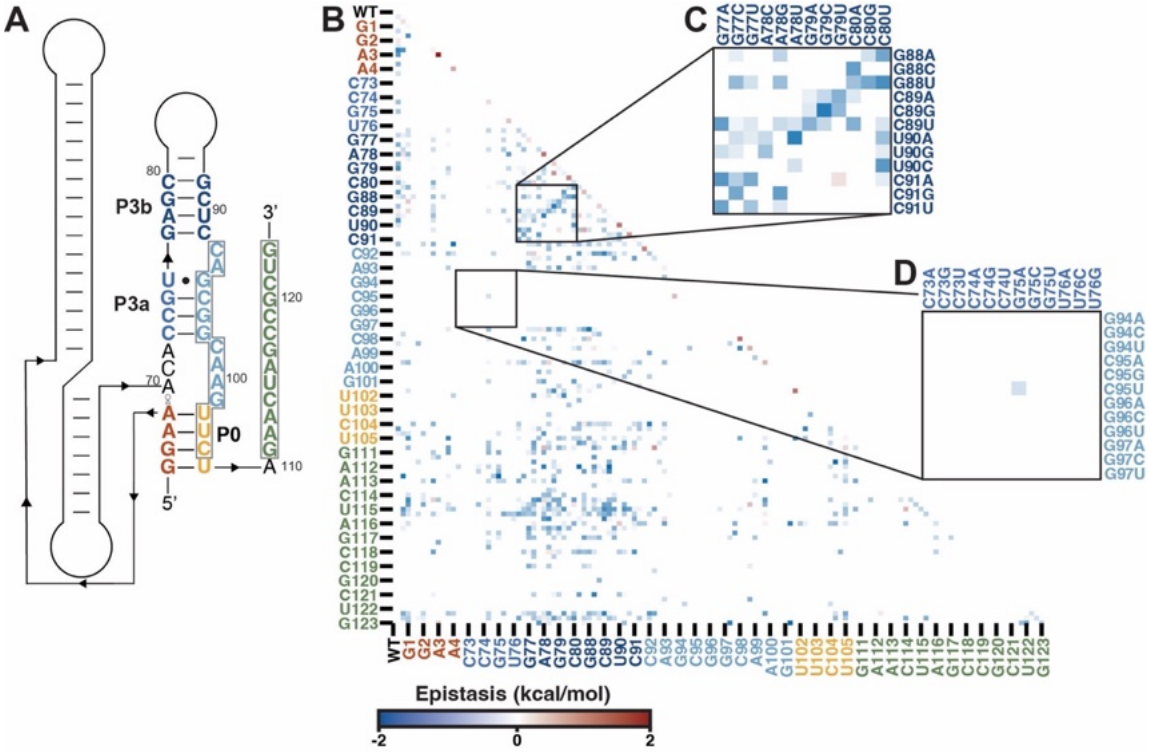

Figure 2.5: SMARTT epistasis and P3b covariation. (A) The SMARTT mutant library focused on ligand contacting positions in P0, P3, and the terminator. (B) A heat map of the epistasis value was generated for all single and double mutants. (C) P3b shows a diagonal perpendicular to the hypotenuse indicative of covariation. (D) P3a contains very few functional mutations although there is a single point of covariation.

assess sequences in an unbiased manner, we generated a functional parameter that combines sensitivity to ligand ($K_{1/2}$) with dynamic range (amplitude) (figure 2.4).

Among the functional sequences, we observe $K_{1/2}$ and amplitude corresponding to the entire measurable range for both values (figure 2.4). This riboswitch appears to be near the maximal functional value that can be achieved within two mutations, although the functional range is significant. The diversity of different functional values showcases the tunability of the riboswitch. Specifically, nearly any sensitivity can be achieved within two mutations. Similar observations about tunability have been made using high-throughput methods on other riboswitch systems [81,82]. As such, tunability may be a universal feature of riboswitches.

The difference between the observed and the expected function was calculated for each double mutant (equation 8) and plotted on a 2-dimensional heat map (figure 2.5B). Sequences where the effects of two mutations are non-additive provide insight into nucleotides that may interact energetically. The majority of mutations resulted in a non-functional riboswitch, indicating that the WT sequence is fairly optimized for riboswitch function. P3b mutations generally show lower function that can be rescued with compensatory mutations that restore base pairing. This is observable as a blue diagonal in the heat map. Base pairing in this region is present in both the ligand-bound and the terminator conformation so these features are particularly distinct (figure 2.5).

### 2.2.3 Riboswitch terminator energetics are tightly controlled

Mutations in P3a are universally detrimental to riboswitch function. P3a is a region that only exists in the ligand-bound state of the full-length riboswitch (figure 2.6A). In the absence of ligand, the 3′-end of P3a forms the base of the terminator hairpin (figure 2.6B). As a consequence, mutations to P3a have impacts on both ligand binding and the energetics of helical switching. These results show that any mutation to P3a abolishes riboswitch function regardless of compensatory mutations that would restore base pairing to the ligand-bound conformation (figure 2.5D). Mutations that rearrange the ligand binding pocket, like those to P3a, are expected to favor the OFF state. However, any mutation to the 3′-end of P3a destabilizes the terminator in addition to the ligand binding pocket. This is sufficient to favor the ON state over the expected OFF state, even in the absence of ligand. In contrast, mutations to the 5′-end of P3a in any context disrupt the ligand binding pocket only and favor the OFF state.

As previously noted, compensatory mutations to restore base pairing in P3a do not rescue riboswitch function (figure 2.5D). Based on the consensus sequence, a covariation diagonal matching the P3b diagonal was expected in P3a. Instead, there is a single point of covariation that changes a GC base pair into an AU base pair (figure 2.5D). Unlike covariation in P3b, which returned full

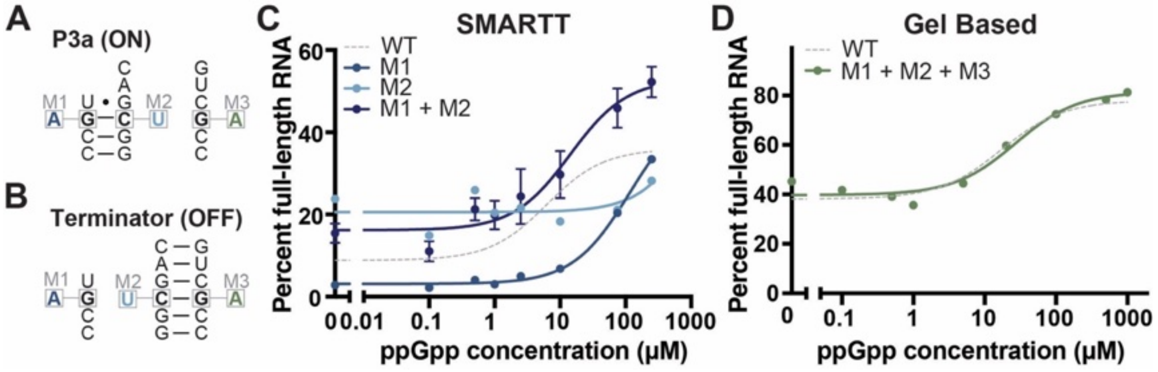

Figure 2.6: Covariation in P3a requires three mutations. (A) In the presence of ligand, the readthrough conformation is stabilized and P3a is formed. (B) In the absence of ligand, the terminator conformation is favored. (C) SMARTT revealed that mutating a GC base pair in P3a to an AU base pair allows for some riboswitch function with a shifted $Y_{min}$. (D) Adding an additional compensatory mutation to the terminator and testing transcription termination by gel returns the $Y_{min}$ to WT levels.

WT function, G75A;C95U returns the sensitivity of the riboswitch to near WT levels, but it shifts the $Y_{min}$ up by 10% (figure 2.6C). This shift in the $Y_{min}$ is due to the state-specific occurrence of P3a, which is present in the ON conformation of the riboswitch but not in the OFF conformation (figure 2.6A & B). Thus, a mutation to both nucleotides in a P3a base pair will result in a single mutation to the terminator base pair (figure 2.6A & B). This decreases the stability of the terminator conformation by introducing a GU wobble in the place of a GC base pair, resulting in the observed shift in $Y_{min}$.

Using the covarying P3a mutation as a base construct, restoring pairing in the terminator stem with a third mutation, G120A, returns the riboswitch to WT function using a gel-based assay (figure 2.6D). This triple mutant, with full base pairing capabilities, has a $K_{1/2}$ = 23±5 μM, $Y_{min}$ = 51±5%, and $Y_{max}$ = 88±5% (figure 2.6D). Other triple mutations in P3a and the terminator were investigated by gel and confirm a third mutation is sufficient to show covariation. The third mutation is necessary to maintain terminator energetics and near WT function of the riboswitch (Table 2.1).

Table 2.1: P3a triple mutants assessed by gel restore function

| **P3a double** | **$K_{1/2}$ (μM)** | **$Y_{min}$ (%)** | **$Y_{max}$ (%)** | **P3a triple** | **$K_{1/2}$ (μM)** | **$Y_{min}$ (%)** | **$Y_{max}$ (%)** |
|---|---|---|---|---|---|---|---|
| WT | 15 | 49 | 85 | WT | 15 | 49 | 85 |
| C73G;G97C | - | 98 | 97 | C73G;G97C;C118G | 74 | 26 | 69 |
| C74G;G96C | - | 97 | 97 | C74G;G96C;C119G | 188 | 73 | 91 |
| G75A;C95U | 6 | 78 | 90 | G75A;C95U;G120A | 23 | 51 | 88 |

### 2.2.4 Native *ykkC* variants have different terminator energies

Based on these results, which indicate the importance of terminator energetics in native ligand recognition, we investigated natural ykkC terminator sequences. The minimum free energy for each terminator stem was calculated, and the length of the poly-uridine pause site was identified for all sequences with predicted rho-independent transcription termination sites. Natural PRPP riboswitches have a stronger terminator hairpin by an average of 5 kcal/mol (figure 2.7A). Both PRPP singlets and tandems are included as they are not sufficiently distinct to merit individual evaluation (figure 2.7B). In keeping with this trend, the *T. oceani* ppGpp riboswitch assessed in this work has a terminator energy of -14.9 kcal/mol, while the *P. micra* PRPP singlet riboswitch assessed in this work and the previously published *T. mathranii* PRPP tandem both have significantly more stable terminators with energies of -31 kcal/mol and -24.7 kcal/mol respectively. An increased terminator hairpin stability is paired with a more uridine-rich pause site in the PRPP riboswitch (figure 2.7D). Additionally, the first three nucleotides of the poly-uridine pause site have been found to be the most essential in efficient transcription termination [89,90]. Analysis of the first three nucleotides in PRPP and ppGpp riboswitches found that PRPP switches have a higher frequency of the preferred UUU sequence (figure 2.7E). Again, the *T. oceani* ppGpp riboswitch assessed in this work matches this trend with the poly-uridine pause site containing 29% non-uridine nucleotides, the first three of which are UGU.

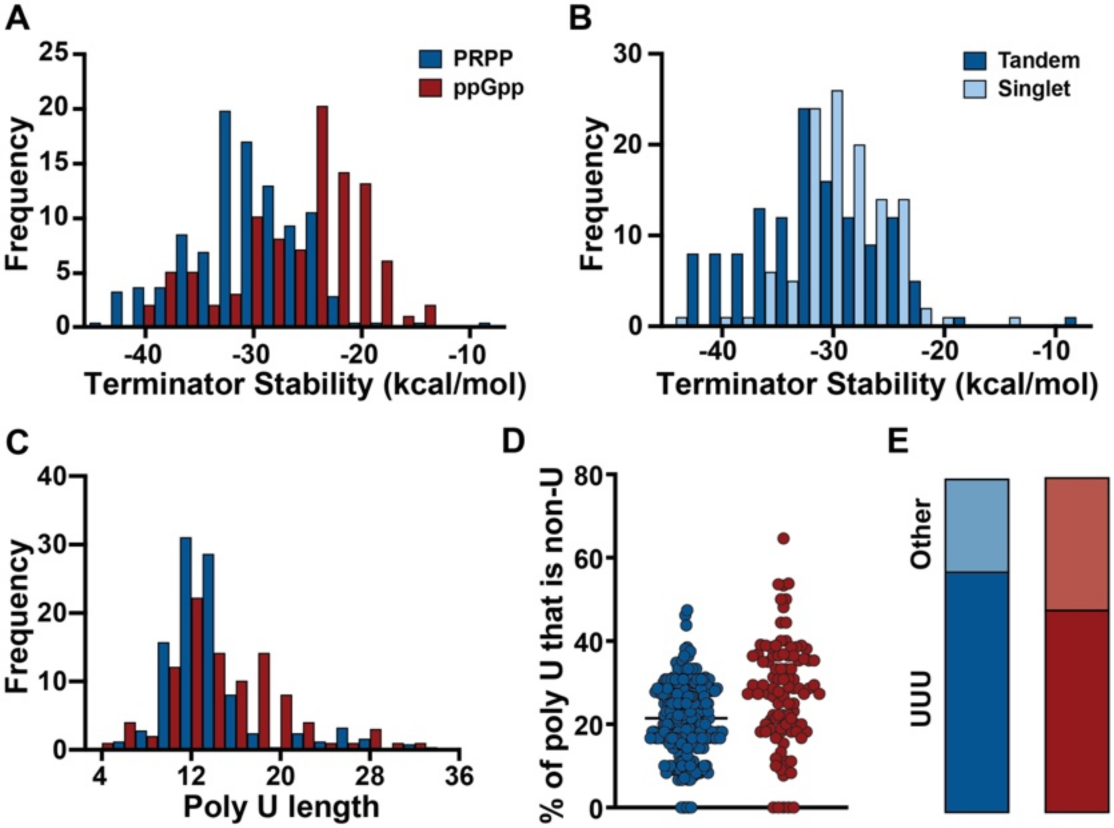

Figure 2.7: Terminator efficiency and stability is distinct between PRPP (blue) and ppGpp (red) riboswitches. (A) Predicated terminator stability of PRPP and ppGpp riboswitches p<0.0001. (B) Terminator energies of singlet and tandem PRPP riboswitches. (C) Poly-uridine lengths of PRPP and ppGpp riboswitches. (D) Uridine content in the poly-uridine pause site p<0.0001. (E) Identity of the first three nucleotides of the poly-uridine pause site p<0.0362.

## 2.3 Discussion

Previous mutational studies on the *ykkC* family of riboswitches have changed the ligand specificity from PRPP to ppGpp using a G93A mutation [64,66]. We have shown that this single nucleotide change is not sufficient to interconvert between ppGpp riboswitches and PRPP switches in a functional assay. Moreover, there was no response to PRPP in the SMARTT dataset despite sequences with the binding site A93G mutation and two additional surrounding mutations. Sequences assessed in this study include the P3 and P0 stems of the PRPP riboswitch consensus sequence. In these cases, the surrounding context in either the paired scaffold regions of P1 and P2 or the terminator must contribute to ligand specificity.

Studies of riboswitches frequently focus on the aptamer domain alone rather than the entire construct [91]. For instance, both binding studies and crystal structures focus solely on the aptamer domain of a riboswitch [64,65]. Notably, mutational studies on the PRPP riboswitch that found altered ligand specificity with G93A used an aptamer-only construct, which is not amenable to functional analysis. While these studies reveal key features of the aptamer, they do not assess the full riboswitch context or functional capabilities. We have shown that the expression platform has an equally important role to play in riboswitch specificity and function. Specifically, it is possible that the single binding site mutation at position 93 could change the ligand binding preference for a PRPP or ppGpp aptamer but the new ligand is incapable of stabilizing the native expression platform against terminator folding. As such, the resulting mutant aptamer would not be functional with the new ligand.

A single GC to GU wobble in the terminator helix is sufficient to increase the riboswitch sensitivity 2-fold while decreasing the amplitude nearly 3-fold. This suggests that while the terminator has a known link to riboswitch readthrough and function, it is also tightly connected with ligand binding. Additionally, there were many sequences in this dataset with terminator mutations that significantly increased riboswitch sensitivity without impacting $Y_{max}$ and $Y_{min}$. These sequences were concentrated in the middle of the terminator stem, where it appears that hairpin breathing is more tolerated.

Furthermore, terminator energetics of natural *ykkC* sequences were found to be ligand specific. PRPP riboswitches, both tandems and singlets, had stronger terminator helices and pause signals than ppGpp riboswitches. Together, these suggest a stronger OFF state that ligand binding must overcome in order to activate gene expression. Compared to PRPP, ppGpp contains an additional phosphate and the guanine base. Despite their structural similarities, the additional components in ppGpp lead to more potential non-covalent interactions with the riboswitch. Overall, ppGpp can supply more energy to stabilizing the bound conformation through these non-covalent interactions. Thus, the result that PRPP switches have stronger termination signals is surprising. This suggests that readthrough transcription in the absence of ligand is less penalized in the ppGpp riboswitches than in the PRPP riboswitches.

The *T. oceani* ppGpp riboswitch used in this study has a $Y_{min}$ of 49±2%, which corresponds to a weaker terminator. Moreover, most ppGpp riboswitches have weaker terminators than PRPP riboswitches. Biologically, this implies that the synthesis of branched-chain amino acids is a less tightly regulated process. Instead, addition of the alarmone ppGpp increases transcription of the full-length product, thus increasing the production of branched-chain amino acids. In contrast, the *P. micra* PRPP riboswitch in this study has a $Y_{min}$ of 18%, suggesting that nucleotide metabolism is much more tightly regulated and synthesis in the absence of PRPP is unfavorable. Tuning a riboswitch sensitivity and response to a biologically relevant level is a key feature of riboswitch-driven gene regulation [91]. This could help explain the diversity of expression platforms observed in nature.

This work has emphasized the importance of the expression platform. This has always been recognized as an integral part of riboswitch function, but the specificity of each expression platform for a single aptamer is often overlooked. The idea of a plug-and-play system in which an aptamer can be added to any expression platform is particularly appealing to synthetic or engineered systems as it reduces the complexity of the system [91–93]. Frequently, aptamers for specific ligands are engineered using SELEX with the hope that an expression platform can be appended later [92,94,95]. Research has already shown that expression platforms are generally not transferrable without sequence modifications [92,93,96,97]. Here, we showed that in addition to transducing a ligand binding signal into a change in gene expression, the expression platform is involved in ligand specificity and, therefore, an integral component of aptamer evolution. In such instances, SMARTT provides an optimal platform from which to build an engineered riboswitch with functional readout.

In addition to highlighting the importance of the expression platform in engineered switches, this work has highlighted a key struggle in variant riboswitch identification. In addition to the *ykkC* family, other variant riboswitches have been identified based on a small number of nucleotide changes, including the purine and cyclic-dinucleotide riboswitch families. Since the discovery of the guanine riboswitch in 2003, variants have been identified that bind adenine, xanthine, or 2′-deoxyguanine [47,48,54,98]. These variants each have unique gene associations and changes to the ligand binding pocket to allow for altered ligand specificity. Similarly, two subclasses of the cyclic-dinucleotide riboswitches have been identified [49,51,99,100]. However, there are examples of these cyclic-dinucleotide riboswitches that show no affinity for either cyclic-di-GMP or cyclic-AMP-GMP, the two known ligands [57,87]. It is possible that these sequences are evidence of an unidentified variant with altered ligand specificity. As seen with the ykkC family, these families may contain further variants that are masked within the current known subclasses.

As seen with the PRPP and ppGpp variants, changes in nucleotides identified through crystal structures and gene association are not always sufficient to distinguish between variants. Sites distal to the ligand binding site, long range interactions, and terminator energetics also play large roles in this differentiation. As such, it is important to use a diverse array of tools to separate subclasses to account for context-dependent sequence differences, un- or mis-annotated gene associations, small variant population, promiscuous subclasses, or subclasses that recognize ligands not yet known to science. Moreover, existing riboswitch classes may contain variants that remain unidentified because the sequence and gene association differences are unknown.

# 3. Classification of a sixth *ykkC* variant class

This chapter is the beginning of a story that takes what I learned from high-throughput sequencing of the *ykkC* riboswitch and builds upon it to identify a new potential variant subclass.

## 3.1 Background

Variant riboswitches are RNAs that are similar in sequence and structure but recognize different ligands [87]. Over ten of the nearly 60 classes of riboswitches are variants [46,88]. The majority of variant families recognize structurally similar ligands [47,48,53,54]. The *ykkC* family is an exception to this. This family encompasses five known variants and recognizes structurally and biochemically diverse ligands, including guanidine, phosphoribosyl pyrophosphate (PRPP), and guanosine tetraphosphate (ppGpp) (figure 3.1) [14,32,58–60].

The *ykkC* variants all share the same scaffold despite large structural differences between ligands. This scaffold consists of two co-axially stacked helices that serve as a docking site for additional ligand recognition helices. The first *ykkC* variant identified, labeled class I, recognizes guanidine and regulates genes for guanidine metabolism and transport (figure 3.1A) [32]. This class has a single helix that docks to the scaffold and recognizes a guanidinium ion. To achieve specificity for the small cation, every face of the ligand is recognized using ionic interactions, cation-$\pi$ interactions, and hydrogen bonding [38,65]. The majority of these interactions use nucleotides in the

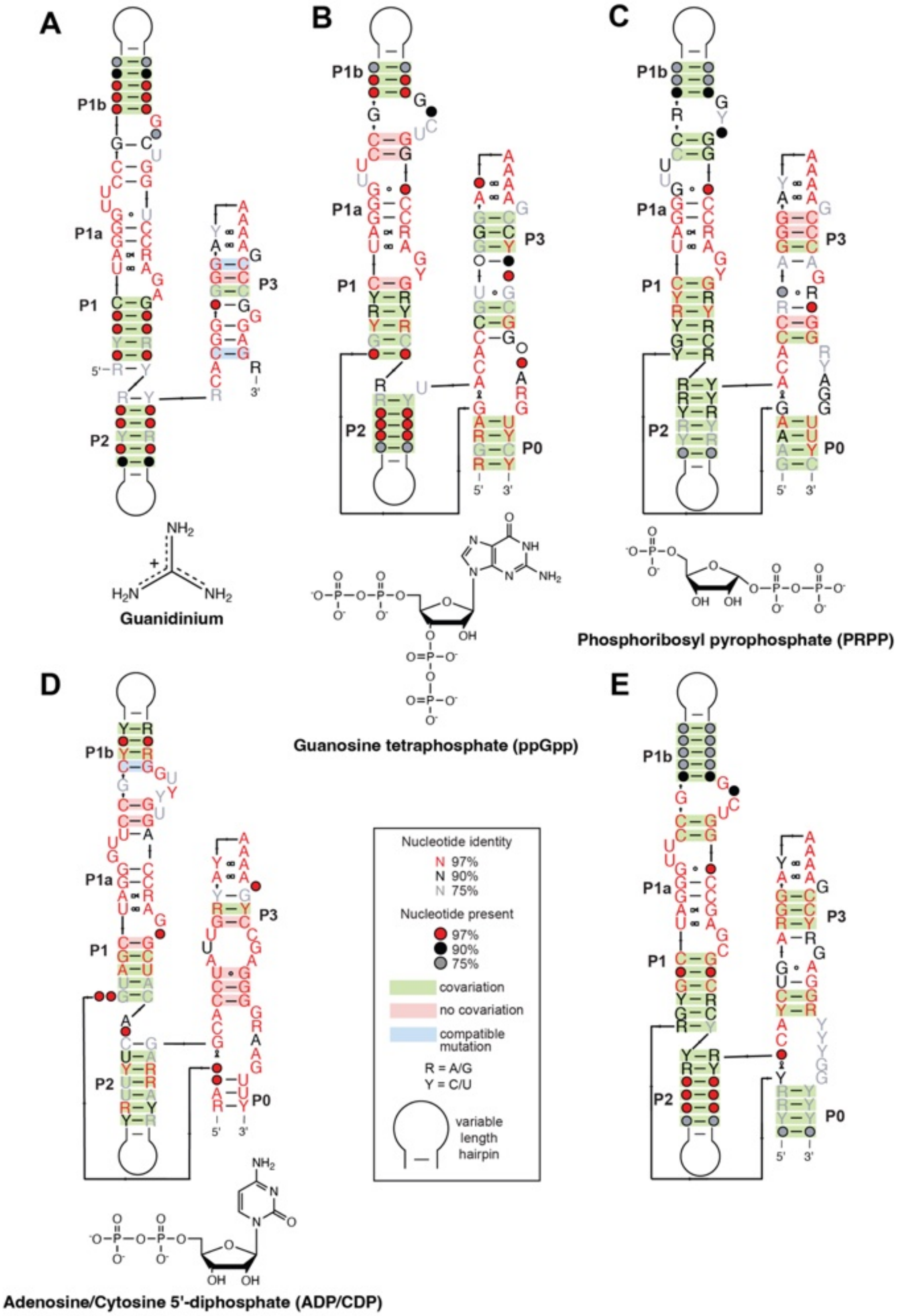

Figure 3.1: Consensus sequence for *ykkC* variants. (A) Consensus for *ykkC* class I variants, guanine riboswitches. (B) Consensus for *ykkC* type 2a variants, ppGpp riboswitches. (C) Consensus for *ykkC* type 2b variants, PRPP riboswitches. (D) Consensus for *ykkC* type 2c variants, ADP/CDP riboswitches. (E) Consensus for *ykkC* type 2d variants.

add-on P3 helix. A single guanine nucleotide from P1 is also involved in direct ligand contact. This guanine is conserved across all five known variants and appears to be involved in direct ligand contact for all crystallized variants. As such, it may serve as a link between the scaffold and the add-on helices to stabilize the bound state of the riboswitches.

All remaining *ykkC* subtypes are classified as class 2. This class is further divided into four subgroups: 2a, 2b, 2c, and 2d. All of the class 2 variants contain an additional P0 helix that docks below the P3 helix found in guanidine riboswitches. This extends the binding pocket and shifts it down to allow for recognition of larger ligands [64,66]. Specifically, class 2a recognizes the alarmone ppGpp, class 2b recognizes the purine precursor PRPP, class 2c recognizes nucleotide diphosphates, and class 2d remains an orphan (figure 3.1B, C, D, & E) [14,59–61]. These larger polyanionic ligands require distinct recognition mechanisms compared to guanidinium. Specifically, coordinated metals are required to recognize the anionic phosphates in ppGpp, PRPP, and nucleotide diphosphates. Similar to guanidine riboswitches, nearly all direct ligand contacts occur using nucleotides in the add-on helices rather than the scaffold. As a consequence, changing ligand specificities most often requires mutations to P3 rather than the scaffold helices P1 and P2.

The high sequence and structural similarities between the *ykkC* variants have made this class difficult to separate. PRPP and ppGpp riboswitches were differentiated based on the identity of a single nucleotide at position 93. In

PRPP sequences, G93 base pairs with C74 as part of the tertiary structure [64,66]. However, in ppGpp sequences, C74 is paired with the ligand, and position 93 can be any nucleotide except a G [59,64,66]. This, along with genetic context, is able to delineate between the two classes. Further classes were differentiated based on genetic context alone and sequence differences were identified based on the bioinformatic consensus diagram generated following separation. Notably, because they were largely separated based on gene association, un- or mis-annotated gene contexts at the time of separation could result in mis-annotated riboswitches. Additionally, given the plasticity of the *ykkC* scaffold and the closeness in sequence space between the class 2 variants, it is plausible that additional variant classes exist.

I applied bioinformatic approaches to investigate the possibility of mis-annotation of the *ykkC* class 2 variants. These sequences contain G93 but are phylogenetically related to ppGpp riboswitches. I confirmed that these sequences bind neither PRPP nor ppGpp despite close sequence similarities to both classes of riboswitches. Instead, these sequences bind XMP and GMP and correspond to a potential sixth variant class.

## 3.2 Results and Discussion

### 3.2.1 *ykkC* variants with G93 associate with a glutamine hydrolyzing GMP synthase gene

Using a high-throughput mutational assay, SMARTT, on a ppGpp riboswitch I found that ppGpp riboswitches that harbor the ligand binding site

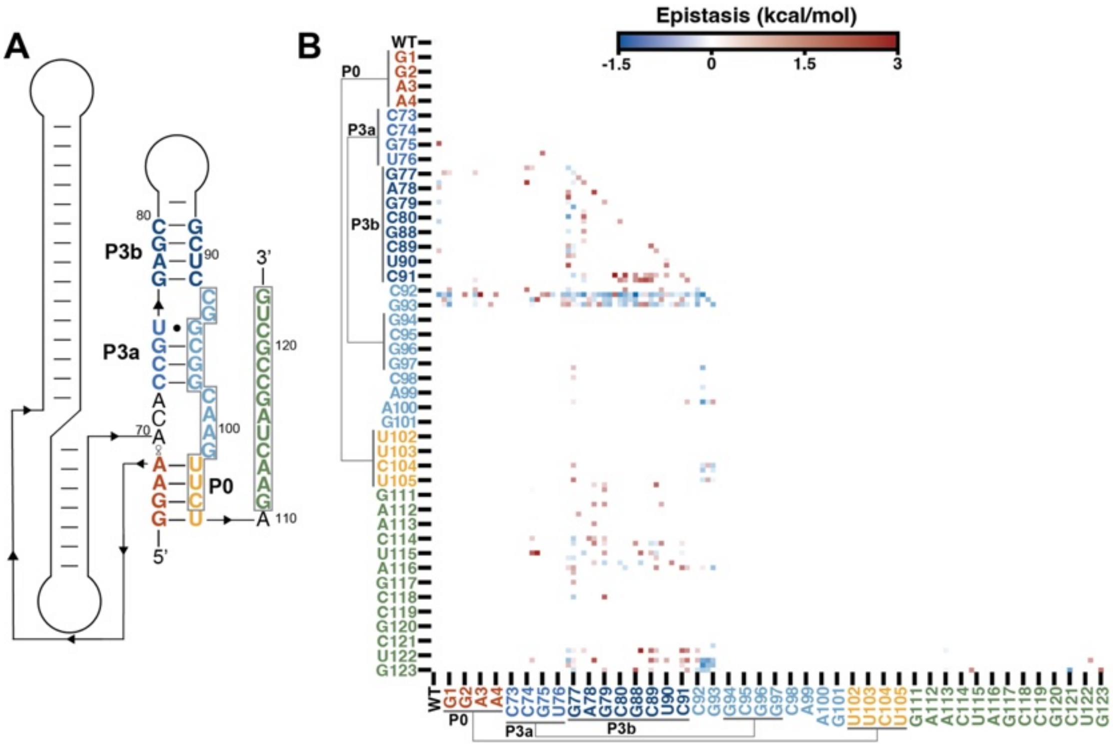

Figure 3.2: G93 does not occlude ppGpp driven function. (A) Secondary structure of the *T. oceani* ppGpp riboswitch with A93G mutation and all other mutated positions colored. (B) Epistasis values for all single and double mutants in the A93G library with ppGpp.

mutation A93G are still capable of responding to ppGpp. This conflicts with the bioinformatic assumptions that G93 riboswitches respond specifically to PRPP because C74 is unavailable to base pair with ppGpp (figure 3.2). Moreover, using a high throughput approach, I identified a number of additional sequence contexts in which G93 remains functional with ppGpp. Based on this result, I hypothesized that there could be natural G93 riboswitches that are not specific for PRPP. A phylogenetic tree was generated using the PRPP and ppGpp variants to assess the relatedness of all annotated sequences. A small cluster of natural G93-containing sequences map preferentially to a ppGpp region over the expected PRPP region (figure 3.3). These sequences could be natural G93 ppGpp riboswitches or they could represent a new *ykkC* variant class.

All of the sequences in this cluster are associated with the guaA gene. GuaA encodes for a glutamine-hydrolyzing GMP synthase, which converts xanthosine 5′-monophosphate (XMP) into guanosine 5′-monophosphate (GMP) (figure 3.4). ppGpp riboswitches have not been found to associate with this gene previously, and it is, therefore, unlikely that ppGpp is the natural ligand for this cluster of sequences. PRPP riboswitches regulate purine biosynthesis. It is possible that this riboswitch responds to PRPP; however, this would be the first instance of a PRPP riboswitch that regulates a process between IMP and GMP. Instead, it is likely that this cluster is a variant RNA with altered ligand specificity.

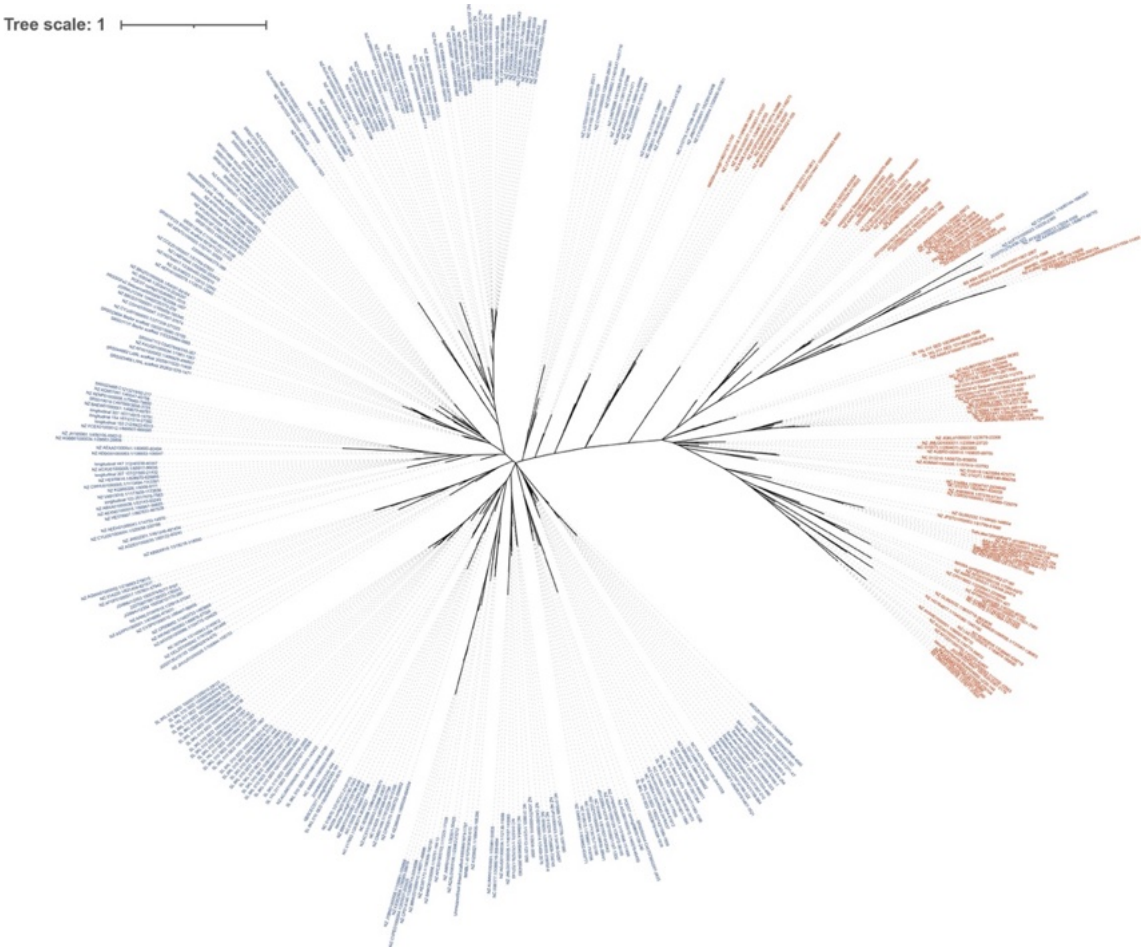

Figure 3.3: Phylogenetic tree of ppGpp and PRPP annotated sequences. ppGpp annotated sequences are shown in red and PRPP sequences are shown in blue. The tree was generated using a maximum likelihood model.

### 3.2.2 *ykkC* variants sense XMP and GMP

To investigate the possibility that this putative RNA aptamer is a variant of the *ykkC* family, I assessed ligand binding using microscale thermophoresis (MST). MST is a technique that measures the molecular interactions of two biomolecules by monitoring changes in the mobility of a single species through a temperature gradient. In this study, the RNA aptamer is fluorescently labeled and its mobility was measured in the presence of various amounts of different ligands. Binding of the ligand to the labeled RNA will induce a measurable change in macromolecular mobility, typically due to changes in hydration shell, charge, or size. In the case of RNA aptamers, ligand binding tends to induce a more compact fold that can be monitored using MST. MST has previously been validated using riboswitch aptamers including *ykkC* variants [65,101].

I designed a minimal construct for binding assays from the *Actinokineospora inagensis* (*A. inagensis*) putative aptamer. This construct contains a G at the 5′-end of the predicted P0 stem to facilitate *in vitro* transcription and ends after the final nucleotide at the 3′-end of the predicted P0 helix. MST results showed that the *A. inagensis* aptamer binds to both XMP and GMP (figure 3.4). Specifically, it binds to XMP with a dissociation constant ($K_D$) = 7±4 μM and it binds to GMP with a $K_D$ = 40±20 μM. Enzymes that recognize XMP and GMP have $K_M$ values of 3 μM and 5-1000 μM, respectively.

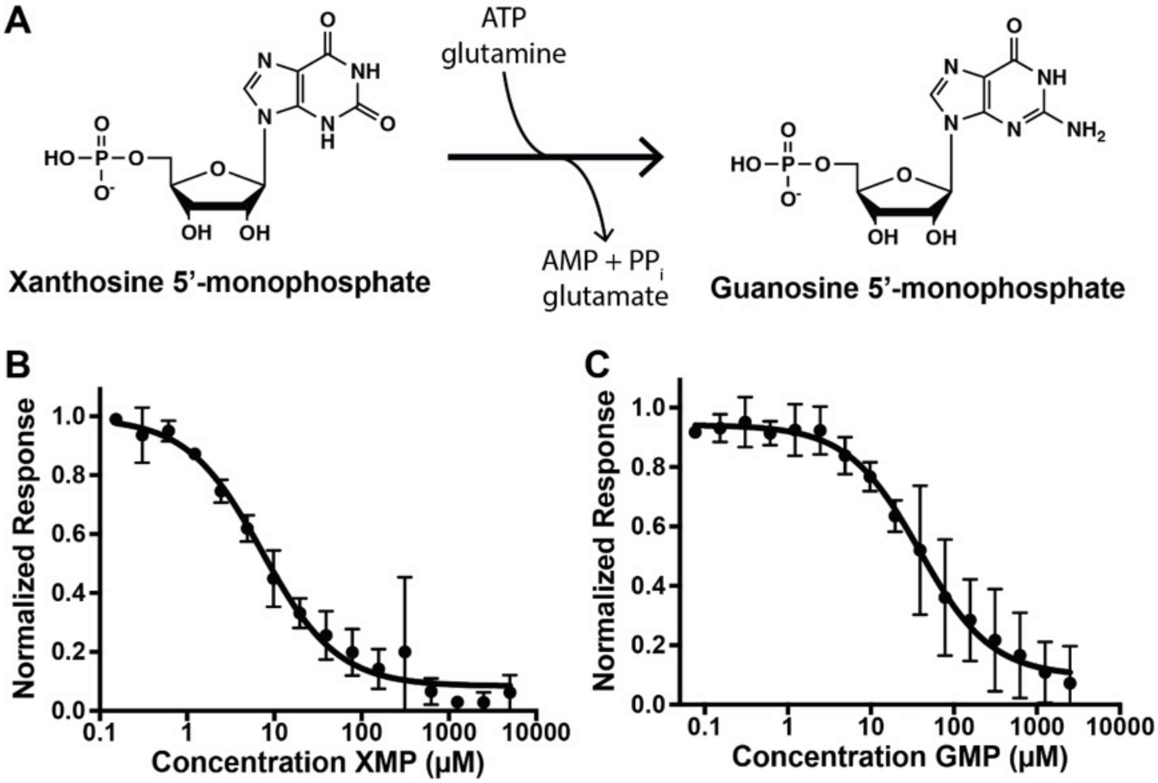

Figure: 3.4: Binding response of the *A. inagensis* aptamer to XMP and GMP. (A) The synthesis of GMP from XMP using a glutamine hydrolyzing GMP synthase. (B) The binding of XMP to the aptamer using MST. (C) The binding of GMP to the aptamer using MST.

This means that the measured $K_D$ for both XMP and GMP are within a biologically relevant range.

In addition to XMP and GMP, compounds with different bases and with different numbers of phosphates were tested for binding. Despite the sequence similarity to both ppGpp and PRPP riboswitches, the *A. inagensis* aptamer rejects both of these compounds. Additionally, this variant aptamer is selective for a single phosphate at the 5′-end and does not bind to guanosine diphosphate (GDP) or guanosine triphosphate (GTP). Moreover, guanine did not bind, indicating a preference for the presence of a sugar. Finally, I tested different base identities. The *A. inagensis* aptamer discriminates against both IMP and AMP. This suggests that hydrogen bonding to a functional group at C2 is essential to ligand recognition.

### 3.2.3 Functional Assays

To prove that these variant aptamers are functional riboswitches, I completed a series of functional assays. Unfortunately, there is no clear regulation mechanism. The full-length construct is predicted to fold such that the Shine-Dalgarno sequence, the ribosome binding site, is occluded; thus, it is possible that this sequence is regulated through translation initiation. A RelE initiation assay found that this is not the case.

In the absence of a clear regulation mechanism, I used a reporter assay with the *A. inagenesis* riboswitch construct to test *in vivo* function. This makes it possible to test the functional switching of a putative riboswitch in cells without knowing the regulation mechanism. The full riboswitch sequence, including the expression platform, is cloned upstream of a β-galactosidase reporter gene and then expressed in *E. coli* cells. Both GMP and XMP are membrane-permeable and I predicted that adding ligand to the media would be sufficient to change the intra-cellular ligand concentrations and generate a change in β-galactosidase expression, which is monitored by a change in enzyme activity. Cells expressing the reporter construct were grown with ligand overnight, and then β-galactosidase activity was monitored using a Miller Assay. I predicted that if the native ligand is GMP, the *A. inagensis* variant is an OFF switch, and if the native ligand is XMP, it is an ON switch. Initial tests show that there is an increase in β-galactosidase activity with the addition of either XMP or GMP, which suggests an ON switch with a native ligand of XMP. However, follow-up experiments cannot confirm this activity. It seems possible that the cellular concentrations of these ligands are disrupting the functioning of the reporter assay and making it difficult to identify the native ligand.

To minimize the impact of the cellular synthesis machinery on the results of this assay, I transformed the *A. inagensis* variant aptamer reporter construct into mutant cell lines that are deficient in the GMP synthesis pathway. I identified three mutants in the GMP synthesis pathway to test: ΔGuaA cannot convert XMP to GMP, ΔGuaB cannot synthesize XMP from IMP, and ΔGuaC cannot convert GMP back into IMP (figure 3.5A). The expectation is that ΔGuaA

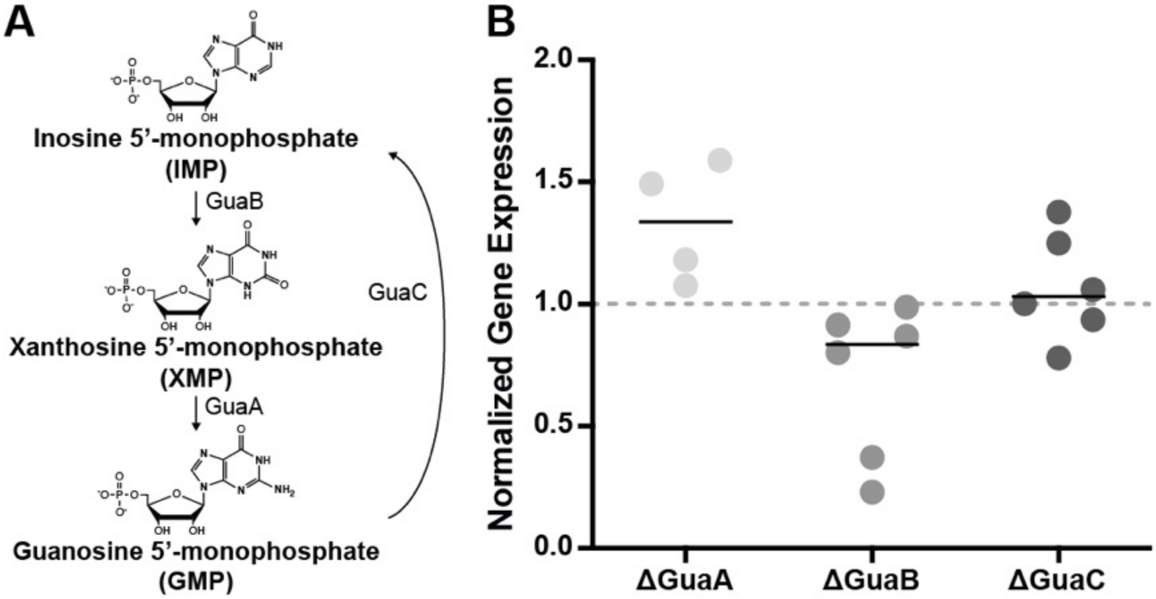

Figure 3.5: β-galactosidase reporter assay of the *A. inagensis* aptamer in GMP synthesis pathway deficient cell lines. (A) The GMP synthesis pathway using GuaB, GuaA, and GuaC. (B) Gene expression in each KO cell line normalized to WT gene expression.

should increase gene expression regardless of the native ligand, while ΔGuaB and ΔGuaC can help differentiate between XMP and GMP as the native ligand.

Preliminary results show that ΔGuaA increases gene expression while ΔGuaB decreases gene expression compared to WT (figure 3.5B). In contrast, ΔGuaC shows minimal change in gene expression. This suggests that the native ligand for this *ykkC* variant is XMP. Confirmation that the addition of XMP increases gene expression is necessary to finalize this conclusion. Additionally, conclusive evidence of riboswitch-ligand interactions via X-ray crystallography and a clear understanding of the regulation mechanism could be used to further support these results.

## 3.3 Discussion:

The *ykkC* scaffold is one of the most evolvable riboswitch scaffolds that has been identified. It is known to contain at least five subclasses that recognize diverse biomolecules, including nucleotide precursors, signaling molecules, and small toxic compounds. These ligands are chemically and structurally distinct, and each subclass uses different nucleotides to specifically recognize a single ligand; however, the classes are similar enough to have been grouped together for over a decade [14,32,58–60].

Recent work investigating the functional relationship between PRPP and ppGpp riboswitches prompted the current investigation into additional *ykkC* subclasses. Phylogenetic analysis of PRPP and ppGpp subclasses identified a small selection of G93 sequences that clustered with ppGpp riboswitches instead of the expected PRPP sequences. Investigation into this cluster confirmed the presence of an unknown aptamer class that regulates genes for a glutamine-hydrolyzing GMP synthase. These aptamers are close enough in sequence space to PRPP aptamers to have been classified with this subgroup upon its discovery in 2017. Additionally, the gene association, while distinct from the purine biosynthesis genes generally associated with PRPP riboswitches, is still related to purine biosynthesis, thus masking the existence of this variant class.

Purine biosynthesis is achieved through many pathways, including recycling of dephosphorylated GTP and ATP; however, *de novo* purine synthesis begins with a pathway that converts PRPP to inosine 5'-monophosphate (IMP). This branch of the pathway is regulated by PRPP riboswitches [14]. Following IMP synthesis, the pathway diverges into GMP or AMP synthesis. To synthesize GMP, IMP-dehydrogenase oxidizes C2 of IMP to make XMP [102]. Then, XMP is converted to GMP using the glutamine-hydrolyzing GMP synthase to aminate the C2 [103,104]. The gene association of this variant class suggests that it is used to regulate GMP biosynthesis, and the identified ligands of GMP and XMP support this finding. It is, however, unclear why this aptamer is binding to both the substrate and the product of the final step in GMP synthesis. Concerningly, the apparent dissociation constants suggest that this aptamer can bind to physiologically relevant concentrations of both of these ligands. Previous binding studies on *ykkC* variant riboswitches have shown that the PRPP aptamer, like the *A. inagensis* aptamer, is capable of binding two ligands. It can bind both PRPP and ppGpp, although with very different dissociation constants. Despite binding studies that suggest promiscuity, a PRPP riboswitch can distinguish between these two molecules functionally. Based on my recent work, I hypothesize that this functional selectivity is related to the expression platform. Specifically, the alternative helices in the expression platform for PRPP riboswitches tend to have a stronger predicted energy than ppGpp riboswitches. The tighter affinity for PRPP compared to ppGpp is predicted to generate the energy required to stabilize the bound confirmation over the alternate helices, making the riboswitch specific. It is possible that this is the case with the GMP/XMP aptamer in this study. Adding an expression platform could result in a specific riboswitch. Unfortunately, this is not a direct comparison because the GMP/XMP variant family has no examples that function through transcription termination. It is also possible that this aptamer, which has no clear recognized regulation mechanism, is a residual riboswitch that has been degraded over time and is no longer a functional gene regulatory RNA. Clarifying the biological role of this aptamer will help determine the functional requirements for specific ligand recognition and genetic switching.

This GMP/XMP aptamer is not the first variant riboswitch that was initially misclassified. The closeness in sequence space combined with unknown gene annotations can make it difficult to fully separate out variant classes.

Specifically, variant sequences with few representatives are easy to overlook, as was the case in this study. These "snugglers" are being identified more often as gene annotations become more complete. A recent example is the 8-oxo-guanine riboswitch, which was initially classified with the purine riboswitches [105]. Another large class of "snugglers" was the cyclic-AMP-GMP variant family [49,50]. In both of these cases, the variant class differs from the parent at a single nucleotide position, making it difficult to separate them without in-depth analysis or new genetic information. The discovery of variant classes frequently leads to the discovery of unique biology, as was the case with the cyclic dinucleotide variants [49,106,107]. The discovery of cyclic-AMP-GMP variants also led to the discovery of cyclic-AMP-GMP and a potential role for this ligand in cells. Classifying more variants accurately could provide a more complete understanding of essential regulated biological processes.

Given the evolvability of the *ykkC* scaffold and the diversity of genes that it regulates and ligands that it binds, it is likely that other variant classes exist as "snugglers". Investigating the sequence requirements for ligand binding in conjunction with the gene association and phylogenetic relatedness could uncover these potential variant classes.

# 4. High-throughput analysis of a cyclic dinucleotide riboswitch

This section is adapted from Barth et al., 2023, an article currently submitted to the journal *Biochemistry*.

## 4.1 Background

Riboswitches are RNA regulatory elements most often located in the 5′ UTR of bacterial genes. They are composed of an aptamer domain, which specifically binds a ligand, and an expression platform, which structurally rearranges in response to ligand binding to manipulate gene expression [108]. The genes regulated by a riboswitch are often tightly linked to the ligand that the riboswitch binds as part of a regulatory circuit [29,109]. For this reason, ligand recognition is highly specific, with nearly all riboswitches being able to respond to a single ligand despite the existence of chemically or structurally similar small molecules.

In some cases, a small number of changes in the nucleotide sequence in the background of a common structural motif is sufficient to alter the ligand specificity of a riboswitch [14,47,49,59,106]. In such instances, the two distinct riboswitch classes are known as variants. Variant riboswitches use the same structural scaffold with minor sequence changes to recognize different ligands [110]. Distinguishing between variant subclasses based on bioinformatics is often difficult due to the high structural and sequence similarity. In many instances, multiple variants were initially classified together and not differentiated into subclasses until much later in the process of discovery [14,32,47,48,53,59,60,110]. Distinct gene association is frequently used to separate variants because of the tight relationship between gene association and ligand identity. However, genes of unknown function or diverse gene associations can make differentiation of variant classes difficult. The cyclic dinucleotide variant family is an example of riboswitches that were difficult to differentiate based on gene association alone.

Cyclic dinucleotides are bacterial second messengers that transmit extracellular signals to receptors within the cell. The best-studied cyclic dinucleotides contain two purines: cyclic-di-GMP (c-GG), cyclic-di-AMP (c-AA), and cyclic-AMP-GMP (c-GA) [111–113]. These dinucleotides regulate various cellular functions, including virulence, biofilm formation, pili biosynthesis, and sporulation [111–116].

Three riboswitch classes bind to these cyclic dinucleotides: *ydaO* and cyclic-di-GMP-I and -II [49,51,58,106,117]. The *ydaO* RNAs are specific for c-AA, while c-GG RNAs show various levels of specificity for c-GG and c-AG. c-GG-I riboswitches can be further divided into two variants subclasses, c-GG-I and c-AG (formerly c-GG-Ib) [49,106]. The c-GG-I subclass is highly conserved and widespread, while the c-AG subclass is found predominantly in geobacter [49,51,106]. The cyclic dinucleotide riboswitches were difficult to differentiate based on gene association because of the diverse set of genes that second messengers regulate [58]. Instead, these variants were separated based on

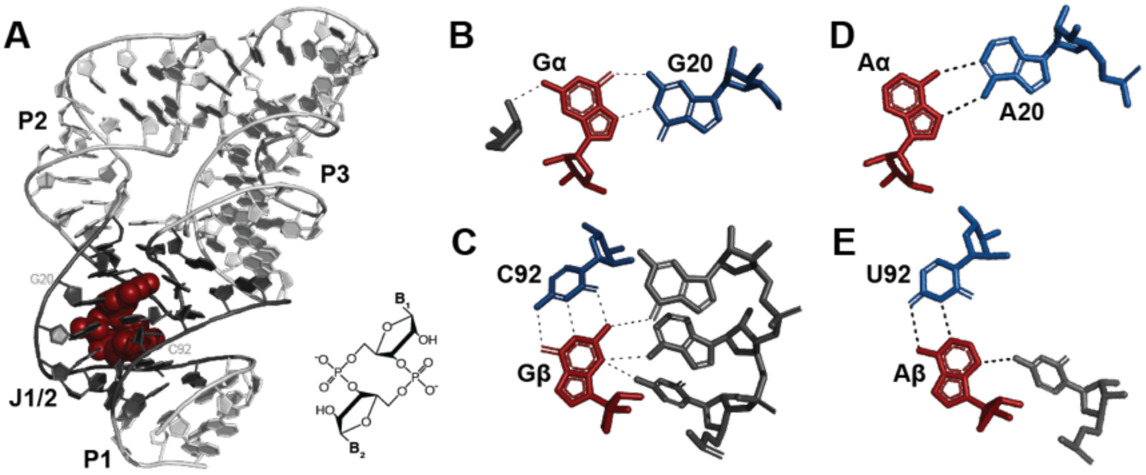

Figure 4.1: Ligand contact with the c-GG variant riboswitches. (A) The cyclic dinucleotide ligand contains two 5′,3′-linked nucleotides where B1 and B2 are purines. It binds at the center of a 3-way-junction in the riboswitch. (B) Gα interacts with G20, (C) Gβ Watson-Crick pairs with C92 of the riboswitch, (D) Aα interacts with A20, or (E) Aβ Watson-Crick pairs with U92.

structure and nucleotide conservation. Specifically, crystal structures identified nucleotides that make contact with the ligand (figure 4.1). These residues showed a surprisingly low level of conservation despite their apparent role in ligand specificity [56,118–125].

Crystal structures of a c-GG-I riboswitch show that G20 and C92 make essential contacts with the ligand [56,99,118,119,125]. G20 interacts with G$\alpha$, the top ligand position, and C92 base pairs with G$\beta$, the bottom ligand position as oriented in figure 4.1B & C. Biochemical mutational studies of a c-GG-I riboswitch verified that mutating either G20A or C92U could change the ligand specificity to an A-containing dinucleotide (figure 4.1D & E) [56,118,119,122–124]. A follow-up to these structural studies investigated the conservation of G20 and C92 and found that, despite their direct role in ligand binding, conservation was relatively low. This prompted research into variant sequences and the discovery of the c-AG subclass [49,106]. The c-AG variants have an A at position 20 instead of the conserved G20 found in c-GG riboswitches [49,106]. This orients the ligand with AMP in the $\alpha$ binding pocket and GMP in the $\beta$ binding pocket (figure 4.1C & D) [49,99,106].

Notably, a subset of the A20-containing riboswitches, expected to be c-AG-responsive riboswitches, showed a response to both c-AG and c-GG [106]. These promiscuous riboswitches have been the focus of biochemical investigations to determine why ligand binding is not specific. Keller et al. found that c-GG binding induces protonation of A20 in a promiscuous c-AG riboswitch, which can hydrogen bond with the Hoogsteen face of G$\alpha$, resulting in the observed promiscuity [100]. Not all A20 sequences are promiscuous, which suggests that there is additional sequence context driving this observation. However, there is no clear way to predict which A20 sequences might be promiscuous based on the nucleotides known to be essential to specificity.

While the discovery of c-AG riboswitches can explain the low conservation at position 20, position 92 also shows surprisingly low conservation, given its direct role in ligand binding. Previous studies have found that some natural variants at position 92 are not functional riboswitches; however, mutational studies indicate that U92 riboswitches can be functional [49,57,123,124]. Moreover, these U92 sequences can bind both c-GG and c-AG with a preference for c-AG [123,124]. A small number of natural U92 riboswitches have been identified [49,57]. It is possible that these U92 riboswitches respond to c-AG despite their classification as c-GG riboswitches. Additionally, these U92 sequences are potential promiscuous sequences based on the results of the mutational studies.

It is possible that riboswitch promiscuity derives from the sequence context and can be explained by nucleotides in the second-shell of ligand interaction. A comprehensive mutational analysis can aid in exploring the contribution of peripheral nucleotides to ligand specificity. Incorporating a functional assay to the study of variants provides a deeper understanding of how small changes in riboswitch sequence impact ligand specificity as well as riboswitch function. SMARTT (sequencing-based mutational assays of RNA transcription termination) is a high-throughput mutational assay that generates quantitative ligand-based transcription termination data for thousands of mutants simultaneously [82]. SMARTT can be used to simultaneously assess the functional consequences of all single and double mutations to a variant riboswitch in the presence of different ligands. This quantitatively assesses how first- and second-shell mutations contribute to altered ligand specificity that is observed in natural variants.

Here, I report that a c-GG-I variant with U at position 92 is promiscuous for both c-GG and c-AG. SMARTT was applied to investigate the effect of thousands of mutants at first- and second-shell residues on riboswitch specificity and promiscuity. These data provide an example of how nucleotides outside the ligand binding pocket influence riboswitch specificity.

## 4.2 Results and Discussion

### 4.2.1 A Cyclic dinucleotide riboswitch with U92 is promiscuous for both cyclic-di-GMP and cyclic-AMP-GMP

To investigate the role of position 92 in ligand recognition and specificity, I assessed a c-GG-I riboswitch with U92 instead of the canonical C92. Unlike G20, which is in the aptamer domain alone, position 92 makes key contacts in both the bound aptamer conformation (figure 4.2A) and the alternate terminator helix (figure 4.2B). As such, a mutation at position 92 can impact both ligand binding and helical switching energetics. The majority of studies on riboswitches have focused on the aptamer, the ligand binding domain, in isolation [49,51,56,99,118,119]. These studies include binding assays and RNA crystal structures. However, riboswitches rely on a helical rearrangement for proper gene regulation *in vivo*. Thus, studying a riboswitch in a functional context, including an expression platform, is essential to understanding function.

An *in vitro* transcription termination assay measures riboswitch function. It leverages the length difference between terminated and full-length transcripts to assess the impact of ligand-driven conformational changes to the RNA. Three key fit parameters are identified from ligand-response curves: $Y_{min}$, $Y_{max}$, and $K_{1/2}$. The $Y_{min}$ describes the apparent equilibrium between the termination state and the readthrough states of the riboswitch in the absence of ligand. The $Y_{max}$ describes the apparent equilibrium between the termination state and the readthrough state extrapolated to infinite ligand concentration. The relationships between these states can be manipulated by changing the stability of either the readthrough or the terminator conformation. The $K_{1/2}$ describes the apparent equilibrium between the bound and unbound conformations. Specifically, the $K_{1/2}$ measures the riboswitch sensitivity to ligand. Increased sensitivity to ligand, a lower $K_{1/2}$, indicates a shift in the apparent equilibrium toward the ligand-bound states. With these three key parameters, the full riboswitch energetics can be described and analyzed.

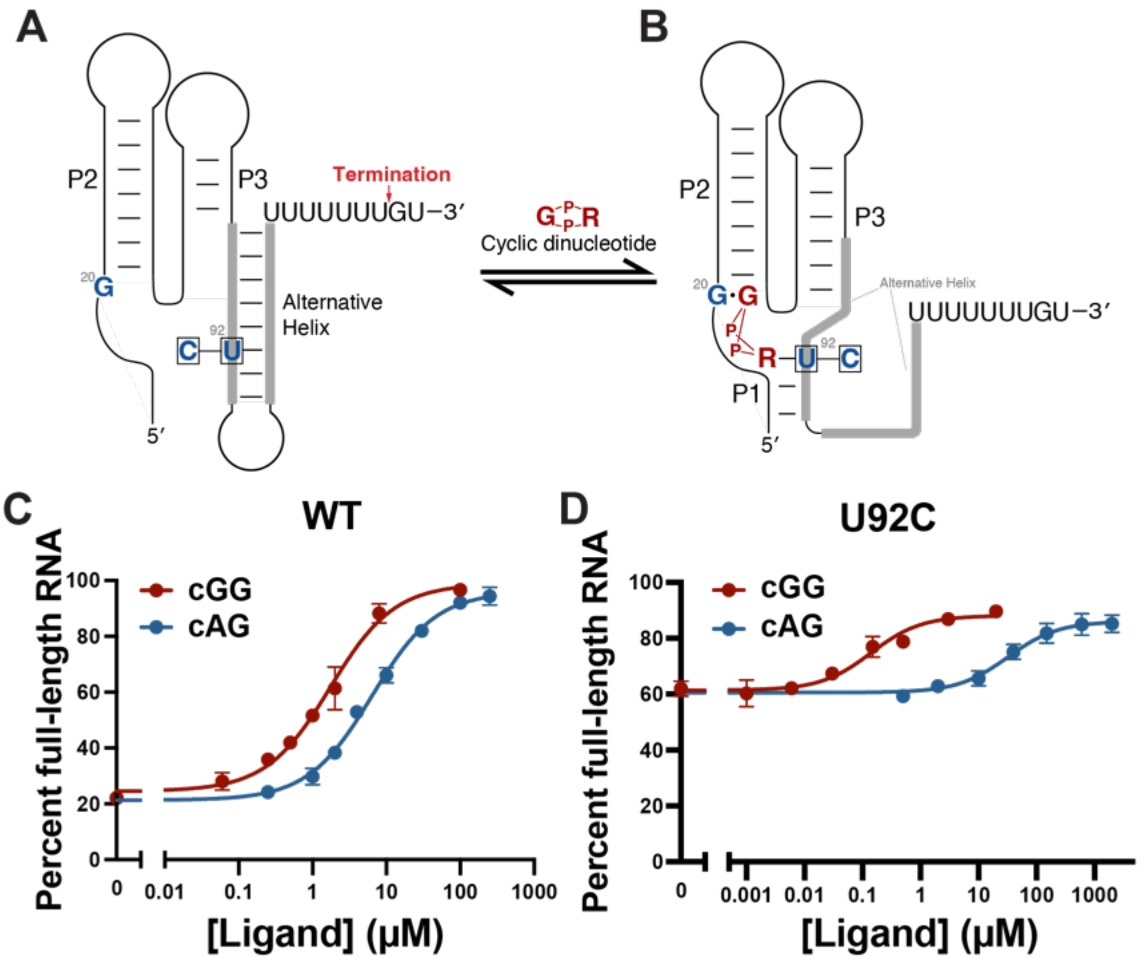

Figure 4.2: Transcription termination of the *A. axanthum* c-GG riboswitch. (A) In the absence of ligand, the riboswitch is in the OFF state, and transcription terminates at position 115. (B) Adding ligand stabilizes the ON readthrough state. (C) The riboswitch responds to c-GG with a $K_{1/2}$ of 1.8 μM and an amplitude of 74% and responds to c-AG with a $K_{1/2}$ of 6.3 μM and an amplitude of 75%. (D) The riboswitch with a single binding site mutation, U92C, is specific for c-GG with a $K_{1/2}$ of 150 nM and an amplitude of 27%.

A c-GG-I riboswitch from *Acholeplasma axanthum* (*A. axanthum*) was selected for evaluation (figure 4.2). The *A. axanthum* transcriptional c-GG-I riboswitch is predicted to be an ON switch that regulates a gene of unknown function. In the absence of ligand, the riboswitch forms a terminator hairpin followed by a poly-uridine tract that terminates transcription at nucleotide 115 (figure 4.2A). In the presence of ligand, an alternate conformation is stabilized that occludes the terminator helix (figure 4.2B). This alternate ON conformation results in full-length transcription of the gene. The *A. ananthum* riboswitch was selected because it contains U92 instead of C92, which is found in most c-GG-I riboswitches. Crystal structures of the cyclic dinucleotide riboswitches show that the nucleotide at position 92 makes a canonical Watson-Crick base pair with the $\beta$ nucleotide of the ligand (figure 4.1). U92-containing switches are, therefore, expected to prefer c-AG instead of the annotated c-GG.

The *A. axanthum* riboswitch was tested with c-AG and found to respond with a half-maximal termination ($K_{1/2}$) at $6.3 \pm 0.3$ $\mu$M and an amplitude of $75 \pm 3$% (figure 4.2C). Transcription termination in the presence of c-GG also resulted in a ligand-dependent response with a half-maximal termination ($K_{1/2}$) at $1.8 \pm 0.4$ $\mu$M and an amplitude of $74 \pm 2$% (figure 4.2C). This riboswitch is promiscuous for both ligands tested with a slight specificity for c-GG (3.5-fold) despite U92 in the ligand binding pocket. This is the first reported sequence for a natural U92 riboswitch that is functional with either c-AG or c-GG. Previous examples of functional U92 cyclic dinucleotide riboswitches were generated synthetically, and all tested natural examples showed no function [49,56,57,118,119,122–124].

In the absence of an annotated gene association, the consequences of a natural promiscuous riboswitch are unclear. It is possible that specificity is unnecessary in the *A. axanthum* riboswitch. No CD-NTases that are known to synthesize c-AG have been identified in *A. axanthum*, which suggests that c-AG is not a naturally occurring signaling molecule in this organism [126]. Thus, there is no evolutionary pressure for specificity in this c-GG riboswitch.

The biological consequence of promiscuity in other naturally occurring riboswitches has not been thoroughly investigated. There are some promiscuous c-AG switches that have been identified [100,106]. Unlike the *A. axanthum* riboswitch used in this study, these promiscuous switches exist in systems that synthesize c-AG and thus are exposed to both potential ligands *in vivo*. Unlike other c-AG switches, which regulate cytochrome genes, these promiscuous switches regulate subtilase family serine protease genes [49,106]. The subtilase family serine protease has been linked to virulence, which corresponds well with the biological role for c-GG [127]. Moreover, c-AG in gammaproteobacterial has been linked to virulence [128,129]. It is possible that these promiscuous c-AG riboswitches are responding to both ligands to regulate bacterial virulence based on the overlapping function of these two cyclic dinucleotides. There is no evolutionary pressure for specificity in these riboswitches if the ligands share a function.

To determine if position 92 is responsible for the observed promiscuity, I tested a U92C mutant. This mutation shifts the primary sequence closer to the consensus c-GG sequence [49,51]. The single U to C mutation in the binding pocket is sufficient to generate a riboswitch with a specific response to c-GG. The U92C mutant shows an improved response to c-GG and a weaker response to c-AG. Specifically, it responds to c-GG with a half-maximal termination ($K_{1/2}$) at 0.15±0.05 μM and an amplitude of 27±3%, and c-AG with a half-maximal termination ($K_{1/2}$) at 26±5 μM and an amplitude of 28±4% (figure 4.2D). This corresponds to nearly 200-fold specificity for c-GG.

Notably, the $Y_{min}$ is significantly increased in the U92C construct. This increase suggests a change in the energy required for helical switching. U92 base pairs with A100 in the terminator helix. U92C, therefore, removes this base pair in the terminator, which destabilizes this alternate helix. This destabilization decreases the energy difference between the bound and unbound conformations. As a result, both c-GG and c-AG provide sufficient energy to generate a response despite the predicted preference against c-AG in the U92C context. The inclusion of a second mutation, A100G, stabilizes the terminator helix and returns the $Y_{min}$ to wild-type levels, which also significantly increases the specificity for c-GG (figure 4.3).

U92C drives specificity toward c-GG, indicating that U92 is partially responsible for the promiscuity observed in the WT riboswitch. U92 can Watson-Crick pair with Aβ of c-AG and wobble base pair with Gβ of c-GG

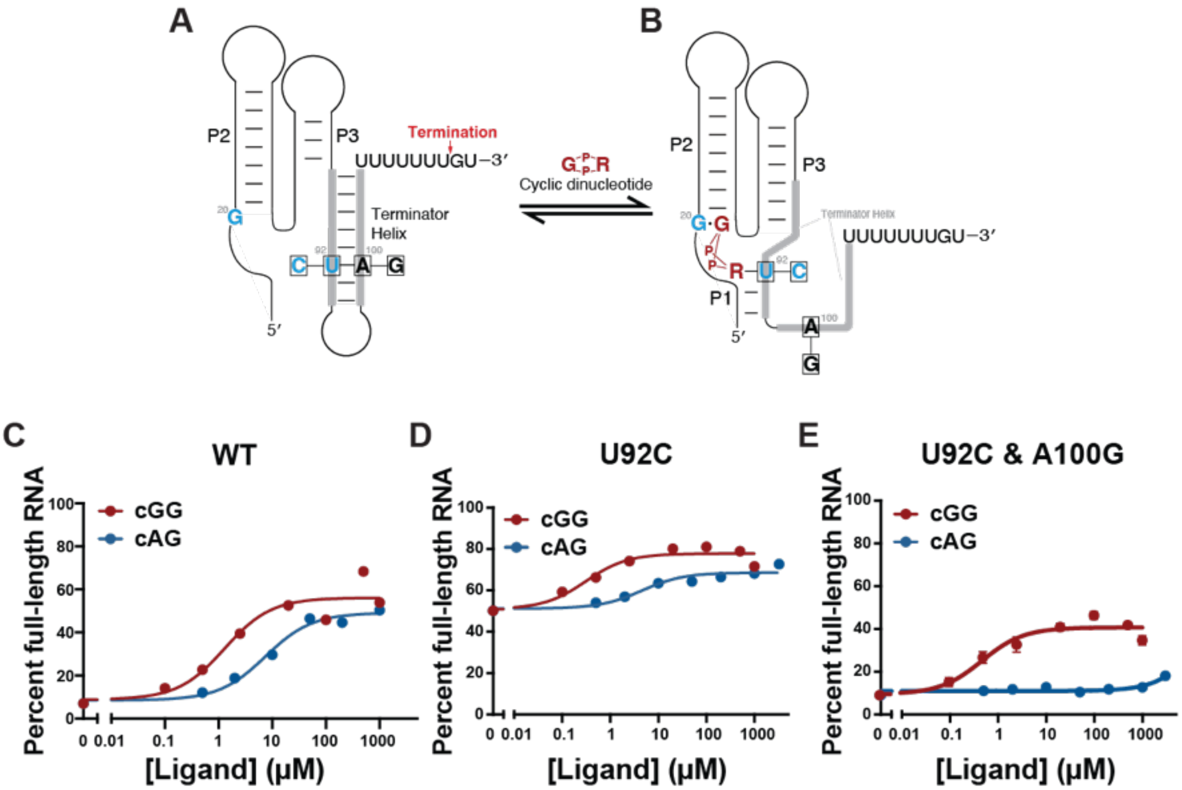

Figure 4.3: Sequencing-based transcription termination of the *A. axanthum* c-GG riboswitch. (A) In the OFF state, U92 base pairs with A100 in the terminator helix. (B) In the ON state, U92 directly interacts with the ligand. (C) The riboswitch responds to c-GG with a $K_{1/2}$ of 1.3 μM and an amplitude of 47% and responds to c-AG with a $K_{1/2}$ of 7.4 μM and an amplitude of 41%. (D) The riboswitch with a single binding site mutation, U92C, is specific for c-GG with a $K_{1/2}$ of 310 nM and an amplitude of 27%. (E) Adding A100G to the U92C mutant stabilizes the terminator and returns the $Y_{min}$ to WT levels. This mutant is specific for c-GG with a $K_{1/2}$ of 480 nM and an amplitude of 31%.

further confirming this result. Using structural arguments, this U92 riboswitch was expected to show a slight preference for c-AG. Specifically, nearest neighbor constraints and helical stacking of P1 and P2 that is bridged by the ligand-U92 interaction both suggest a preference for c-AG [130–132]. In contrast, the WT switch shows a slight preference for c-GG, which cannot be explained by the nucleotides at positions 20 and 92 alone.

### 4.2.2 The cyclic dinucleotide riboswitch is highly tunable

To assess the contribution of nucleotides in the first- and second-shell of ligand contact to riboswitch specificity, I used SMARTT (sequencing-based mutational analysis of RNA transcription termination) (figure 4.4). I mutated nucleotides expected to be in the first- or second-shell of ligand contact as observed in crystal structures.

Full transcription termination ligand-response curves were generated for both c-GG and c-AG for over 7,000 variants encompassing all single mutants and most double mutants in addition to most triple mutants that contain U92C. The transcription termination curves for the wild-type (WT) sequence as generated by SMARTT are comparable to curves generated using the traditional gel-based transcription termination method. For c-GG, I observe a $K_{1/2} = 1.2 \pm 0.4 \mu M$ and an amplitude of 55%, and the response for c-AG has a $K_{1/2} = 3.8 \pm 1.7 \mu M$ and an amplitude of 48% (figure 4.5A & B).

I extracted functional sequences for either c-GG or c-AG. To assess sequences in an unbiased manner, I used a functional parameter that

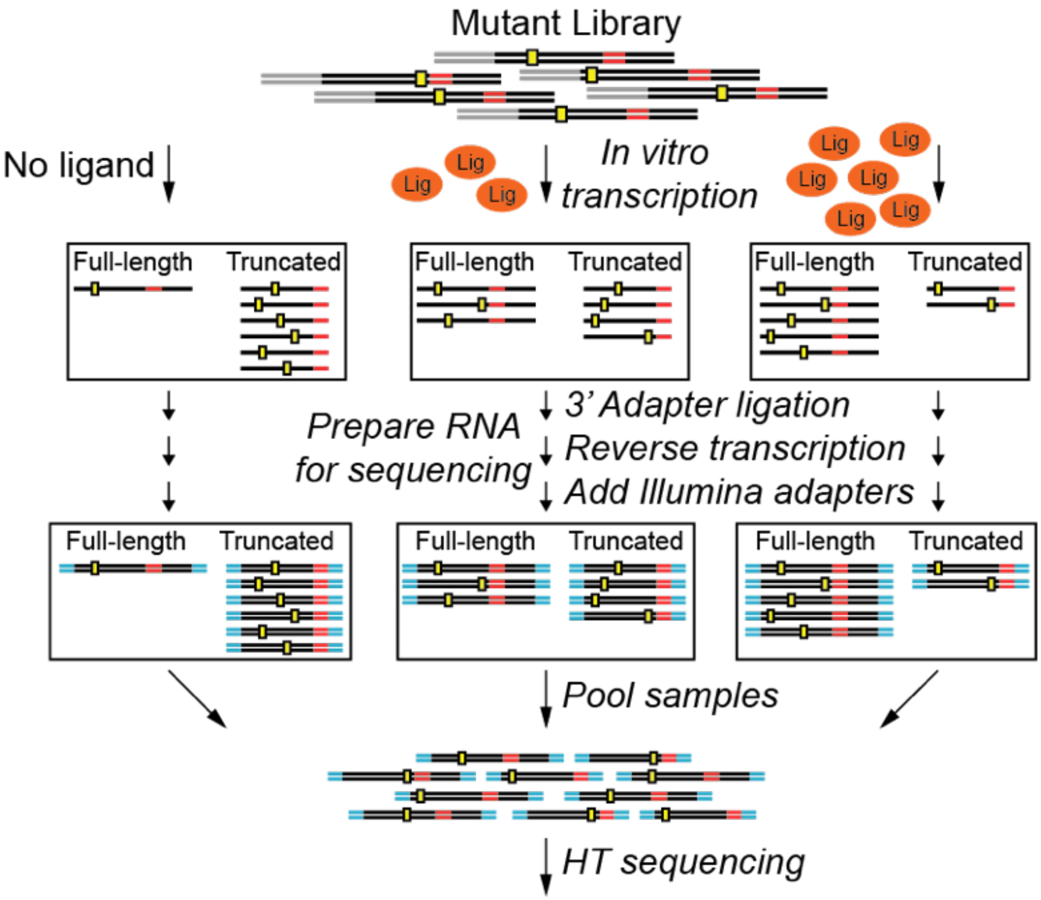

Figure 4.4: The SMARTT workflow begins with a mutant library followed by *in vitro* transcription and preparation for sequencing.

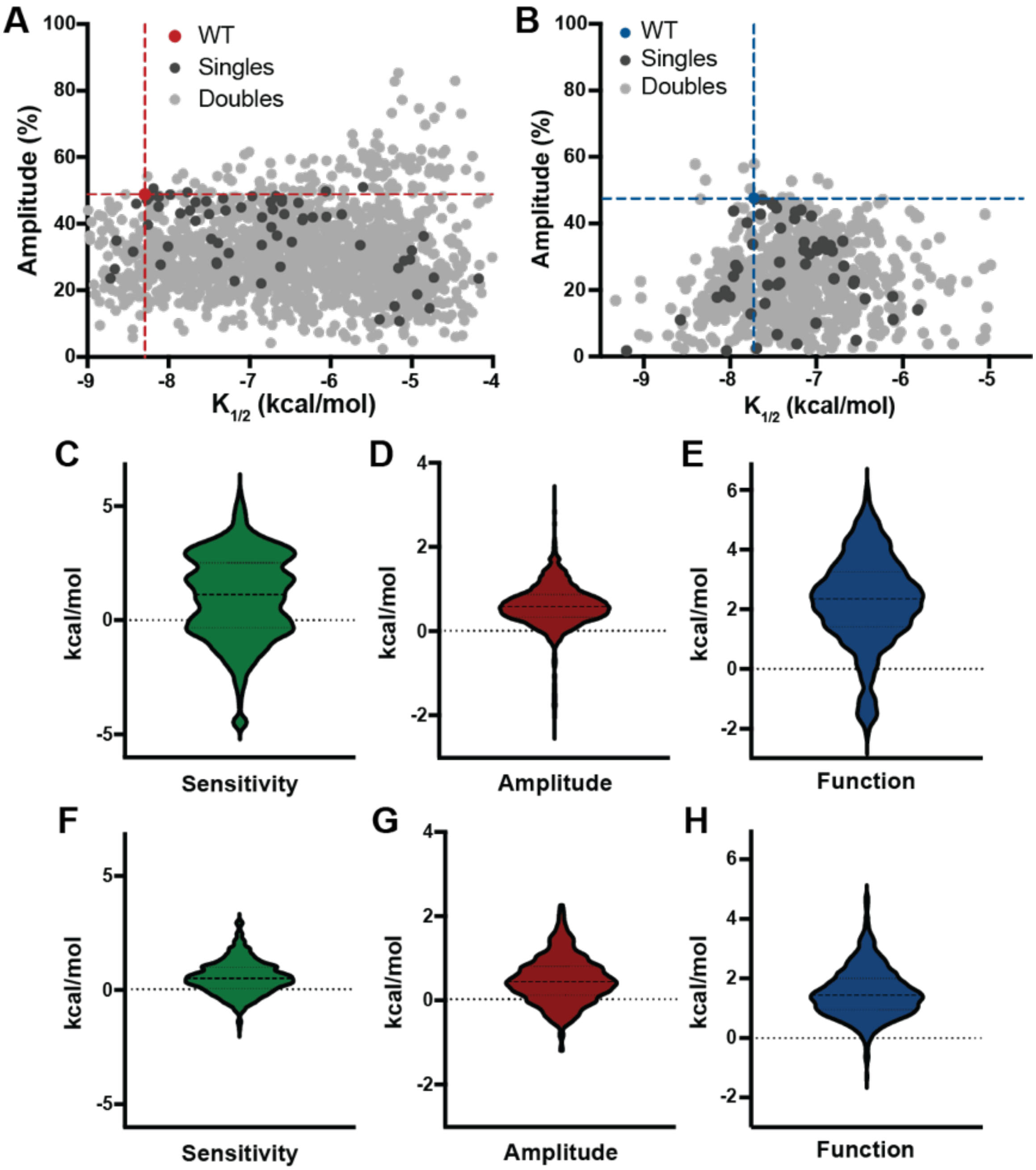

Figure 4.5: Sensitivity and amplitude distribution for *A. axanthum* CDN riboswitch mutants. A wide range of functional c-GG (A) and c-AG (B) riboswitches are identified with SMARTT. Double mutants are shown in dark gray, single mutants are black, and WT is red (c-GG) or blue (c-AG). The sensitivity ($K_{1/2}$) (C), amplitude (D), and function (E) are determined for each sequence in the SMARTT c-GG dataset. Additionally, sensitivity ($K_{1/2}$) (F), amplitude (G), and function (H) are determined for each sequence in the SMARTT c-AG dataset.

combines sensitivity to ligand ($K_{1/2}$) with dynamic range (amplitude) (Figure 4.5) [81]. Combining these parameters facilitated the efficient analysis of thousands of variants. A functional riboswitch shows a robust response to ligand at physiologically relevant concentrations. Low amplitude or weak sensitivity does not correspond to a functional riboswitch. The generated functional parameter integrates $K_{1/2}$ with amplitude in a single value to assess physiological relevance. To combine these metrics, $K_{1/2}$ was transformed into free energy, and amplitude was converted into a pseudo-energy. These energies were then compared to the WT values to generate $\Delta\Delta G_{sensitivity}$ and $\Delta\Delta G_{amplitude}$, which are similar in magnitude (figure 4.5) and thus weighted equally. The function parameter (f) combines these values and is therefore centered around a WT value of 0. Variants with increased sensitivity or amplitude have $f < 0$, and those with weaker sensitivity or lower amplitude have $f > 0$.

Among the functional sequences with c-GG, I observe $K_{1/2}$ values that span five orders of magnitude with amplitudes ranging from 5.5-99% (figure 4.5A). In contrast, while the c-AG responsive sequences access a wide distribution of sensitivities and amplitudes, the observed span is smaller. We observe $K_{1/2}$ values that span four orders of magnitude with amplitudes ranging from 5.5-60% (figure 4.5B). This showcases the high tunability of this riboswitch. Specifically, within two mutations, nearly any combination of sensitivity and amplitude can be achieved (figure 4.5A & B). Similar observations about tunability have been made using high-throughput methods on other riboswitch systems [81,82]. This may be a common feature of riboswitches. Responding to biologically relevant ligand concentrations with appropriate sensitivity is an essential component of riboswitch-driven gene regulation. Tuning sensitivity to a single organism's natural environment could explain the diversity of helical lengths and expression platforms seen in nature.

### 4.2.3 The c-GG functional landscape is robust

The difference between the observed and the expected function was calculated for each double mutant (equation 8) to determine whether a double mutation behaves similarly to the sum of each single mutation alone. Sequences where the effects are non-additive provide insight into nucleotides that may interact energetically. The epistasis values for each mutation were then plotted on a 2-dimensional heat map (figure 4.6B). For c-GG, the majority of mutations result in a riboswitch with some function. However, mutations to G20, A47, and nucleotides in P2 generally show lower function. This is seen in the 2-dimensional heatmap as a large region of non-functional sequences (figure 4.6B, outlined). This is consistent with previous studies showing the importance of both G20 and A47 [56].

In contrast to the robust functional landscape in the presence of c-GG, the functional landscape for c-AG is much more limited. Fewer sequences respond to c-AG, and those that do respond are weaker (figure 4.**6D**). This supports the natural riboswitch preference for c-GG observed in gel-based transcription termination. Similar to c-GG, mutations to P2 are poorly tolerated

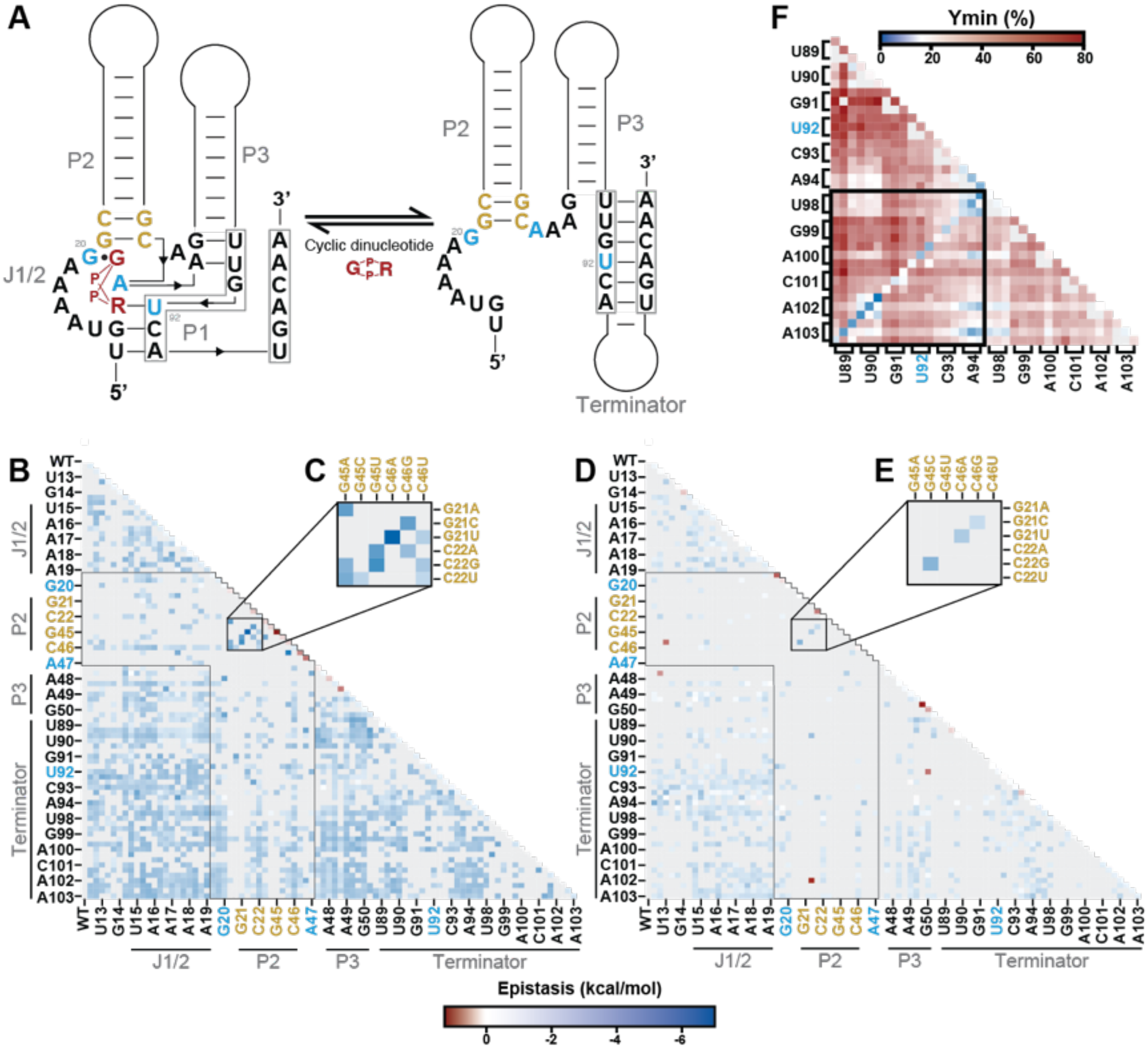

Figure 4.6: SMARTT epistasis and P2 covariation with c-GG. (A) The SMARTT mutant library focused on ligand contacting positions in P1, P2, P3, and the terminator. Nucleotides shown in blue directly interact with the ligand, and nucleotides shown in gold are in P2. (B) A heat map of the epistasis value was generated for all single and double mutants for c-GG responsive riboswitches. Gray corresponds to sequences that are non-functional. (C) Inset of P2 covariation. (D) A heat map of the epistasis value was generated for all single and double mutants for c-AG responsive riboswitches. (E) Inset of P2 covariation. (F) Heat map of the $Y_{min}$ for terminator mutants, base paired nucleotides are highlighted by the box.

by c-AG. However, c-AG is sensitive to GU wobble pairing in P2, while c-GG remains functional (figure 4.6C & E).

In both SMARTT datasets, the majority of mutations weaken the overall response of the riboswitch by either decreasing the amplitude or increasing the $K_{1/2}$ (figure 4.5A & B). This can be explained by evaluating the terminator stability for the mutant riboswitches. Manipulating the stability of the terminator helix shifts the apparent equilibrium between the ON and OFF states when the ligand concentration is constant. This shift is most noticeable in the $Y_{min}$ and $Y_{max}$ values. With the exception of the nucleotides at the loop end of the terminator, single and double mutations to the terminator dramatically shift the $Y_{min}$ higher, corresponding to an increase in full-length product in the absence of ligand (figure 4.6F). This correlates with a weaker terminator. Restoring the terminator stability by compensatory mutation returns the $Y_{min}$ to near wild-type levels. Mutations to A94 or U98 at the loop end of the terminator hairpin have a much smaller impact on the $Y_{min}$ and $Y_{max}$ values. This suggests that the loop end of the terminator can tolerate the removal of a single base pair without impacting terminator energetics. This is a common feature of terminator hairpins that has been observed in several systems [82,133,134].

While most sequences show functional switching in the presence of c-GG, any mutation to the terminator destabilizes the OFF state and thus reduces the overall amplitude of the riboswitch. This explains the general trend of weakened function observed with SMARTT, as the majority of sequences contain at least one mutation to the terminator helix.

### 4.2.4 Building a biochemical consensus for U92 c-GG and c-AG riboswitches

I used the single and double mutant data collected with SMARTT to generate a biochemical conservation diagram for both c-GG and c-AG responsive riboswitches. These consensus diagrams can be compared to phylogenetic conservation diagrams for this variant class of riboswitches [49,51]. Additionally, comparing these consensus diagrams to each other identifies sequences that impact specificity for either c-GG or c-AG. I determined the fraction of riboswitches containing a particular nucleotide identity that were functional and used this ratio to calculate the conservation of each nucleotide at each position (equations 9 & 10).

The biochemical dataset is limited by the number of positions mutated but generally shows high levels of conservation in the ligand-contacting positions and sequence flexibility throughout the rest of the riboswitch for both ligands (figure 4.5). Similarly, the previously reported phylogenetic conservation shows high levels of covariation in P1, P2, and P3 with minimal sequence conservation [49]. The key distinction between the biochemical and bioinformatic consensus diagrams for c-GG is the conservation of the closing base pair in P2 (figure 4.7A). High conservation of a GC base pair in this position is observed, which is absent in the bioinformatic consensus. SMARTT

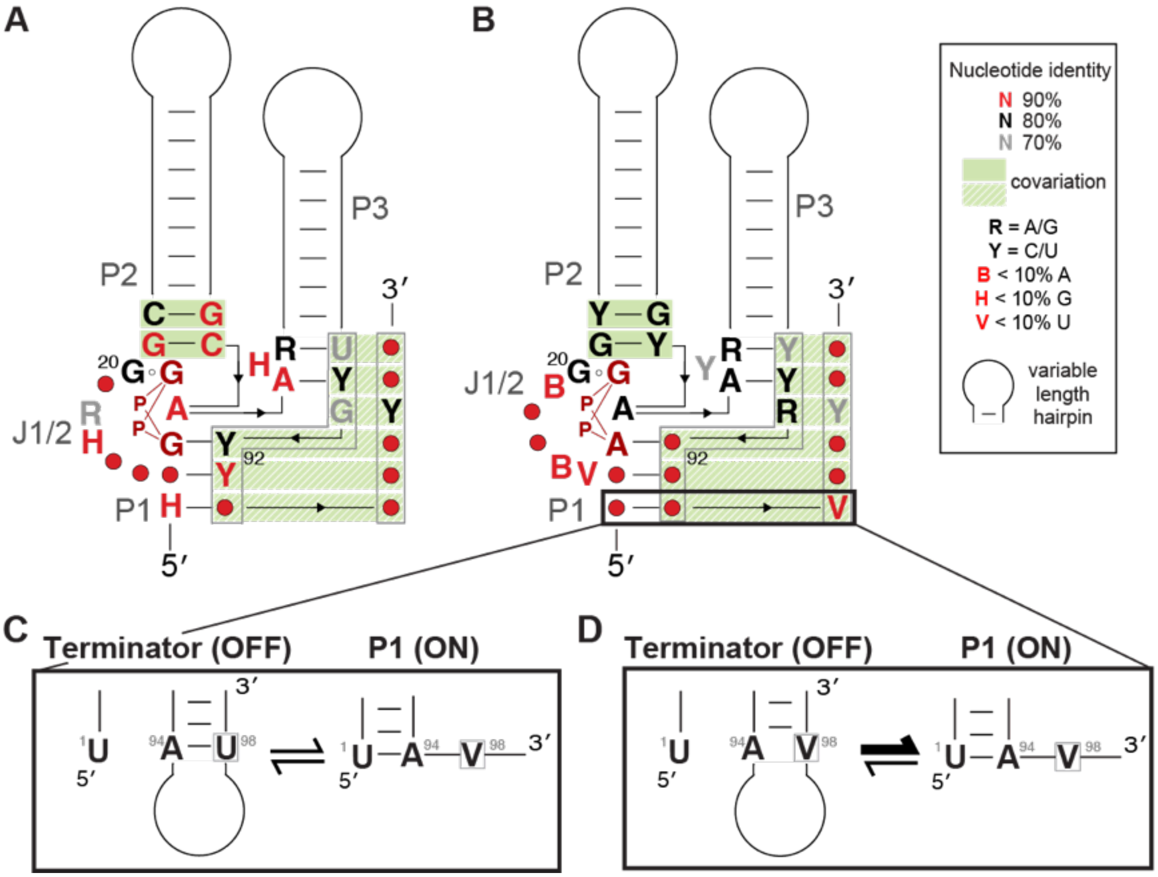

Figure 4.7: Biochemical consensus generated using SMARTT data. (A) Biochemical consensus of c-GG responsive riboswitches. (B) Biochemical consensus of c-AG responsive riboswitches. (C) A model of base pairing in the ON and OFF state with U98. (D) A model of base pairing in the ON and OFF state with V98.

also highlights the importance of covariation in the terminator, which the bioinformatic consensus excludes due to variability in the terminator between riboswitch constructs.

### 4.2.5 Peripheral nucleotides impact ligand selectivity

Unlike the c-GG-I riboswitch, which is very well characterized, there is no published consensus diagram for a G20 c-AG riboswitch. Due to the decreased affinity for c-AG, the functional landscape proved less robust than that for c-GG; however, the distinctions in consensus models for the two ligands are significant (figure 4.7B). Compared to the c-GG consensus, the c-AG consensus shows lower conservation at all three ligand-contacting nucleotides, G20, A47, and U92. This is due to the weaker constraints that were applied in screening the c-AG dataset based on the limited functional landscape with c-AG.

The most striking feature of the c-AG consensus sequence is the changes in sequence conservation within J1/2. J1/2 links P1 to P2 in the ligand-bound state of the riboswitch. These nucleotides are within 10Å of the ligand binding site, and some are involved in non-specific ligand interactions. However, the nucleotides at both ends of J1/2 are not predicted to interact with the ligand. Both the biochemical and bioinformatic c-GG consensus sequences show low conservation of the first two nucleotides in J1/2 with moderate conservation of adenosine or purine in the remaining nucleotides. In contrast, the c-AG consensus sequence has a strong preference against uridine in the first junction nucleotide and against adenine in the second and last junction positions. This preference is in direct conflict with the native riboswitch sequence at those selected J1/2 positions. All sequences that showed an improved response to c-AG contained at least one mutation in J1/2. From this, I conclude that junction positions are driving the observed specificity for c-GG despite U92, which has a marginal preference for base pairing with Aβ instead of Gβ [130–132]. These sequence preferences observed with SMARTT correspond to nucleotides of low conservation in the c-AG riboswitch consensus sequence. This highlights the importance of second-shell nucleotides that, due to their low biological conservation, were not previously considered for a role in ligand specificity.

Moreover, the sequence preferences of these junction nucleotides could be used to predict other promiscuous sequences. Previous research has determined that c-AG riboswitch promiscuity can be attributed to protonation of A20 in some sequences [100]. Notably, all promiscuous sequences previously tested conflict with the bioinformatic c-AG consensus sequence generated in this study at one or more junction positions [100]. This explains the decreased affinity for c-AG in comparison to c-GG in these riboswitches.

### 4.2.6 Energetic requirements for riboswitch switching

The other positions that have a strong preference for non-WT sequences are at the loop end of the terminator helix. Specifically, the nucleotide on the 3′ end of the loop shows a strong preference for anything except a uridine. Based on energetic calculations, this mutation is predicted to slightly weaken the terminator, reducing the minimum free energy by about 0.5 kcal/mol; however, analysis of the terminator helix shows that this mutation had a minimal impact on the $Y_{min}$ and $Y_{max}$. Instead, the major impact is expected to be on the bound conformation. Introducing this mutation is predicted to liberate A94 to base pair with U1 instead of U98 (figure 4.7C & D). This is expected to make it easier for P1 to form, which stabilizes the bound state. This may lessen the energy differential between the predicted bound and termination conformations to promote c-AG-driven function without inhibiting the termination state of the riboswitch. This same position can increase the sensitivity of c-GG responsive riboswitches as well; however, the impact on function is smaller due to the increased sensitivity of the parent construct for c-GG.

Other distal nucleotides in a number of riboswitch classes have previously been demonstrated to have an effect on riboswitch function independent of specificity [135–138]. For example, in the c-GG riboswitch, the P2 tetraloop and P3 tetraloop receptor are essential to riboswitch folding and function [56,122,136]. In these instances, distal nucleotides have a clear impact on riboswitch structure and impacts on riboswitch function are achieved by stabilizing or destabilizing different conformations. In contrast, the sequence preferences at the distal positions in this study do not have a clear impact on conformation stability. Specifically, crystal structures of both c-GG and c-AG riboswitches do not provide structural insight into how the J1/2 nucleotides are responsible for the observed ligand specificity [56,99,125]. Instead, it is possible that these junction nucleotides are involved in forming an intermediate structure that inhibits proper global folding. This is particularly relevant as transcriptional riboswitches are kinetically, rather than thermodynamically, regulated. However, these results suggest a ligand-dependent sequence preference in J1/2 that cannot be explained by riboswitch misfolding. Riboswitch studies have not yet identified similar sequence preferences in second-shell nucleotides that impact ligand specificity. Instead, research into enzyme-substrate specificity provides an alternate possible explanation for the role of J1/2 in ligand specificity.

Some protein enzymes use residues far away from the active site to manipulate substrate binding. Studies that investigate the mechanisms by which these distal residues can impact substrate binding have found that mutations to these positions impact the millisecond and microsecond motions of the protein as well as the global architecture [139–143]. Macromolecular motions are an essential component of substrate recognition as they enable the proper positioning of essential residues while minimally impacting the global protein fold.

Due to their flexible nature, protein loops play a key role in altering ligand-biding specificity. In addition to RNase A, the loops in trypsin, chymotrypsin, immune receptors, and angiogenin, among others, have been investigated and contain substrate-determining sequences [139–143]. Substitutions to loop regions can be made that sufficiently modulate the local environment to disrupt substrate binding or enzymatic efficiency. Similar, structurally independent, specificity-determining loops have not yet been identified in RNA. This work has shown that specific J1/2 sequences in a c-GG riboswitch decrease c-AG-driven function without any clear impact on riboswitch structure or c-GG-driven function. It is possible that this junction serves a similar role as protein loops regarding molecular motions and ligand specificity. For instance, a nucleotide substitution in the second-shell could shift the positioning of first-shell nucleotides that are in direct contact with the ligand, thereby altering ligand specificity.

## 4.3 Conclusion

The cyclic dinucleotide riboswitches are not the only riboswitch classes that differentiate between highly similar compounds. Other riboswitch variants use small sequence changes to selectively respond to chemically similar ligands. Additionally, all riboswitches must discriminate against structurally and chemically similar compounds found in cells. Research into RNA specificity mechanisms has focused heavily on direct ligand-contacting nucleotides identified using crystallography. However, crystal structures present a static snapshot of a dynamic molecule and narrowing the field of view to a handful of nucleotides in direct interaction with the ligand overlooks potential key elements that impact specificity. Investigating nucleotides in the shells surrounding the ligand provides insight into unpredicted functionally relevant sequences.

In the absence of direct contacts, the contributions of these distal nucleotides to ligand specificity are difficult to determine without in-depth mutational analyses. The SMARTT dataset generated here highlighted the impact that these remote contacts have on the specificity and function of the c-GG riboswitch. Applying similar high-throughput mutational analyses to other variant families could provide insights into nucleotides that impact riboswitch specificity outside the ligand binding pocket.

# 5. Cyclic Dinucleotide Riboswitch Evolution

This chapter is the beginning of an exciting story that takes what I learned from high-throughput sequencing of the CDN riboswitch and builds upon it to evolve new functions.

## 5.1 Background

Designing riboswitches against a target ligand has been a core area of research for synthetic gene regulation. Riboswitches are ideal tools for this because they are simple systems that result in reversible modulation of gene expression [67,68]. Unlike many other gene regulatory systems, riboswitches do not require protein factors and are thus easier to reliably insert into cells. A riboswitch upstream of a gene of interest can be designed to respond to a specific ligand at a specific concentration to modulate gene expression. Supplementation of the ligand in the growth media will thus induce a riboswitch-driven change in gene expression.

In addition to functioning as gene regulatory elements, riboswitches can serve as small molecule sensors. A riboswitch regulated by a ligand of interest upstream of a reporter can inform on the relative concentrations of that ligand in the cellular environment [69,71,144–148]. Recent development of fluorogenic aptamers has made it possible to use an entirely RNA-based reporter system rather than accounting for changes in protein expression, although proteins such as luciferase or β-galactosidase are also used as reporters.

The development of synthetic riboswitches relies heavily on *in vitro* selection for the development of RNA aptamers that can serve as the first domain of a riboswitch. SELEX, Systemic Evolution of Ligands by EXponential enrichment, has been used successfully to identify aptamers that bind tightly and specifically to a ligand of interest [72]. Classic methods enrich for aptamers that are not ideal riboswitches because they are not amenable to the structural rearrangement required for gene regulation. Instead, modifications have been made to the SELEX protocol to increase the probability of evolving a functional riboswitch aptamer. Capture SELEX relies on helical competition to select for RNAs that bind a molecule of interest preferentially over forming alternative secondary structure, as would be seen in a riboswitch expression platform [94,149]. While this method is more likely to evolve an aptamer for a functional riboswitch, the addition of a functional expression platform still poses some challenges.

The alternative to selecting a riboswitch aptamer *de novo* is to begin with a riboswitch scaffold and synthetically evolve an alternate function. While this is not the traditional approach to aptamer evolution, more studies are being conducted that build off a natural riboswitch scaffold to develop new aptamers and riboswitches [69,75]. These graftamers can then be returned to a native context, including the expression platform, to form synthetic riboswitches. Scaffolded aptamer evolution has also been used to create reporter systems where the evolved ligand is fluorogenic when bound to RNA.

Using a natural riboswitch parent construct increases the probability of finding an aptamer that meets the energetic requirements for ligand-driven helical switching. Graftamers have been used for the development of synthetic riboswitches for caffeine and quinine, as well as a fluorogenic reporter RNA [69,75,150].

All graftamer studies have used the purine riboswitch as a parent construct. The first purine riboswitches were identified in 2003 and responded to guanine [48]. Shortly thereafter, some purine riboswitches that responded to adenine were found [47]. This altered ligand specificity is a hallmark of variant riboswitches. Variants use the same structural scaffold with small sequence changes to recognize distinct ligands [87]. Following the initial discovery of variant riboswitches, many more variant classes have been identified, including other purine variants, the *ykkC* family, which consists of at least five variants, and the cyclic dinucleotide (CDN) family [14,32,49,51,59,60,106]. Due to the high evolvability of variant scaffolds – as demonstrated by the altered ligand specificity – they are ideal parent constructs for *in vitro* evolution.

The CDN family consists of two known variant classes, the cyclic-di-GMP and the cyclic-AMP-GMP riboswitches. This family was first identified as an orphan riboswitch class in 2007 [20]. Shortly thereafter, the ligand was found to be cyclic-di-GMP [51]. Crystal structures of a cyclic-di-GMP riboswitch identified two key nucleotides responsible for ligand specificity, G20 and C92 [56,118,122,151]. Specifically, G20 contacts the Hoogsteen face of G$\alpha$, and

C92 Watson-Crick pairs with Gβ. However, low conservation at these positions hinted at the existence of variant classes. This was confirmed with the discovery of the cyclic-AMP-GMP variant subclass [49,106]. This variant uses A20 to contact the Hoogsteen face of Aα [99]. While these two nucleotides in direct contact with the ligand tend to indicate specificity, there are many sequences that do not bind the predicted ligand [49,57]. Recent work has shown that in addition to these two nucleotides, sequences in regions adjacent to the ligand binding pocket also impact specificity (4.2.3). As a consequence, other variant classes that differ from the consensus in these adjacent positions might exist.

When the CDN riboswitches were initially discovered, there were two cyclic dinucleotides known to exist in nature: cyclic-di-AMP and cyclic-di-GMP. The cyclic-AMP-GMP variant confirmed the existence of the third cyclic dinucleotide. All three of these molecules are second messengers in bacteria. They transmit an extracellular signal into the cell and regulate expression or activity of a number of genes associated with that signal. Cyclic-di-GMP regulates the transition between sessile and motile life, while cyclic-di-AMP is more often associated with either heat or osmotic stress response [111,114–116]. Cyclic-AMP-GMP is less well studied, and an understanding of its roles and functions comes largely from analyzing the genes downstream of known cyclic-AMP-GMP riboswitches [49,106,152].

Shortly after the discovery of cyclic-AMP-GMP riboswitches in cells, pyrimidine-containing cyclic dinucleotides were discovered in nature [126].

Previously, only purine-containing CDNs were known. These pyrimidine-containing CDNs are synthesized by rare dinucleotide cyclases, cGAS/DncV-like nucleotidyltransferases (CD-NTases), found in many different phyla of bacteria [126]. The Kranzusch lab showed that all four possible uridine-containing CDNs could be synthesized, with the most prevalent products being cyclic-AMP-UMP and cyclic-di-UMP [126]. These CDNs are not present in high abundance, and their cellular concentrations remain unknown. Moreover, the function of these molecules has not been identified. Based on their similarity to purine-containing CDNs, it is likely that these molecules are also second messengers; however, the pathways that they regulate are not yet known.

Investigating the biological function of these molecules would typically involve identifying an effector in the same pathway. However, this has not yet yielded results. A possible alternative method is to assess biological function using pyrimidine-containing CDN-responsive riboswitches. The CDN riboswitch class recognizes similar ligands to these pyrimidine-containing CDNs. It is possible that undiscovered variants of this class exist that recognize these ligands. Alternatively, a synthetic riboswitch evolved from a CDN parent construct could be used to recognize these ligands in a reporter system to monitor changes in cellular concentrations *in vivo* [144,153]. Such a reporter system could be used to identify environmental conditions that promote or inhibit the synthesis of pyrimidine-containing CDNs and inform on their possible *in vivo* functions.

A riboswitch library based on a promiscuous cyclic-di-GMP parent construct was used to evolve a pyrimidine-containing CDN-responsive riboswitch. I have used a high-throughput sequencing-based functional assay to identify sequences that show a robust response to pyrimidine-containing CDNs. I found that mutations to both the ligand contacting positions and the adjacent nucleotides can change the ligand specificity.

## 5.2 Results and Discussion

### 5.2.1 A promiscuous CDN riboswitch is specific for purine-containing CDNs

I selected the previously reported promiscuous CDN riboswitch from *Acholeplasma axanthum* (*A. axanthum*) for this study (figure 5.1A & B). This riboswitch has U92 in the β ligand binding position and is promiscuous for both c-AG and c-GG. I hypothesized that, due to the increased ligand flexibility, this switch may be easier to evolve to new functions. Crystal structures of the CDN riboswitches suggest that they are not expected to respond to any pyrimidine-containing CDNs based on binding pocket geometry. To confirm that the *A. axanthum* riboswitch responds to only purine-containing CDNs, I performed transcription termination in the presence of all ten possible CDNs.

The purine-containing CDNs, c-GG, c-AG, and c-AA, were synthesized enzymatically using a diguanylate cyclase (tDGC) for c-GG and an AMP-GMP

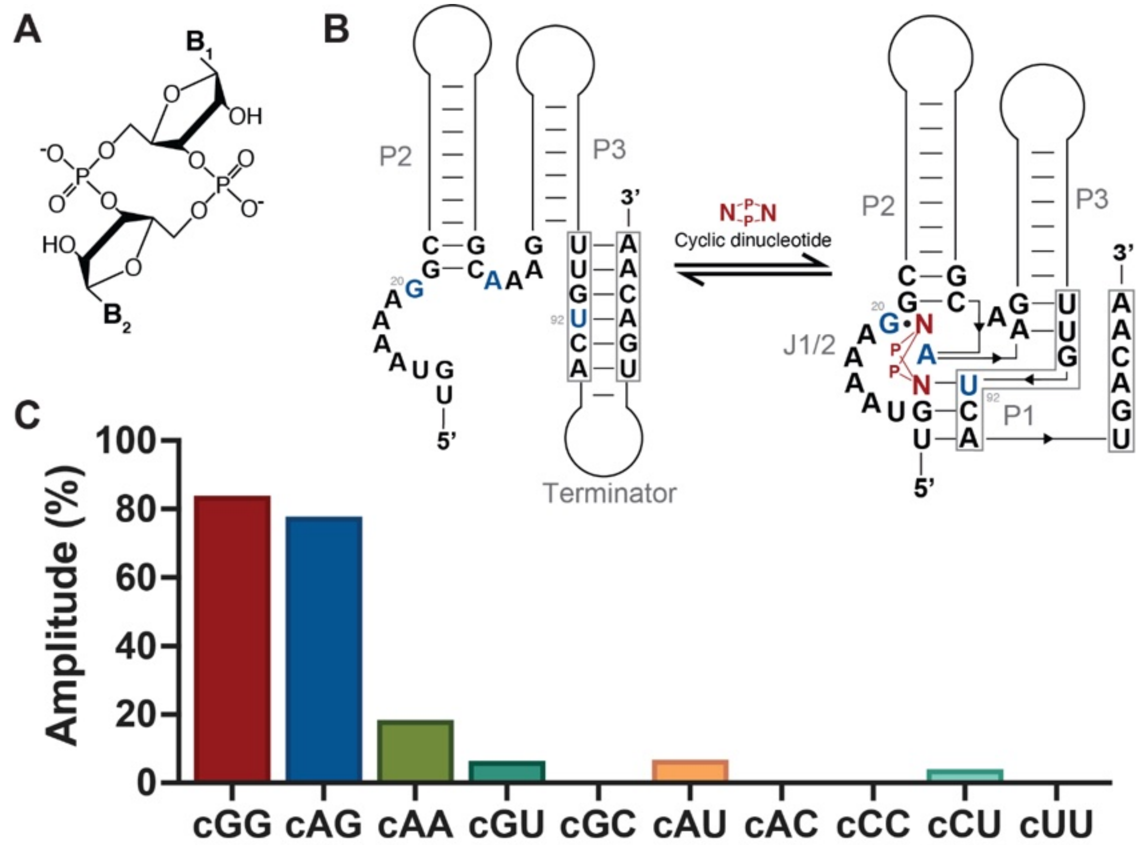

Figure 5.1: Promiscuous CDN riboswitch responds to purine-containing CDNs. (A) CDN where B1 and B2 are any nucleotide. (B) Schematic of A. axanthum riboswitch in the OFF and ON states. (C) The transcription termination amplitude of a CDN riboswitch in the presence of each CDN.

cyclase (DncV) for c-AG and c-AA. The pyrimidine-containing CDNs were chemically synthesized following previously reported protocols for one-pot synthesis. All CDNs were verified using mass spectrometry.

The *A. axanthum* CDN variant riboswitch was previously reported to respond to c-AG and c-GG with approximately equal affinities and a robust amplitude. Transcription termination in the presence of c-AG and c-GG confirms this result with an amplitude of 80% (figure 5.1C). Transcription with c-AA is also capable of modulating transcription length, although with much weaker sensitivity and an amplitude of only 20%. In contrast to the response seen with purine-containing CDNs, all pyrimidine-containing CDNs are poor modulators of transcript length. The C-containing and c-UU CDNs are unable to generate any function, while the hybrid U-containing CDNs show very low function. Structurally, it is possible that c-GU or c-AU are recognized with the purine in the $\beta$ binding pocket base paired with U92 and U in the $\alpha$ binding pocket. Specifically, the hoogsteen face of U can be contacted with a single hydrogen bond from the Watson-Crick face of G20. This could explain the small amount of modulation that is seen with c-GU and c-AU, although the sensitivity is significantly weakened commensurate with the weaker interaction between the ligand and the riboswitch.

### 5.2.2 Functional analysis of a mutated aptamer produces potential variants

I used the mutant library synthesized previously to begin evolving a new ligand specificity for the *A. axanthum* variant riboswitch. I used a Sequencing-based Mutational Analysis of RNA Transcription Termination, or SMARTT, to assess the functionality of thousands of mutants in the presence of pyrimidine-containing CDNs.

The pyrimidine-containing CDNs were pooled in two groups, C-containing and U-containing, for a preliminary SMARTT run. This increased the number of reads per ligand concentration tested and allowed for observation of select triple mutants. To maximize the probability of a positive hit, both pools went through a single round of selection. After transcription termination at the high ligand concentration, full-length reads were selected for using a reverse transcription primer targeted to the region downstream of the termination site. This selects transcripts that were turned ON in the presence of ligand, which includes sequences that were activated by ligand as well as sequences that are broken in the ON state. Following selection, standard protocols were followed for SMARTT with three tested concentrations: no ligand, high C-containing pooled ligand, and high U-containing pooled ligand. This data is sufficient to calculate an amplitude for each sequence with each ligand pool (figure 5.2A & D).

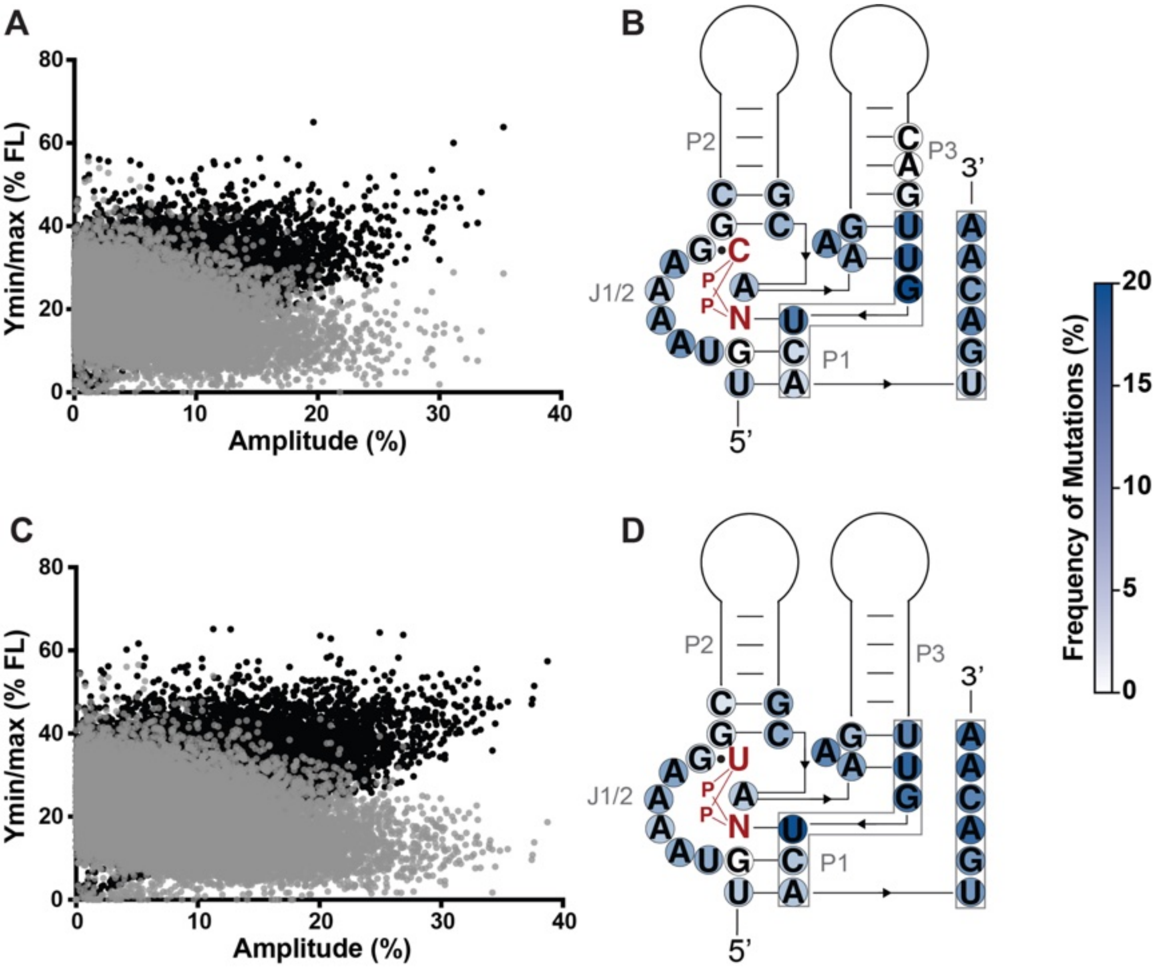

Figure 5.2: SMARTT amplitude and mutation frequency. (A) The $Y_{min}$ (gray) and $Y_{max}$ (black) for each sequence plotted against the calculated amplitude for C-containing CDNs. (B) Mutation frequencies for sequences with an amplitude greater than 15% for C-containing CDNs (ligand orientation unknown). (C) The $Y_{min}$ and $Y_{max}$ for each sequence plotted against the calculated amplitude for U-containing CDNs. (D) Mutation frequencies for sequences with an amplitude greater than 15% for U-containing CDNs (ligand orientation unknown).

The majority of sequences measured appear to be unresponsive, with an amplitude of less than 10%. However, both C-containing and U-containing CDN pools produced sequences with amplitudes greater than 20% as well. Sequences with an amplitude greater than 15% were compiled, and the positions with the highest frequency of mutation were identified for each pool (figure 5.2B & D). The most frequently mutated positions for C-containing CDNs are at the 3′-end of P3. These nucleotides are also part of the terminator helix in the alternate conformation, and mutating them destabilizes both the bound and unbound conformations. It is likely, therefore, that these mutations are not ligand specific. The next highest frequencies are found at U92. This is expected to have a strong impact on ligand specificity because it is in direct contact with the ligand. In contrast to C-containing CDNs, the U-containing CDNs have a higher frequency of mutation in the 3′-end of the terminator. This decreases the stability of the OFF state without impacting the ON state. This suggests that the CDN scaffold is capable of binding to U-containing CDNs, but the binding energy is insufficient to stabilize the ON state in the absence of destabilizing mutations. Finally, the mutation frequency in J1/2 is also increased for both C-containing and U-containing CDNs. This agrees with the previously reported finding that nucleotides in J1/2, despite no clear contact with the ligand, can impact ligand specificity.

### 5.2.3 Candidates identified through SMARTT modulate transcript length in a gel-based assay

Three candidates from each pooled ligand experiment were selected for investigation using a gel-based assay. The selected candidates have amplitudes over 30% in the SMARTT assay and mutations distributed throughout the riboswitch. Additionally, sequences with three mutations were only selected if each ligand concentration contained over 100 reads. Preliminary transcription termination results show that two constructs from the selected sequences have a reasonable ligand-driven response to a pyrimidine-containing CDN.

Mutant 1 (M1, figure 5.3A) was selected for investigation from the C-containing CDN sequences. In the SMARTT assay, M1 has an amplitude of 30% with the pooled C-containing ligands and <5% with the pooled U-containing ligands. This suggests that it is specific for a CDN containing C but not U. Using a gel-based assay, M1 has an amplitude of 24% and is specific for cAC over all other C-containing CDNs (figure 5.3B). This construct has moderate sensitivity for cAC with a $K_{1/2}$ of 320 μM. M1 appears to be the most promising candidate functionally because the $Y_{min}$ is approximately 15%, indicating a strong terminator. Surprisingly, none of the mutations are in positions that directly interact with the ligand. G50 is in the P3 region and is predicted to form a wobble base pair with U89. Mutating G50C disrupts this base pair, however, A49, which is predicted to be bulged out, can replace G50 and Watson-Crick

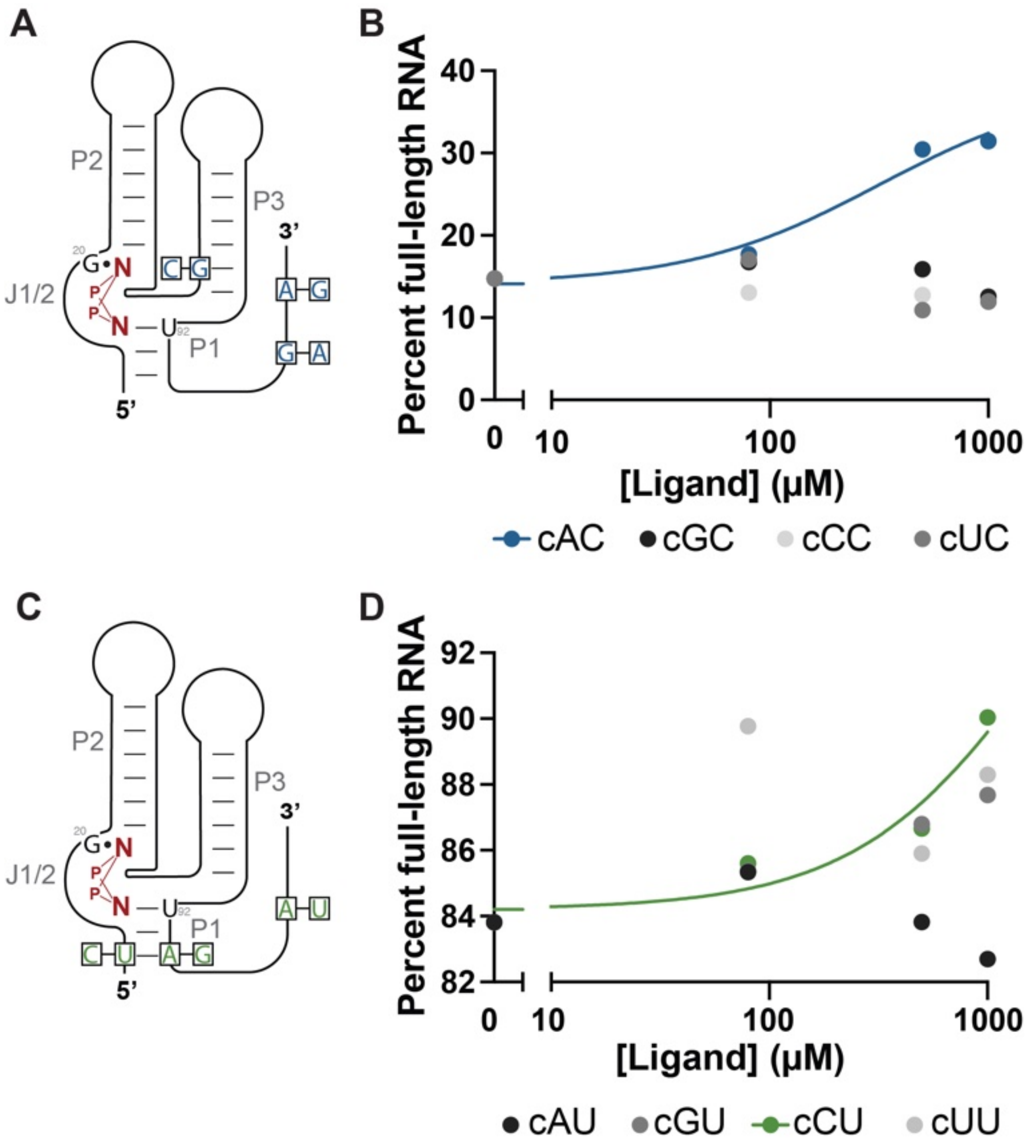

Figure 5.3 Gel-based transcription termination of pyrimidine-containing CDN candidates. (A) Secondary structure of a CDN riboswitch with M1 marked in blue. (B) Transcription termination curve for M1 with all C-containing CDNs. (C) Secondary structure of the a CDN riboswitch with M2 marked in green. (D) Transcription termination curve for M2 with all U-containing CDNs.

pair with U89. This maintains the stability of P3, which is essential to formation of the ON state. The remaining two mutations are in the 3′-end of the terminator helix. The first, G101A, is at the loop end of the terminator, and previous results have shown that mutations to this region are tolerated by the CDN riboswitch (CDN paper). The second, A104G, replaces an AU base pair with a GU wobble. This is predicted to reduce the energy of the terminator helix, which could explain the observation that cAC can modulate transcript length despite minimal rearrangement of the ligand binding pocket. However, ligand specificity cannot be explained structurally or energetically for this construct.

Mutant 2 (M2, figure 5.3C) was selected from the U-containing CDN pool, where it has an amplitude of 30%. M2 was also seen in the functional candidates from the C-containing pool, where it has an amplitude of 15%. Based on this, it was expected to respond to cCU. Using a gel-based transcription termination assay, M2 was found to have an amplitude of approximately 15%, although the $Y_{min}$ is significantly higher than expected at 84% (figure 5.3D). Additionally, M2 was not as selective for a single ligand; it showed a response to cGU as well as cCU. M2 stabilizes P1 by converting a native AU base pair into a GC base pair. This stabilizes the ON conformation, which explains the increased $Y_{min}$ in this construct. Additionally, M2 has a mutation in the terminator, A102U. This mutation disrupts the base pair that would form between U92 and the terminator in the OFF state. This is predicted to destabilize the terminator while simultaneously leaving U92 free to pair with the ligand. As such, it is unsurprising that the response seen in M2 is not ligand specific.

M1 and M2 both require further optimization for ligand specific gene regulation. M1 is a strong candidate for a cAC riboswitch, and building a mutant library from this construct will explore the functional space for cAC and could increase the riboswitch sensitivity. M2 is not yet specific for a single ligand, but minor sequence changes could be sufficient to generate specificity. Unlike the original parent construct, M2 is showing a response to pyrimidine-containing CDNs and it is, thus, a reasonable parent construct for a second round of SMARTT.

## 5.3 Conclusions

I have confirmed that SMARTT can be used to evolve new ligand specificity for a CDN variant riboswitch. Unlike current methods, which evolve an aptamer and expression platform separately, this method identifies a functional riboswitch in a single assay. However, this method has not been optimized for large mutation libraries and thus cannot access the sequence diversity that is found in the development of graftamers or capture-SELEX yet. Sequence space is limited by the number of reads per sequence that are required for data analysis. This is compounded with the absence of a counter-selection step to remove sequences that are broken ON. The addition of a counter-selection step would increase the probability of sequencing functional mutants in the future.

Previous high-throughput functional analyses of riboswitches have found that both sensitivity and amplitude are highly tunable. SMARTT can be used after the initial riboswitch sequence selection to fine-tune the response to match desired levels. In this case, careful library design, built around the selected mutants in this study, could yield functional riboswitches that are highly specific for a new CDN ligand of choice.

Riboswitch tuning can also be done in sequence with graftamer evolution to optimize the appended expression platform for the newly evolved aptamer. Functional riboswitches for quinine and caffeine have been developed by the Breaker lab using a graftamer approach. However, the quinine riboswitch had low activity compared to the parent guanine riboswitch. To address this, the Breaker lab made a single mutation that was predicted to increase the riboswitch activity. In this case, using bioinformatic data was sufficient to generate a responsive riboswitch. However, SMARTT can be used to identify functional sequences with a range of sensitivities and magnitudes simultaneously and thus tune any evolved aptamer to a desired ligand concentration. Using a joint graftamer-SMARTT approach, a riboswitch tuned to specific concentrations for any ligand can be developed.

The evolution of new ligand binding capabilities within two mutations of a native CDN riboswitch exhibits the evolvability of the CDN variant scaffold. It is possible, given the proximity of these variants in sequence space, that natural variants have already evolved to recognize the pyrimidine-containing CDNs. To search for potential natural variants, it is necessary to understand the requirements for pyrimidine-containing CDN-driven function. This can be done with the development of a biochemical consensus sequence generated through SMARTT, as reported previously (Focht 2023).

Other similar variants, including the purine riboswitches, have previously been synthetically evolved to recognize new ligands. Additionally, research into the *ykkC* scaffold, which contains at least five known variants, has highlighted the evolvability of this structure. It seems reasonable that variant riboswitches will be the basis for more *in vitro* evolution in the future as they have proven to be highly evolvable functional RNAs that do not require the same functional optimization as *de novo* aptamers.

# 6. Conclusions

The RNA world theory suggests that the earliest organisms on earth evolved to function using only RNA rather than the complex system of DNA, RNA, and proteins employed by organisms today [88,154]. There is strong evidence that suggests that RNA is capable of self-replication, which is a necessary step in furthering the RNA world theory as it provides a pathway by which genetic material can be passed down using only RNA [155,156]. Another important feature of RNA that supports this theory is the ability to recognize and respond to the environment. Riboswitches exemplify this feature. The RNA world theory predicts that the earliest building blocks were nucleic acids rather than the more complex amino acids. The disproportionate number of riboswitches that recognize nucleic acid-derived compounds, including cofactors and signaling molecules, provides further support for the RNA world theory [12,29].

Notably, riboswitch function is based on both sequence specificity and tertiary folds. Frequently, mutations to an aptamer scaffold are tolerated as long as the 3-dimensional structure remains intact. As such, riboswitches can be mutated while maintaining function and can thus access a larger sequence space. A consequence of this sequence flexibility is the development of new functions with the introduction of a few mutations, in other words, the development of variant riboswitches.

In this work, I have used a high-throughput sequencing-based approach to investigate variant riboswitches. I have highlighted the sequence differences and similarities between variant classes, especially in regions that are distal from the ligand binding pocket. Using high-throughput techniques, I have identified or evolved new variant sequences for the *ykkC* and CDN riboswitches. This speaks to the versatility and adaptability of variant riboswitch scaffolds, a requirement of the RNA world theory.

Ligand specificity is an essential component of a functional riboswitch. Previous high-throughput work has shown that expression platform energetics can be involved in ligand recognition and that the overstabilization of alternate structures can occlude ligand-driven modulation [82]. I show that in variant riboswitches, the energetic landscape of the expression platform contributes to ligand specificity. Moreover, nucleotides outside the ligand binding pocket appear to influence ligand identity. As the field of variant riboswitches develops further, investigating the entire functional region can help differentiate between related classes.

# 7. Methods

## 7.1 Materials and Methods: *ykkC*

### 7.1.1 Design of plasmids and template DNA for gel-based in vitro transcription termination

The *P. micra* (NZ_AXUQ01000034.1/1861-1967) and *T. oceani* (NC_014377.1/ 866146-866258) riboswitches were each cloned into a pUC19 plasmid. Both riboswitches were PCR amplified using Phusion HF polymerase (NEB or Thermo Fisher Scientific) with primers that added promoters upstream as well as native downstream sequences [157]. Mutations were inserted via PCR of the aptamer using Phusion HF polymerase and primers with nucleotide changes. The PCR products were then gel purified (Thermo Fisher Scientific gel extraction), and promoters and downstream sequences were added with PCR using Phusion HF polymerase.

### 7.1.2 Preparation of mutagenized DNA library

A mutagenized *T. oceani* library was prepared by chemical oligonucleotide synthesis with mixed base incorporation (Keck oligonucleotide synthesis facility, Yale). Mutations were restricted to the P0 and P3 stem loops as well as the terminator stem, as these regions show the highest occurrence of mutation between ppGpp and PRPP consensus sequences. The library was designed to maximize single and double mutations using equation 1, where P is the probability of n mutations occurring in a construct of length L with a mutation rate of f [158].

$$P = \ {}_{L}C_{n} * f^{n} * (1-f)^{n} \qquad (1)$$

Probabilities were then divided by the number of possible sequences with n mutations (equation 2) to find the frequency of each sequence occurring ($\nu_{seq}$) in the library (equation 3).

$$S = \ 3^{n} * \ {}_{L}C_{n} \qquad (2)$$

$$\nu_{seq} = \frac{P}{S} \qquad (3)$$

A mutation rate of 3% for each non-wild type base was used over 43 positions in P0, P3, and the terminator, which resulted in 1.5% wild type, 7% single mutants, 15% double mutants, and 76.5% with 3 or more mutations. With 43 positions, there are 8127 possible double mutants. With this mutation rate, approximately 200 reads per condition are expected for each double mutant. Due to length restrictions in mixed base oligonucleotide synthesis, 2 primers were ordered and then combined using an overlap extension protocol with SSII reverse transcriptase (Invitrogen). Following overlap extension, a promoter upstream and 80 nucleotides downstream were added to the riboswitch with PCR using Phusion HF polymerase. This final product was used as template DNA in the SMARTT transcription termination assay.

### 7.1.3 Gel-based in vitro transcription termination assays

20μL transcription reactions were prepared on ice as reported in Torgerson et al. 2018 [82]. Reactions contained 40 mM Tris-HCl (pB 7.5), 150 mM KCl, 10μg/mL BAS, 1% glycerol, 15 mM $MgCl_2$, 1 mM DTT, 50 μM each ATP, CTP, UTP, and GTP, 30 nM DNA template, approximately 2.5 μCi [α-$^{32}$P] GTP, 0.5 units *E. coli* RNA Polymerase Holoenzyme (NEB), and varying concentrations of ligand (PRPP or ppGpp). Reactions were incubated at 37°C for 1 hour and then transferred to ice for 2 minutes. 2 volumes of loading buffer (25 mM EDTA, <0.1% Xylene Cyanol, <0.1% Bromophenol Blue, ~93% formamide) were added to each reaction. Full-length and truncated RNA transcripts were separated on a 10% denaturing polyacrylamide gel (7 M Urea) and imaged with a Typhoon FLA (GE Healthcare). Individual bands were quantified with ImageQuant (GE Healthcare) and normalized based on the guanosine content of each sequence. Data from a single ligand titration were fit to equation 4:

$$\%\ Full\text{-}Length = (Y_{max} - Y_{min})\left(\frac{X}{K_{1/2}+X}\right) + Y_{min} \tag{4}$$

where X is ligand concentration, $Y_{min}$ is the percentage of full-length RNA with no ligand, $Y_{max}$ is the percentage of full-length RNA at saturating ligand concentrations, and $K_{1/2}$ is the concentration at half-maximal termination. Ligand titrations were performed in duplicate.

### 7.1.4 Preparation of RNA for high-throughput sequencing

A set of 100 μL transcription termination reactions was prepared on ice following Torgerson et al. 2018 [82]. Samples contained the same reagents as the gel-based assay described above except for radiolabeled GTP, which was not included. Each ligand concentration was prepared separately, incubated at

37°C for 1 hour, and then placed on ice for 2 minutes. Template DNA was removed with DNase treatment. The full transcription sample was combined with 2.2 µL water, 2.5 µL of 10x TURBO DNase buffer (Thermo Fisher Scientific), and 0.3 µL TURBO DNase (Thermo Fisher Scientific) and incubated at 37°C for 20 minutes. The RNA was then diluted 10-fold and purified using an Oligo Clean & Concentrator Kit (Zymo).

A preadenylated DNA adapter was ligated onto the 3′ end of the RNA (/5rApp/NNNNNCTGTAGGCACCATCAAT/3ddC/ ordered from IDT) using T4 RNA Ligase 2, KQ (NEB). The final reaction mixture contained 1x T4 RNA Ligase reaction buffer, 25% PEG8000, 100 µM DNA adapter, 10 U/µL T4 RNA Ligase 2, KQ (NEB), and ~50% of the purified RNA. The reaction was incubated at 12°C overnight and then purified using solid phase reversible immobilization (SPRI) magnetic beads (Backman Coulter).

The ligated adapter provides a primer annealing site for reverse-transcription into cDNA. 36.2 nM of primer (5′GATTGATGGTGCCTACAG) was annealed to the RNA in the presence of 725 µM dNTPs for 5 minutes at 65°C and then cooled to 4°C for 1 minute before the remaining reagents were added. The final reaction mixture containing 25 nM primer, 500 µM dNTPs, 1x SSIV buffer, 5 mM DTT, 1 U/µL RNaseOUT (Invitrogen), and 10 U/µL Superscript IV (Thermo Fisher Scientific) was incubated at 55°C for 10 minutes and then transferred to 80°C for 10 minutes to inactivate the Superscript IV enzyme. The RNA was then degraded by incubation with RNase A (0.5 µg/µL; Thermo Fisher

Scientific) and RNase H (0.05 U/µL; Invitrogen) at 37°C for 30 minutes. The resulting cDNA was purified using SPRI magnetic beads (Beckman Coulter).

Illumina sequencing adapters, spacers to increase library diversity, and a unique index were added to the cDNA from each sample condition with a single round of PCR using Phusion HF Polymerase (NEB or Thermo Fisher Scientific) and primers with large 5′ overhangs (supplemental). 8 cycles of PCR were used to amplify the cDNA. Final DNA concentration was obtained using a Qubit high sensitivity dsDNA detection (Thermo Fisher Scientific). Samples were pooled with equal final concentrations before sequencing at the Yale Center for Genome Analysis (YCGA). Sequencing was performed on an Illumina HiSeq 4000 (2 x 150). The 30 samples produced by the two mutant libraries and two ligands were sequenced with 30% of a lane (~1% of a lane per sample).

### 7.1.5 Analysis of sequencing results

Analysis proceeded similarly to Torgerson et al. 2018 [82]. The region upstream of the aptamer and the 3′ adapter were removed using CutAdapt [159]. A minimum adapter overlap length of 10 bases and a minimum final sequence length of 80 nucleotides was applied. Sequences missing adapters were discarded. Bowtie2 was used to align sequences to a full-length WT template [160]. Dovetailing was permitted, but discordant sequences were removed.

Custom Python scripts previously reported were used to determine the fraction of full-length and truncated reads for all variants with 0-2 mutations [82]. Sequences were classified as terminated if the last nucleotide sequenced was between 130-200 (inclusive), which is consistent with termination in the poly-uridine tract. The minimum full-length sequence was 200 nucleotides, consistent with readthrough transcription of the full construct. The number of full-length and truncated reads were then counted for each variant to determine percent full-length RNA at each ligand concentration. The data were fit to equation 4 and curves were plotted and visualized in Prism 9.

Once equations of fit were established (equation 4), $K_{1/2}$ was converted to a free energy using equation 5:

$$\Delta G_{sensitivity} = -RTln(K_{1/2}) \tag{5}$$

where R is the ideal gas constant (0.001987 kcal $K^{-1}$ $mol^{-1}$), T is the temperature of the transcription termination reaction (310 K), and $K_{1/2}$ is the measured value. The $\Delta\Delta G_{sensitivity}$ was found by subtracting the WT $\Delta G_{sensitivity}$ from the mutant value. Amplitude was transformed into a pseudoenergy using equation 6:

$$\Delta G_{switching} = -RTln(\frac{Amp}{Amp_{max}-Amp}) \tag{6}$$

where Amp is the measured value and $Amp_{max}$ is the maximum amplitude measured in the sequencing sample). Similarly, the $\Delta\Delta G_{switching}$ was found by subtracting the WT $\Delta G_{switching}$ from the mutant value.

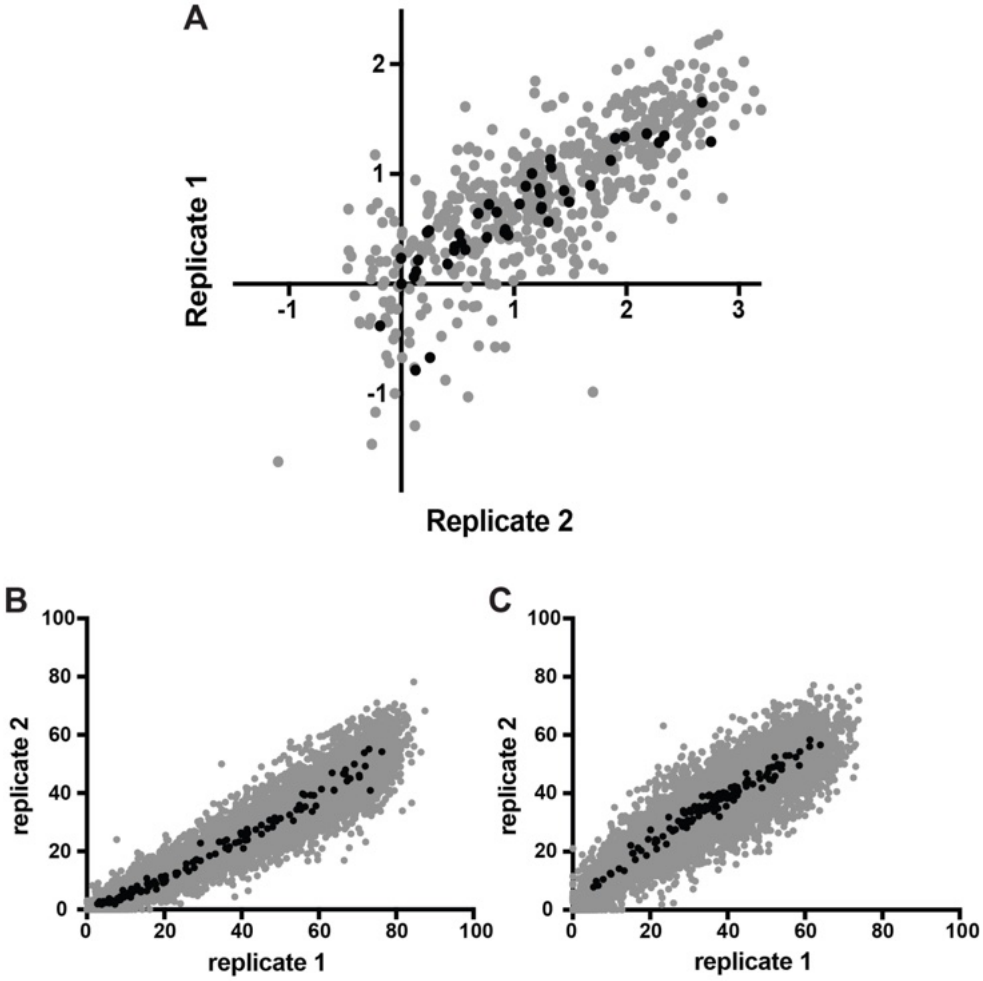

Figure 7.1: Replicate reproducibility of SMARTT. (A) Functional value for replicates with fits over 5% amplitude and measurable sensitivity. (B) Percent full-length for each mutant with 0 μM ligand. (C) Percent full-length for each mutant with 250 μM ligand.

### 7.1.6 Calculation of Function

Data with 0.5 μM < $K_{1/2}$ or $K_{1/2}$ > 500 μM or with amplitude < 10% were considered too noisy to fit and discarded (figure 7.1). All other data were analyzed using a functional parameter described in equation 7:

$$Func\ (f) = \Delta\Delta G_{sensitivity,Mut} + \Delta\Delta G_{switching,Mut} \tag{7}$$

where $\Delta\Delta G_{sensitivity}$ and $\Delta\Delta G_{switching}$ can be found using equations 5 and 6. With this parameter, any f < 0 describes a sequence that is more functional than WT with either a tighter sensitivity or larger amplitude while f > 0 represents a less functional riboswitch.

Further calculations were completed to describe the relationship between mutations in the functional landscape using equation 8:

$$Epistasis = f_{AB} - (\bar{f_A} + \bar{f_B}) \tag{8}$$

where $\bar{f_A}$ and $\bar{f_B}$ are the average function for all sequences that contain mutation A or mutation B respectively and $f_{AB}$ is the function of the double mutant that contains both mutation A and mutation B. This epistasis value was then visualized on a heat map using Prism 9.

### 7.1.7 Bioinformatic energetic analysis

Sequences for ppGpp and PRPP riboswitches were obtained from published Stockholm files [14,59]. The sequences were manually assessed to identify potential rho independent transcription termination sites. These sites contained at least seven uridines preceded by an RNA hairpin. Hairpin energies were calculated using RNAfold [161]. Poly-uridine lengths were counted from the base of the hairpin to the first set of two consecutive non-uridine nucleotides. Statistical significance was evaluated using an unpaired t-test. The data were visualized using Prism 9.

## 7.2 Materials and Methods: *ykkC* variant

### 7.2.1 Preparation of mutagenized DNA library

A mutagenized *T. oceani* library was prepared from a mutated A93G parent as previously indicated (7.1.2).

### 7.2.2 Preparation of RNA for high-throughput sequencing

The mutagenized *T. oceani* library was prepared for sequencing as previously indicated (7.1.4).

### 7.2.2 Analysis of sequencing results

Analysis proceeded as previously described (7.1.5).

### 7.2.3 Calculation of Function

The function parameter was calculated as previously described (7.1.6).

### 7.2.4 Phylogenetic analysis

A phylogenetic tree was generated using the published RNA sequences from the ykkC variants [14,32,59,60]. Fasta files were aligned using CMfinder [162] and viewed using Ralee [163]. Nucleotide positions for which greater than 90% of the sequences contained gaps were masked before relatedness was assessed to reduce the weight attributed to highly variable regions. An approximately-maximum-likelihood phylogenetic tree was generated using

FastTree [164]. The tree was visualized using the interactive tree of life (iTOL) online interface and annotated using a custom template.

### 7.2.5 Microscale Thermophoresis Binding assay

Due to length restrictions in oligonucleotide synthesis, riboswitch constructs for binding were generated via overlap extension with SSII reverse transcriptase (Invitrogen). Following overlap extension, the T7 promoter was added upstream and native sequence was added downstream of the riboswitch with PCR using Phusion HF polymerase. The resulting DNA was transcribed using T7 polymerase and purified on a 10% denaturing acrylamide gel. The purified RNA was labeled at the 3′ end with fluorescein 5-thiosemicarbazide as described previously [165]. MST was completed using a NanoTemper Monolith NT.115 and untreated capillaries. A stock of 200nM RNA was prepared in assay buffer [50 mM HEPES-KOH (pH 7.5), 200 mM KCl, and 20 mM $MgCl_2$] with 0.1% Tween 20. This was diluted 1:1 with ligand stocks generated with serial dilution. The normalized fluorescence at 5 s was used as the readout for ligand binding. Data were fit using Prism 9.

### 7.2.6 Riboswitch Reporter assay

Liquid-based reporter assays were conducted using BW25113 cells. The full-length *A. inagensis* riboswitch, including the first six codons, was cloned upstream of a β-galactosidase LacZ gene in the Simons lab pRS414 plasmid [166]. This plasmid was then transformed into BW25113 cells which were grown on ampicillin plates to select for colonies that contained the plasmid.

Liquid cultures were grown overnight in minimal media (M9) supplemented with 50 ug/mL ampicillin. XMP or GMP was added to the overnight culture where appropriate. Following overnight growth, 1 mL of the culture was resuspended in a permeabilization buffer [100 mM $Na_2HPO_4$, 20 mM KCl, 20 nM $MgSO_4$, 0.003% SDS, 0.8 mg/mL CTAB, 0.4 mg/mL Sodium deoxycholate, and 5.4 μL/mL BME]. Cells were left to permeabilize for half an hour at 37°C then 0.5 volumes of substrate solution [60 mM $Na_2HPO_4$, 40 mM $NaH_2PO_4$, 4 mg/mL ONPG, 2.7 μL/mL BME] were added. Reactions were left to incubate at room temperature for 6 hr before stop buffer [1 M $Na_2CO_3$] was added and A420 was measured. Absorbance was normalized to the $OD_{600}$ for each sample.

Knockout strains were acquired from the Keio knockout library [167]. The same Miller assay protocol were used for the mutant strains with the addition of kanamycin to the overnight culture.

## 7.3 Materials and Methods: CDN specificity

### 7.3.1 Design of plasmids and template DNA for gel-based in vitro transcription termination assays

The *A. axanthum* (LR215048.1/1047420-1047511) riboswitch was cloned into a pUC19 plasmid. The riboswitch, including an upstream promoter and downstream native sequence, was PCR amplified using Phusion HF polymerase (NEB or Thermo Fisher Scientific) [157]. Targeted mutations were inserted via PCR of the aptamer using Phusion HF polymerase and primers with nucleotide changes. The PCR products were then gel purified (Thermo Fisher Scientific gel extraction), and promoters and downstream sequences were added with PCR using Phusion HF polymerase.

### 7.3.2 Preparation of mutagenized DNA library

A mutagenized *A. axanthum* library was prepared by chemical oligonucleotide synthesis with mixed base incorporation (Keck oligonucleotide synthesis facility, Yale). Mutations were restricted to the nucleotides that are within 10 Å of the ligand binding site as well as the terminator stem, as these regions are most likely to impact ligand binding. The library was designed to maximize single and double mutations using equation 1, where P is the probability of n mutations occurring in a construct of length L with a mutation rate of f [158].

$$P = \ {}_{L}C_{n} * f^{n} * (1-f)^{n} \tag{1}$$

Probabilities were then divided by the number of possible sequences with n mutations (equation 2) to find the frequency of each sequence occurring ($\nu_{seq}$) in the library (equation 3).

$$S = \ 3^{n} * \ {}_{L}C_{n} \tag{2}$$

$$\nu_{seq} = \frac{P}{S} \tag{3}$$

A mutation rate of 4% for each non-wild type base was used over 28 positions in first and second shell ligand contact as well as the terminator, which resulted in 2.5% wild type, 11% single mutants, 20% double mutants, and 66.5% with 3 or more mutations. With 28 positions, there are 3402 possible double mutants. With this mutation rate, approximately 400 reads per condition were expected for each double mutant. Due to length restrictions in mixed base oligonucleotide synthesis, two primers were combined using an overlap extension protocol with SSII reverse transcriptase (Invitrogen). Following overlap extension, a promoter upstream and 40 nucleotides downstream were added to the riboswitch with PCR using Phusion HF polymerase. This final product was used as template DNA in the SMARTT transcription termination assay [82].

### 7.3.3 Gel-based in vitro transcription termination assays

Transcription reactions were prepared on ice as reported in Torgerson et al. 2018 [82]. Reactions contained 40 mM Tris-HCl (pB 7.5), 150 mM KCl, 10μg/mL BAS, 1% glycerol, 10 mM $MgCl_2$, 1 mM DTT, 50 μM each ATP, CTP, UTP, and GTP, 30 nM DNA template, approximately 2.5 μCi [$\alpha$-$^{32}$P] GTP, 0.5 units *E. coli* RNA Polymerase Holoenzyme (NEB), and varying concentrations of ligand (c-GG or c-AG). Reactions were incubated at 37°C for 1 hour and then transferred to ice for 2 min and quenched with two volumes of loading buffer. Full-length and truncated RNA transcripts were separated on a 10% denaturing polyacrylamide gel (7 M Urea) and imaged with a Typhoon FLA (GE Healthcare). Individual bands were quantified with ImageQuant (GE Healthcare) and normalized based on the guanosine content of each sequence. Data from a single ligand titration were fit to equation 4:

$$\%\ \mathit{Full\text{-}Length} = (Y_{max} - Y_{min})\left(\frac{X}{K_{1/2}+X}\right) + Y_{min} \tag{4}$$

where X is ligand concentration, $Y_{min}$ is the percentage of full-length RNA with no ligand, $Y_{max}$ is the percentage of full-length RNA at saturating ligand concentrations, and $K_{1/2}$ is the concentration at half-maximal termination. Ligand titrations were performed in duplicate.

### 7.3.4 Preparation of RNA for high-throughput sequencing

A set of 100 μL transcription termination reactions was prepared on ice following Torgerson et al. 2018 [82]. Samples contained the same reagents as the gel-based assay described above except the radiolabeled GTP was not included. Each ligand concentration was prepared separately, incubated at 37°C for 1 hr, and then placed on ice for 2 minutes. The reaction was quenched and template DNA was removed with DNase treatment. The full transcription sample was combined with 2.2 μL water, 2.5 μL of 10x TURBO DNase buffer (Thermo Fisher Scientific), and 0.3 μL TURBO DNase (Thermo Fisher Scientific) and incubated at 37°C for 20 minutes. The RNA was then purified using an Oligo Clean & Concentrator Kit (Zymo).

A preadenylated DNA adapter was ligated onto the 3′ end of the RNA (/5rApp/NNNNNCTGTAGGCACCATCAAT/3ddC/ ordered from IDT) using T4 RNA Ligase 2, KQ (NEB). The final reaction mixture contained 1x T4 RNA Ligase reaction buffer, 25% PEG8000, 100 μM DNA adapter, 10 U/μL T4 RNA Ligase 2, KQ (NEB), and ~50% of the purified RNA. The reaction was incubated at 12°C for three hours and then purified using solid phase reversible immobilization (SPRI) magnetic beads (Backman Coulter).

The ligated adapter provides a primer annealing site for reverse-transcription into cDNA. 36.2 nM of primer (5′GATTGATGGTGCCTACAG) was annealed to the RNA in the presence of 725 μM dNTPs for 5 minutes at 65°C and then cooled to 4°C for 1 minute before the remaining reagents were added. The final reaction mixture containing 25 nM primer, 500 μM dNTPs, 1x SSIV buffer, 5 mM DTT, 1 U/μL RNaseOUT (Invitrogen), and 10 U/μL Superscript IV (Thermo Fisher Scientific) was incubated at 55°C for 10 minutes and then transferred to 80°C for 10 minutes to inactivate the Superscript IV enzyme. The RNA was then degraded by incubation with RNase A (0.5 μg/μL; Thermo Fisher Scientific) and RNase H (0.05 U/μL; Invitrogen) at 37°C for 30 minutes. The resulting cDNA was purified using SPRI magnetic beads (Beckman Coulter).

Illumina sequencing adapters, spacers to increase library diversity, and a unique index were added to the cDNA from each sample condition with a single round of PCR using Phusion HF Polymerase (NEB or Thermo Fisher Scientific) and primers with large 5′ overhangs (supplemental). 12 cycles of PCR were used to amplify the cDNA. Final DNA concentration was obtained using a Qubit high sensitivity dsDNA detection (Thermo Fisher Scientific). Samples were pooled with equal final concentrations before sequencing at the Yale Center for Genome Analysis (YCGA). Sequencing was performed on an Illumina HiSeq 4000 (2 x 150). The 30 samples produced by the two mutant libraries and two ligands were sequenced with 30% of a lane (~1% of a lane per sample).

### 7.3.5 Analysis of sequencing results

Analysis proceeded similarly to Torgerson et al. 2018 [82]. The region upstream of the aptamer and the 3′ adapter were removed using CutAdapt [159]. Bowtie2 was used to align sequences to a full-length WT template [160]. Dovetailing was permitted, but discordant sequences were removed. Custom Python scripts previously reported were used to determine the fraction of full-length and truncated reads for all variants with 0-2 mutations [82]. Sequences were classified as terminated if the last nucleotide sequenced was between 104-117 (inclusive), which is consistent with termination in the poly-uridine tract. The minimum full-length sequence was 144, consistent with readthrough transcription of the full construct. The number of full-length and truncated reads were then counted for each variant to determine percent full-length RNA at each ligand concentration. The data were fit to equation 4 and curves were plotted and visualized in Prism 9.

Once equations of fit were established (equation 4), $K_{1/2}$ was converted to a free energy using equation 5:

$$\Delta G_{sensitivity} = -RTln(K_{1/2}) \qquad (5)$$

where R is the ideal gas constant, T is the temperature of the transcription termination reaction (310 K), and $K_{1/2}$ is the measured value. The $\Delta\Delta G_{sensitivity}$ was found by subtracting the WT $\Delta G_{sensitivity}$ from the mutant value. Amplitude was transformed into a pseudoenergy using equation 6:

$$\Delta G_{switching} = -RTln(\frac{Amp}{Amp_{max}-Amp}) \qquad (6)$$

where Amp is the measured value and $Amp_{max}$ is the maximum amplitude measured in the sequencing sample. Similarly, the ΔΔGswitching was found by subtracting the WT ΔGswitching from the mutant value.

### 7.3.6 Calculation of Function

All data were analyzed using a functional parameter described previously [81]. With this parameter, any $f < 0$ describes a sequence that is more functional than WT with either a tighter sensitivity or larger amplitude while $f > 0$ represents a less functional riboswitch. The functional value for two replicate sequencing runs was averaged and sequences with a standard deviation greater than 1 were removed from further analysis. This excludes data with $K_{1/2} < 0.5$ µM, $K_{1/2} > 500$ µM. These data are difficult to fit due to the ligand concentrations tested. The filtered data does include mutants with amplitude less than 5% which would otherwise be excluded (figure 7.2). Further calculations were completed to describe the relationship between mutations in the functional landscape using equation 8:

$$Epistasis = f_{AB} - (\bar{f}_A + \bar{f}_B) \tag{8}$$

where $\bar{f}_A$ and $\bar{f}_B$ are the average function for all sequences that contain mutation A or mutation B respectively and $f_{AB}$ is the function of the double mutant that contains both mutation A and mutation B. This epistasis value was then visualized on a heat map using Prism 9.

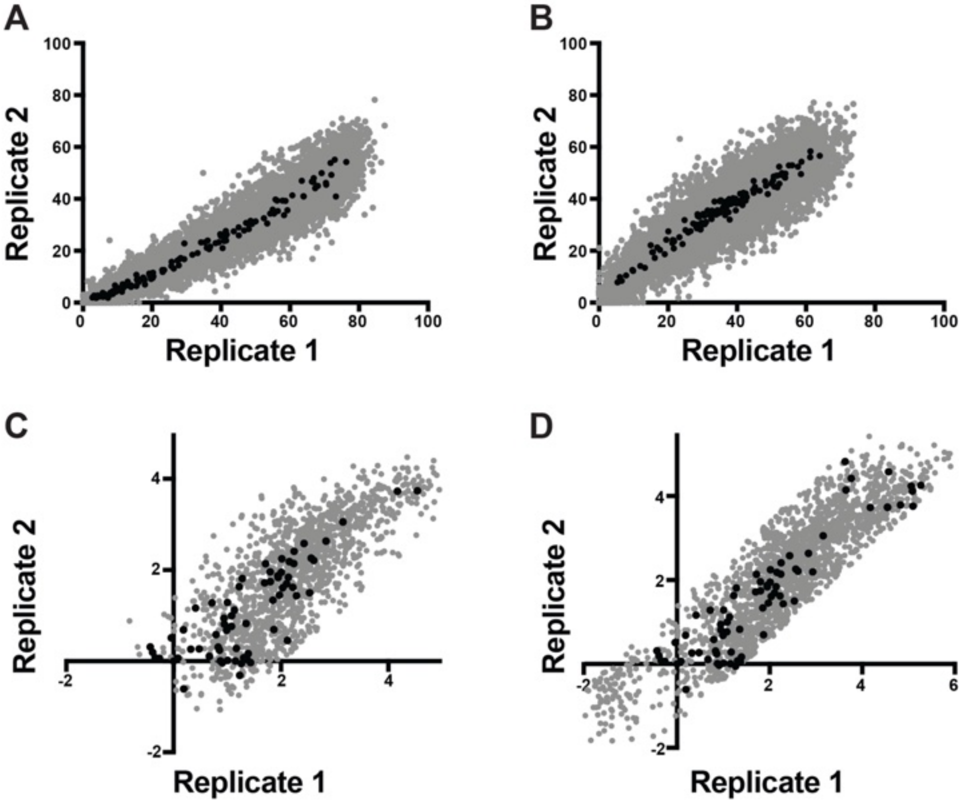

Figure 7.2: Replicate reproducibility of SMARTT. (A) Percent full-length for each mutant with 0 μM ligand. (B) Percent full-length for each mutant with 250 μM ligand. (C) Functional value for replicates with fits over 5% amplitude and measurable sensitivity. (D) Functional value for replicates where the standard deviation of the average of two replicates is less than 1.

### 7.3.7 Biochemical Consensus Diagram

Statistical analysis was performed using Prism 9. Sequences with functional values below the mean functional value of the dataset were selected to generate a pool of functional single and double mutants. The probability of each nucleotide at each position was calculated using previously published methods [81]. Cutoffs for nucleotide identity and purine/pyrimidine distinctions were made at 10% intervals between 70-90%. The two replicates were analyzed independently, and the least restrictive consensus diagram was used for analysis.

## 7.4 Materials and Methods: CDN evolution

### 7.4.1 Design of plasmids and template DNA for gel-based in vitro transcription termination

The *A. axanthum* (LR215048.1/1047420-1047511) riboswitch was prepared as previously indicated (7.2.1).

### 7.4.2 Preparation of mutagenized DNA library

A mutagenized *A. axanthum* library was prepared as previously indicated (7.2.2).

### 7.4.3 Chemical synthesis of all pyrimidine-containing cyclic dinucleotides

The things with the things – reference the hartig paper

### 7.4.4 Enzymatic synthesis of purine cyclic dinucleotides

The things with the things again – use references

### 7.4.5 Gel-based in vitro transcription termination assays

Transcription reactions were prepared on ice as reported in Torgerson et al. 2018 [82]. Reactions contained 40 mM Tris-HCl (pB 7.5), 150 mM KCl, 10μg/mL BAS, 1% glycerol, 10 mM $MgCl_2$, 1 mM DTT, 50 μM each ATP, CTP, UTP, and GTP, 30 nM DNA template, approximately 2.5 μCi [$\alpha$-$^{32}$P] GTP, 0.5 units *E. coli* RNA Polymerase Holoenzyme (NEB), and varying concentrations of ligand (cyclic dinucleotides with all base combinations). Reactions were incubated at 37°C for 1 hour and then transferred to ice for 2 min and quenched with two volumes of loading buffer. Full-length and truncated RNA transcripts were separated on a 10% denaturing polyacrylamide gel (7 M Urea) and imaged with a Typhoon FLA (GE Healthcare). Individual bands were quantified with ImageQuant (GE Healthcare) and normalized based on the guanosine content of each sequence. Data from a single ligand titration were fit to equation 4:

$$\%\, Full\text{-}Length = (Y_{max} - Y_{min})\left(\frac{X}{K_{1/2}+X}\right) + Y_{min} \tag{4}$$

where X is ligand concentration, $Y_{min}$ is the percentage of full-length RNA with no ligand, $Y_{max}$ is the percentage of full-length RNA at saturating ligand concentrations, and $K_{1/2}$ is the concentration at half-maximal termination. Ligand titrations were performed in duplicate.

### 7.4.6 Preparation of RNA for high-throughput sequencing

A set of 100 μL transcription termination reactions was prepared on ice following Torgerson et al. 2018 [82]. Samples contained the same reagents as the gel-based assay described above except the radiolabeled GTP was not included. To maximize sequencing depth, all uridine-containing cyclic dinucleotides were pooled and a single high concentration was assayed. Similarly, all cytidine-containing cyclic dinucleotides were pooled and a single high concentration was assayed. Each ligand concentrationwas prepared separately, incubated at 37°C for 1 hr, and then placed on ice for 2 minutes. The reaction was quenched and template DNA was removed with DNase treatment. The full transcription sample was combined with 2.2 μL water, 2.5 μL of 10x TURBO DNase buffer (Thermo Fisher Scientific), and 0.3 μL TURBO DNase (Thermo Fisher Scientific) and incubated at 37°C for 20 minutes. The RNA was then purified using an Oligo Clean & Concentrator Kit (Zymo).

Sequences that turned on in the presence of pyrimidine-containing cyclic dinucleotides were selected for using a primer for reverse transcription that anneals to the full-length transcript. 36.2 nM of primer was annealed to the RNA in the presence of 725 μM dNTPs for 5 minutes at 65°C and then cooled to 4°C for 1 minute before the remaining reagents were added. The final reaction mixture containing 25 nM primer, 500 μM dNTPs, 1x SSIV buffer, 5 mM DTT, 1 U/μL RNaseOUT (Invitrogen), and 10 U/μL Superscript IV (Thermo Fisher Scientific) was incubated at 55°C for 10 minutes and then transferred to 80°C for 10 minutes to inactivate the Superscript IV enzyme. The RNA was then degraded by incubation with RNase A (0.5 μg/μL; Thermo Fisher Scientific) and RNase H (0.05 U/μL; Invitrogen) at 37°C for 30 minutes. The resulting cDNA was purified using SPRI magnetic beads (Beckman Coulter). An upstream promoter was added to the cDNA. The riboswitch cDNA was PCR amplified using Phusion HF polymerase (NEB or Thermo Fisher Scientific) with primers that added promoters upstream [157]. This selected DNA library was then used for a second round of transcription termination following the same steps outlined above. After DNAase treatment for removal of template DNA and column purification, the transcribed RNA proceeded to sequencing preparation.

A preadenylated DNA adapter was ligated onto the 3′ end of the RNA (/5rApp/NNNNNCTGTAGGCACCATCAAT/3ddC/ ordered from IDT) using T4 RNA Ligase 2, KQ (NEB). The final reaction mixture contained 1x T4 RNA Ligase reaction buffer, 25% PEG8000, 100 μM DNA adapter, 10 U/μL T4 RNA Ligase 2, KQ (NEB), and ~50% of the purified RNA. The reaction was incubated at 12°C for three hours and then purified using solid phase reversible immobilization (SPRI) magnetic beads (Backman Coulter).

The ligated adapter provides a primer annealing site for reverse-transcription into cDNA. 36.2 nM of primer (5′GATTGATGGTGCCTACAG) was annealed to the RNA in the presence of 725 μM dNTPs for 5 minutes at 65°C and then cooled to 4°C for 1 minute before the remaining reagents were added. The final reaction mixture containing 25 nM primer, 500 μM dNTPs, 1x SSIV buffer, 5 mM DTT, 1 U/μL RNaseOUT (Invitrogen), and 10 U/μL Superscript IV (Thermo Fisher Scientific) was incubated at 55°C for 10 minutes and then transferred to 80°C for 10 minutes to inactivate the Superscript IV enzyme. The RNA was then degraded by incubation with RNase A (0.5 μg/μL; Thermo Fisher Scientific) and RNase H (0.05 U/μL; Invitrogen) at 37°C for 30 minutes. The resulting cDNA was purified using SPRI magnetic beads (Beckman Coulter).

Illumina sequencing adapters, spacers to increase library diversity, and a unique index were added to the cDNA from each sample condition with a single round of PCR using Phusion HF Polymerase (NEB or Thermo Fisher Scientific) and primers with large 5′ overhangs (supplemental). 12 cycles of PCR were used to amplify the cDNA. Final DNA concentration was obtained using a Qubit high sensitivity dsDNA detection (Thermo Fisher Scientific). Samples were pooled with equal final concentrations before sequencing at the Yale Center for Genome Analysis (YCGA). Sequencing was performed on an Illumina HiSeq 4000 (2 x 150). The 30 samples produced by the two mutant libraries and two ligands were sequenced with 30% of a lane (~1% of a lane per sample).

### 7.4.7 Analysis of sequencing results

Analysis proceeded similarly to Torgerson et al. 2018 [82]. The region upstream of the aptamer and the 3′ adapter were removed using CutAdapt [159]. Bowtie2 was used to align sequences to a full-length WT template [160].

Dovetailing was permitted, but discordant sequences were removed. Custom Python scripts previously reported were used to determine the fraction of full-length and truncated reads for all variants with 0-2 mutations [82]. Sequences were classified as terminated if the last nucleotide sequenced was between 104-117 (inclusive), which is consistent with termination in the poly-uridine tract. The minimum full-length sequence was 144 nucleotides, consistent with readthrough transcription of the full construct. The number of full-length and truncated reads were then counted for each variant to determine percent full-length RNA at each ligand concentration. The high and low ligand data were subtracted to calculate an approximate amplitude for each mutant sequence. The data were visualized in Prism 9.

# 8. Appendix

## 8.1 DNA sequences for transcription termination

*P. micra* Wild type – PRPP
TTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGCTAAGTGATATTGATAACTCA
AAGAGAAAGTGCACATAGGGTTCCGGAGATTTTTCTTGCAGGTCCAAGCTGTG
CAGGTGCTATGCACTACACCTAAGGGAGAAAAGCCCAGAAGGTAGGTTTCGC
TTGAAGCCTACTTTTTGGGCGTTTATTTTTTTTAGATGTTATTTTTTAAGAG

*P. micra* G93A – PRPP
TTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGCTAAGTGATATTGATAACTCA
AAGAGAAAGTGCACATAGGGTTCCGGAGATTTTTCTTGCAGGTCCAAGCTGTG
CAGGTGCTATGCACTACACCTAAGGGAGAAAAGCCCAAAAGGTAGGTTTCGC
TTGAAGCCTACTTTTTGGGCGTTTATTTTTTTTAGATGTTATTTTTTAAGAG

*T. oceani* Wild type – ppGpp
TTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGCTAAGCATAATTAGGAAGTG
TACCTTAGGGTTCCGGCCATAAGGCGTCAGCGACCGAGCGGTACAATCCGGG
GAAACCCGGAACACCGTGAGCATAAAAGGCTCCAGCGGCAAGTTCTTAAAA
GAACTAGCCGCTGTTTTTTTATAATCAAAAACCGCAAAAATGTGATGAAGGAG
ACGGAATTAATGATCACGGTCAAAGAGAGCCGGGGAAGGT

*T. oceani* A93G – ppGpp
TTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGCTAAGCATAATTAGGAAGTG
TACCTTAGGGTTCCGGCCATAAGGCGTCAGCGACCGAGCGGTACAATCCGGG
GAAACCCGGAACACCGTGAGCATAAAAGGCTCCGGCGGCAAGTTCTTAAAA
GAACTAGCCGCTGTTTTTTTATAATCAAAAACCGCAAAAATGTGATGAAGGAG
ACGGAATTAATGATCACGGTCAAAGAGAGCCGGGGAAGGT

A. axanthum Wild type – cyclic dinucleotide
TTACACTTTATGCTTCCGGCTCGTATAATGTGTGTATATATGAGCATCGAAAGAA
TGTAAAAGGCAAACCAAGGGTAACCTTGGGACGCAAAGCTATAGGGTCCTAA
AATTGGATAGCCAGTTGTCATTATGACAACTGGCTTTTTTTGTCAATTTAGGAGG
TGATTTGGATGCGGTTATCTGAAATATG

A. axanthum U92C – cyclic dinucleotide
TTACACTTTATGCTTCCGGCTCGTATAATGTGTGTATATATGAGCATCGAAAGAA
TGTAAAAGGCAAACCAAGGGTAACCTTGGGACGCAAAGCTATAGGGTCCTAA
AATTGGATAGCCAGTTGCCATTATGACAACTGGCTTTTTTTGTCAATTTAGGAG
GTGATTTGGATGCGGTTATCTGAAATATG

## 8.2 DNA sequences for SMARTT

*T. oceani* Wild type – ppGpp
TTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGCTAAG (J23119 promoter) –
ACCAGTCTAAAGCATAATTA (5′ constant region) –
GGAAGTGTACCTTAGGGTTCCGGCCATAAGGCGTCAGCGACCGAGCGGTACA
ATCCGGGGAAACCCGGAACACCGTGAGCATAAAAGGCTCCAGCGGCAAGTT
CTTAAAAGAACTAGCCGCTG (mutated region) – TTTTTTTA (termination site) –
TAATCAAAAACCGCAAAAATGTGATGAAGGAGACGGAATTAATGATCACGGTC
AAAGAGAGCCGGGGAAGGT (3′-downstream region)

*T. oceani* A93G – ppGpp
TTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGCTAAG (J23119 promoter) –
ACCAGTCTAAAGCATAATTA (5′ constant region) –
GGAAGTGTACCTTAGGGTTCCGGCCATAAGGCGTCAGCGACCGAGCGGTACA
ATCCGGGGAAACCCGGAACACCGTGAGCATAAAAGGCTCCGGCGGCAAGTT
CTTAAAAGAACTAGCCGCTG (mutated region) – TTTTTTTA (termination site) –
TAATCAAAAACCGCAAAAATGTGATGAAGGAGACGGAATTAATGATCACGGTC
AAAGAGAGCCGGGGAAGGT (3′-downstream region)

*A. axanthum* Wild type – cyclic dinucleotide
TTACACTTTATGCTTCCGGCTCGTATAATG (lacUV5 promoter) –
TGTGTATATATGAGCATCGAAAG (5′ constant region) –
AATGTAAAAGGCAAACCAAGGGTAACCTTGGGACGCAAAGCTATAGGGTCCT
AAAATTGGATAGCCAGTTGTCATTATGACAACTGGC (mutated region) –
TTTTTTTG (termination site) –
TCAATTTAGGAGGTGATTTGGATGCGGTTATCTGAAATATG (3′ downstream region)

A. axanthum U92C – cyclic dinucleotide
TTACACTTTATGCTTCCGGCTCGTATAATG (lacUV5 promoter) –
TGTGTATATATGAGCATCGAAAG (5′ constant region) –
AATGTAAAAGGCAAACCAAGGGTAACCTTGGGACGCAAAGCTATAGGGTCCT
AAAATTGGATAGCCAGTTGCCATTATGACAACTGGC (mutated region) –
TTTTTTTG (termination site ) –
TCAATTTAGGAGGTGATTTGGATGCGGTTATCTGAAATATG (3′ downstream region)

3′ Ligated Adaptor
5′-rApp-NNNNNCTGTAGGCACCATCAAT-ddC

RT Primer
5′-GATTGATGGTGCCTACAG

## 8.3 RNA sequences for MST

*A. inagensis* WT – GMP/XMP
GACCGACGGUGGUGACUAGGGUUCCGCCGCGCCGAACGCGGGCGCGGG
UCUGGUUCGAGCGUCACGGAAAGCCCGCGACCUGCGGGCAGGCAUCU
CCGGGAGAGAAACCCAGGGAGGGACCGGUCGG

## 8.4 DNA insert for reporter assay

*A. inagensis* WT – GMP/XMP
TACGACGAATTCCAAGAATAATGTTGATCCTTTTAAATAAGTCTGATAAAATGTG
AACTAAGACCGACGGTGGTGACTAGGGTTCCGCCGCGCCGAACGCGGGCG
CGGGTCTGGTTCGAGCGTCACGGAAAGCCCGCGACCTGCGGGCAGGCA
TCTCCGGGAGAGAAACCCAGGGAGGGACCGGTCGGCGCACAGACATGTCTC
GACATGGACGCTGCGCCGATGGGGACCGCTCCCCGCGTGCTCGCGGCTTCC
GTGCCAACGGAAGGGTGAGGCCAGATGACGGAGAGTCAACGCCCGGTCCT
GGATCCTCAGC

# 9. References

1. Starosta, A.L., Lassak, J., Jung, K., and Wilson, D.N. (2014). The bacterial translation stress response. FEMS Microbiol. Rev. *38*, 1172–1201. 10.1111/1574-6976.12083.

2. Cummins, E.P., and Taylor, C.T. (2005). Hypoxia-responsive transcription factors. Pflüg. Arch. *450*, 363–371. 10.1007/s00424-005-1413-7.

3. Lepock, J.R. (2005). How do cells respond to their thermal environment? Int. J. Hyperthermia *21*, 681–687. 10.1080/02656730500307298.

4. Orphanides, G., and Reinberg, D. (2002). A Unified Theory of Gene Expression. Cell *108*, 439–451. 10.1016/S0092-8674(02)00655-4.

5. Strasser, A., O'Connor, L., and Dixit, V.M. (2000). Apoptosis Signaling. Annu. Rev. Biochem. *69*, 217–245. 10.1146/annurev.biochem.69.1.217.

6. Zalkin, H., and Dixon, J.E. (1992). De Novo Purine Nucleotide Biosynthesis. In Progress in Nucleic Acid Research and Molecular Biology, W. E. Cohn and K. Moldave, eds. (Academic Press), pp. 259–287. 10.1016/S0079-6603(08)60578-4.

7. Marles-Wright, J., and Lewis, R.J. (2007). Stress responses of bacteria. Curr. Opin. Struct. Biol. *17*, 755–760. 10.1016/j.sbi.2007.08.004.

8. Storz, G., Vogel, J., and Wassarman, K.M. (2011). Regulation by Small RNAs in Bacteria: Expanding Frontiers. Mol. Cell *43*, 880–891. 10.1016/j.molcel.2011.08.022.

9. Lotz, T.S., and Suess, B. (2018). Small-Molecule-Binding Riboswitches. Microbiol. Spectr. *6*, 10.1128/microbiolspec.rwr-0025–2018. 10.1128/microbiolspec.rwr-0025-2018.

10. Ariza-Mateos, A., Nuthanakanti, A., and Serganov, A. (2021). Riboswitch Mechanisms: New Tricks for an Old Dog. Biochem. Mosc. *86*, 962–975. 10.1134/S0006297921080071.

11. Bédard, A.-S.V., Hien, E.D.M., and Lafontaine, D.A. (2020). Riboswitch regulation mechanisms: RNA, metabolites and regulatory proteins. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1863*, 194501. 10.1016/j.bbagrm.2020.194501.

12. Kavita, K., and Breaker, R.R. (2023). Discovering riboswitches: the past and the future. Trends Biochem. Sci. *48*, 119–141. 10.1016/j.tibs.2022.08.009.

13. Sherlock, M.E., Higgs, G., Yu, D., Widner, D.L., White, N.A., Sudarsan, N., Sadeeshkumar, H., Perkins, K.R., Mirihana Arachchilage, G., Malkowski, S.N., *et al.* (2022). Architectures and complex functions of tandem riboswitches. RNA Biol. *19*, 1059–1076. 10.1080/15476286.2022.2119017.

14. Sherlock, M.E., Sudarsan, N., Stav, S., and Breaker, R.R. (2018). Tandem riboswitches form a natural Boolean logic gate to control purine metabolism in bacteria. eLife *7*, e33908. 10.7554/eLife.33908.

15. Sudarsan, N., Hammond, M.C., Block, K.F., Welz, R., Barrick, J.E., Roth, A., and Breaker, R.R. (2006). Tandem Riboswitch Architectures Exhibit Complex Gene Control Functions. Science *314*, 300–304. 10.1126/science.1130716.

16. Gusarov, I., and Nudler, E. (1999). The Mechanism of Intrinsic Transcription Termination. Mol. Cell *3*, 495–504. 10.1016/S1097-2765(00)80477-3.

17. Yarnell, W.S., and Roberts, J.W. (1999). Mechanism of Intrinsic Transcription Termination and Antitermination. Science *284*, 611–615. 10.1126/science.284.5414.611.

18. You, L., Omollo, E.O., Yu, C., Mooney, R.A., Shi, J., Shen, L., Wu, X., Wen, A., He, D., Zeng, Y., *et al.* (2023). Structural basis for intrinsic transcription termination. Nature *613*, 783–789. 10.1038/s41586-022-05604-1.

19. Breaker, R.R. (2018). Riboswitches and Translation Control. Cold Spring Harb. Perspect. Biol. *10*, a032797. 10.1101/cshperspect.a032797.

20. Weinberg, Z., Barrick, J.E., Yao, Z., Roth, A., Kim, J.N., Gore, J., Wang, J.X., Lee, E.R., Block, K.F., Sudarsan, N., *et al.* (2007). Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. Nucleic Acids Res. *35*, 4809–4819. 10.1093/nar/gkm487.

21. Stav, S., Atilho, R.M., Mirihana Arachchilage, G., Nguyen, G., Higgs, G., and Breaker, R.R. (2019). Genome-wide discovery of structured noncoding RNAs in bacteria. BMC Microbiol. *19*. 10.1186/s12866-019-1433-7.

22. Brewer, K.I., Greenlee, E.B., Higgs, G., Yu, D., Mirihana Arachchilage, G., Chen, X., King, N., White, N., and Breaker, R.R. (2021). Comprehensive discovery of novel structured noncoding RNAs in 26 bacterial genomes. RNA Biol. *18*, 2417–2432. 10.1080/15476286.2021.1917891.

23. Yao, Z., Barrick, J., Weinberg, Z., Neph, S., Breaker, R., Tompa, M., and Ruzzo, W.L. (2007). A Computational Pipeline for High- Throughput

Discovery of cis-Regulatory Noncoding RNA in Prokaryotes. PLOS Comput. Biol. *3*, e126. 10.1371/journal.pcbi.0030126.

24. Weinberg, Z., and Breaker, R.R. (2011). R2R - software to speed the depiction of aesthetic consensus RNA secondary structures. BMC Bioinformatics *12*, 3. 10.1186/1471-2105-12-3.

25. Baker, J.L., Sudarsan, N., Weinberg, Z., Roth, A., Stockbridge, R.B., and Breaker, R.R. (2012). Widespread Genetic Switches and Toxicity Resistance Proteins for Fluoride. Science *335*, 233–235. 10.1126/science.1215063.

26. Regulski, E.E., and Breaker, R.R. (2008). In-Line Probing Analysis of Riboswitches. In Post-Transcriptional Gene Regulation Methods In Molecular Biology™., J. Wilusz, ed. (Totowa, NJ: Humana Press), pp. 53–67. 10.1007/978-1-59745-033-1_4.

27. White, N., Sadeeshkumar, H., Sun, A., Sudarsan, N., and Breaker, R.R. (2022). Lithium-sensing riboswitch classes regulate expression of bacterial cation transporter genes. Sci. Rep. *12*, 19145. 10.1038/s41598-022-20695-6.

28. White, N., Sadeeshkumar, H., Sun, A., Sudarsan, N., and Breaker, R.R. (2022). Na+ riboswitches regulate genes for diverse physiological processes in bacteria. Nat. Chem. Biol. *18*, 878–885. 10.1038/s41589-022-01086-4.

29. Breaker, R.R. (2022). The Biochemical Landscape of Riboswitch Ligands. Biochemistry *61*, 137–149. 10.1021/acs.biochem.1c00765.

30. Serganov, A., Yuan, Y.-R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A.T., Hobartner, C., Micura, R., Breaker, R.R., and Patel, D.J. (2004). Structural Basis for Discriminative Regulation of Gene Expression by Adenine- and Guanine-Sensing mRNAs. Chem. Biol. *11*, 1729–1741. 10.1016/j.chembiol.2004.11.018.

31. Price, I.R., Gaballa, A., Ding, F., Helmann, J.D., and Ke, A. (2015). Mn2+-Sensing Mechanisms of yybP-ykoY Orphan Riboswitches. Mol. Cell *57*, 1110–1123. 10.1016/j.molcel.2015.02.016.

32. Nelson, J.W., Atilho, R.M., Sherlock, M.E., Stockbridge, R.B., and Breaker, R.R. (2017). Metabolism of Free Guanidine in Bacteria Is Regulated by a Widespread Riboswitch Class. Mol. Cell *65*, 220–230. 10.1016/j.molcel.2016.11.019.

33. Salvail, H., Balaji, A., Yu, D., Roth, A., and Breaker, R.R. (2020). Biochemical Validation of a Fourth Guanidine Riboswitch Class in Bacteria. Biochemistry *59*, 4654–4662. 10.1021/acs.biochem.0c00793.

34. Sherlock, M.E., Malkowski, S.N., and Breaker, R.R. (2017). Biochemical Validation of a Second Guanidine Riboswitch Class in Bacteria. Biochemistry *56*, 352–358. 10.1021/acs.biochem.6b01270.

35. Sherlock, M.E., and Breaker, R.R. (2017). Biochemical Validation of a Third Guanidine Riboswitch Class in Bacteria. Biochemistry *56*, 359–363. 10.1021/acs.biochem.6b01271.

36. Lenkeit, F., Eckert, I., Hartig, J.S., and Weinberg, Z. (2020). Discovery and characterization of a fourth class of guanidine riboswitches. Nucleic Acids Res. *48*, 12889–12899. 10.1093/nar/gkaa1102.

37. Reiss, C.W., and Strobel, S.A. (2017). Structural basis for ligand binding to the guanidine-II riboswitch. RNA *23*, 1338–1343. 10.1261/rna.061804.117.

38. Reiss, C.W., Xiong, Y., and Strobel, S.A. (2017). Structural Basis for Ligand Binding to the Guanidine-I Riboswitch. Structure *25*, 195–202. 10.1016/j.str.2016.11.020.

39. Huang, L., Wang, J., Wilson, T.J., and Lilley, D.M.J. (2017). Structure of the Guanidine III Riboswitch. Cell Chem. Biol. *24*, 1407-1415.e2. 10.1016/j.chembiol.2017.08.021.

40. Huang, L., Wang, J., and Lilley, D.M.J. (2017). The Structure of the Guanidine-II Riboswitch. Cell Chem. Biol. *24*, 695-702.e2. 10.1016/j.chembiol.2017.05.014.

41. Ren, A., Rajashankar, K.R., and Patel, D.J. (2012). Fluoride ion encapsulation by Mg2+ ions and phosphates in a fluoride riboswitch. Nature *486*, 85–89. 10.1038/nature11152.

42. Chawla, M., Credendino, R., Poater, A., Oliva, R., and Cavallo, L. (2015). Structural Stability, Acidity, and Halide Selectivity of the Fluoride Riboswitch Recognition Site. J. Am. Chem. Soc. *137*, 299–306. 10.1021/ja510549b.

43. Sudarsan, N., Barrick, J.E., and Breaker, R.R. (2003). Metabolite-binding RNA domains are present in the genes of eukaryotes. RNA *9*, 644–647. 10.1261/rna.5090103.

44. Serganov, A., Polonskaia, A., Phan, A.T., Breaker, R.R., and Patel, D.J. (2006). Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. Nature *441*, 1167–1171. 10.1038/nature04740.

45. Sudarsan, N., Cohen-Chalamish, S., Nakamura, S., Emilsson, G.M., and Breaker, R.R. (2005). Thiamine Pyrophosphate Riboswitches Are Targets for the Antimicrobial Compound Pyrithiamine. Chem. Biol. *12*, 1325–1335. 10.1016/j.chembiol.2005.10.007.

46. Sherlock, M.E., and Breaker, R.R. (2020). Former orphan riboswitches reveal unexplored areas of bacterial metabolism, signaling, and gene control processes. RNA *26*, 675–693. 10.1261/rna.074997.120.

47. Mandal, M., and Breaker, R.R. (2004). Adenine riboswitches and gene activation by disruption of a transcription terminator. Nat. Struct. Mol. Biol. *11*, 29–35. 10.1038/nsmb710.

48. Mandal, M., Boese, B., Barrick, J.E., Winkler, W.C., and Breaker, R.R. (2003). Riboswitches Control Fundamental Biochemical Pathways in Bacillus subtilis and Other Bacteria. Cell *113*, 577–586. 10.1016/S0092-8674(03)00391-X.

49. Nelson, J.W., Sudarsan, N., Phillips, G.E., Stav, S., Lünse, C.E., McCown, P.J., and Breaker, R.R. (2015). Control of bacterial exoelectrogenesis by c-AMP-GMP. Proc. Natl. Acad. Sci. *112*, 5389–5394. 10.1073/pnas.1419264112.

50. Kellenberger, C.A., Wilson, S.C., Hickey, S.F., Gonzalez, T.L., Su, Y., Hallberg, Z.F., Brewer, T.F., Iavarone, A.T., Carlson, H.K., Hsieh, Y.-F., *et al.* (2015). GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP. Proc. Natl. Acad. Sci. *112*, 5383–5388. 10.1073/pnas.1419328112.

51. Sudarsan, N., Lee, E.R., Weinberg, Z., Moy, R.H., Kim, J.N., Link, K.H., and Breaker, R.R. (2008). Riboswitches in eubacteria sense the second messenger cyclic di-GMP. Science *321*, 411–413. 10.1126/science.1159519.

52. Batey, R.T., Gilbert, S.D., and Montange, R.K. (2004). Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. Nature *432*, 411–415. 10.1038/nature03037.

53. Hamal Dhakal, S., Panchapakesan, S.S.S., Slattery, P., Roth, A., and Breaker, R.R. (2022). Variants of the guanine riboswitch class exhibit altered ligand specificities for xanthine, guanine, or 2′-deoxyguanosine. Proc. Natl. Acad. Sci. *119*, e2120246119. 10.1073/pnas.2120246119.

54. Kim, J.N., Roth, A., and Breaker, R.R. (2007). Guanine riboswitch variants from Mesoplasma florum selectively recognize 2′-deoxyguanosine. Proc. Natl. Acad. Sci. *104*, 16092–16097. 10.1073/pnas.0705884104.

55. Edwards, A.L., and Batey, R.T. (2009). A Structural Basis for the Recognition of 2′-Deoxyguanosine by the Purine Riboswitch. J. Mol. Biol. *385*, 938–948. 10.1016/j.jmb.2008.10.074.

56. Smith, K.D., Lipchock, S.V., Ames, T.D., Wang, J., Breaker, R.R., and Strobel, S.A. (2009). Structural basis of ligand binding by a c-di-GMP riboswitch. Nat. Struct. Mol. Biol. *16*, 1218–1223. 10.1038/nsmb.1702.

57. Wang, C., Sinn, M., Stifel, J., Heiler, A.C., Sommershof, A., and Hartig, J.S. (2017). Synthesis of All Possible Canonical (3′–5′-Linked) Cyclic Dinucleotides and Evaluation of Riboswitch Interactions and Immune-Stimulatory Effects. J. Am. Chem. Soc. 10.1021/jacs.7b06141.

58. Barrick, J.E., Corbino, K.A., Winkler, W.C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I., *et al.* (2004). New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. Proc. Natl. Acad. Sci. *101*, 6421–6426. 10.1073/pnas.0308014101.

59. Sherlock, M.E., Sudarsan, N., and Breaker, R.R. (2018). Riboswitches for the alarmone ppGpp expand the collection of RNA-based signaling systems. Proc. Natl. Acad. Sci. *115*, 6052–6057. 10.1073/pnas.1720406115.

60. Sherlock, M.E., Sadeeshkumar, H., and Breaker, R.R. (2019). Variant Bacterial Riboswitches Associated with Nucleotide Hydrolase Genes Sense Nucleoside Diphosphates. Biochemistry *58*, 401–410. 10.1021/acs.biochem.8b00617.

61. Greenlee, E.B., Stav, S., Atilho, R.M., Brewer, K.I., Harris, K.A., Malkowski, S.N., Mirihana Arachchilage, G., Perkins, K.R., Sherlock, M.E., and Breaker, R.R. (2018). Challenges of ligand identification for the second wave of orphan riboswitch candidates. RNA Biol. *15*, 377–390. 10.1080/15476286.2017.1403002.

62. Dalebroux, Z.D., and Swanson, M.S. (2012). ppGpp: magic beyond RNA polymerase. Nat. Rev. Microbiol. *10*, 203–212. 10.1038/nrmicro2720.

63. Gaca, A.O., Colomer-Winter, C., and Lemos, J.A. (2015). Many Means to a Common End: the Intricacies of (p)ppGpp Metabolism and Its Control of Bacterial Homeostasis. J. Bacteriol. *197*, 1146–1156. 10.1128/JB.02577-14.

64. Knappenberger, A.J., Reiss, C.W., and Strobel, S.A. (2018). Structures of two aptamers with differing ligand specificity reveal ruggedness in the functional landscape of RNA. eLife *7*, e36381. 10.7554/eLife.36381.

65. Knappenberger, A.J., Reiss, C.W., Focht, C.M., and Strobel, S.A. (2020). A Modular RNA Domain That Confers Differential Ligand Specificity. Biochemistry *59*, 1361–1366. 10.1021/acs.biochem.0c00117.

66. Peselis, A., and Serganov, A. (2018). ykkC riboswitches employ an add-on helix to adjust specificity for polyanionic ligands. Nat. Chem. Biol. *14*, 887–894. 10.1038/s41589-018-0114-4.

67. Kelvin, D., and Suess, B. (2023). Tapping the potential of synthetic riboswitches: reviewing the versatility of the tetracycline aptamer. RNA Biol. *20*, 457–468. 10.1080/15476286.2023.2234732.

68. Groher, F., and Suess, B. (2014). Synthetic riboswitches — A tool comes of age. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1839*, 964–973. 10.1016/j.bbagrm.2014.05.005.

69. Dey, S.K., Filonov, G.S., Olarerin-George, A.O., Jackson, B.T., Finley, L.W.S., and Jaffrey, S.R. (2022). Repurposing an adenine riboswitch into a fluorogenic imaging and sensing tag. Nat. Chem. Biol. *18*, 180–190. 10.1038/s41589-021-00925-0.

70. Su, Y., and Hammond, M.C. (2020). RNA-based fluorescent biosensors for live cell imaging of small molecules and RNAs. Curr. Opin. Biotechnol. *63*, 157–166. 10.1016/j.copbio.2020.01.001.

71. Litke, J.L., You, M., and Jaffrey, S.R. (2016). Chapter Fourteen - Developing Fluorogenic Riboswitches for Imaging Metabolite Concentration Dynamics in Bacterial Cells. Methods Enzymol. *572*, 315–333. 10.1016/bs.mie.2016.03.021.

72. Ellington, A.D., and Szostak, J.W. (1990). In vitro selection of RNA molecules that bind specific ligands. Nature *346*, 818–822. 10.1038/346818a0.

73. Hoetzel, J., and Suess, B. (2022). Structural Changes in Aptamers are Essential for Synthetic Riboswitch Engineering. J. Mol. Biol. *434*, 167631. 10.1016/j.jmb.2022.167631.

74. Kaiser, C., Schneider, J., Groher, F., Suess, B., and Wachtveitl, J. (2021). What defines a synthetic riboswitch? – Conformational dynamics of ciprofloxacin aptamers with similar binding affinities but varying regulatory potentials. Nucleic Acids Res. 10.1093/nar/gkab166.

75. Mohsen, M.G., Midy, M.K., Balaji, A., and Breaker, R.R. (2023). Exploiting natural riboswitches for aptamer engineering and validation. Nucleic Acids Res. *51*, 966–981. 10.1093/nar/gkac1218.

76. Andreasson, J.O.L., Savinov, A., Block, S.M., and Greenleaf, W.J. (2020). Comprehensive sequence-to-function mapping of cofactor-dependent RNA catalysis in the glmS ribozyme. Nat. Commun. *11*. 10.1038/s41467-020-15540-1.

77. Kobori, S., Nomura, Y., Miu, A., and Yokobayashi, Y. (2015). High-throughput assay and engineering of self-cleaving ribozymes by sequencing. Nucleic Acids Res. *43*, e85–e85. 10.1093/nar/gkv265.

78. Kobori, S., and Yokobayashi, Y. (2016). High-Throughput Mutational Analysis of a Twister Ribozyme. Angew. Chem. Int. Ed., n/a-n/a. 10.1002/anie.201605470.

79. Kobori, S., Takahashi, K., and Yokobayashi, Y. (2017). Deep Sequencing Analysis of Aptazyme Variants Based on a Pistol Ribozyme. ACS Synth. Biol. *6*, 1283–1288. 10.1021/acssynbio.7b00057.

80. Kobori, S., and Yokobayashi, Y. (2018). Analyzing and Tuning Ribozyme Activity by Deep Sequencing To Modulate Gene Expression Level in Mammalian Cells. ACS Synth. Biol. *7*, 371–376. 10.1021/acssynbio.7b00367.

81. Focht, C.M., Hiller, D.A., Grunseich, S.G., and Strobel, S.A. (2023). Translation regulation by a Guanidine-II riboswitch is highly tunable in sensitivity, dynamic range, and apparent cooperativity. RNA, rna.079560.122. 10.1261/rna.079560.122.

82. Torgerson, C.D., Hiller, D.A., Stav, S., and Strobel, S.A. (2018). Gene regulation by a glycine riboswitch singlet uses a finely tuned energetic landscape for helical switching. RNA *24*, 1813–1827. 10.1261/rna.067884.118.

83. Torgerson, C.D., Hiller, D.A., and Strobel, S.A. (2020). The asymmetry and cooperativity of tandem glycine riboswitch aptamers. RNA *26*, 564–580. 10.1261/rna.073577.119.

84. Chatterjee, S., Chauvier, A., Dandpat, S.S., Artsimovitch, I., and Walter, N.G. (2021). A translational riboswitch coordinates nascent transcription–translation coupling. Proc. Natl. Acad. Sci. *118*. 10.1073/pnas.2023426118.

85. Strobel, E.J., Cheng, L., Berman, K.E., Carlson, P.D., and Lucks, J.B. (2019). A ligand-gated strand displacement mechanism for ZTP riboswitch

transcription control. Nat. Chem. Biol. *15*, 1067–1076. 10.1038/s41589-019-0382-7.

86. Ray, S., Chauvier, A., and Walter, N.G. (2019). Kinetics coming into focus: single-molecule microscopy of riboswitch dynamics. RNA Biol. *16*, 1077–1085. 10.1080/15476286.2018.1536594.

87. Weinberg, Z., Nelson, J.W., Lünse, C.E., Sherlock, M.E., and Breaker, R.R. (2017). Bioinformatic analysis of riboswitch structures uncovers variant classes with altered ligand specificity. Proc. Natl. Acad. Sci. *114*, E2077–E2085. 10.1073/pnas.1619581114.

88. Breaker, R.R. (2012). Riboswitches and the RNA World. Cold Spring Harb. Perspect. Biol. *4*. 10.1101/cshperspect.a003566.

89. Reynolds, R., and Chamberlin, M.J. (1992). Parameters affecting transcription termination by Escherichia coli RNA: II. Construction and analysis of hybrid terminators. J. Mol. Biol. *224*, 53–63. 10.1016/0022-2836(92)90575-5.

90. Peters, J.M., Vangeloff, A.D., and Landick, R. (2011). Bacterial Transcription Terminators: The RNA 3′-End Chronicles. J. Mol. Biol. *412*, 793–813. 10.1016/j.jmb.2011.03.036.

91. Aboul-ela, F., Huang, W., Abd Elrahman, M., Boyapati, V., and Li, P. (2015). Linking aptamer-ligand binding and expression platform folding in riboswitches: prospects for mechanistic modeling and design. Wiley Interdiscip. Rev. RNA *6*, 631–650. 10.1002/wrna.1300.

92. Berens, C., and Suess, B. (2015). Riboswitch engineering — making the all-important second and third steps. Curr. Opin. Biotechnol. *31*, 10–15. 10.1016/j.copbio.2014.07.014.

93. Ceres, P., Garst, A.D., Marcano-Velázquez, J.G., and Batey, R.T. (2013). Modularity of Select Riboswitch Expression Platforms Enables Facile Engineering of Novel Genetic Regulatory Devices. ACS Synth. Biol. *2*, 463–472. 10.1021/sb4000096.

94. Boussebayle, A., Torka, D., Ollivaud, S., Braun, J., Bofill-Bosch, C., Dombrowski, M., Groher, F., Hamacher, K., and Suess, B. (2019). Next-level riboswitch development—implementation of Capture-SELEX facilitates identification of a new synthetic riboswitch. Nucleic Acids Res. *47*, 4883–4895. 10.1093/nar/gkz216.

95. Etzel, M., and Mörl, M. (2017). Synthetic Riboswitches: From Plug and Pray toward Plug and Play. Biochemistry *56*, 1181–1198. 10.1021/acs.biochem.6b01218.

96. Ceres, P., Trausch, J.J., and Batey, R.T. (2013). Engineering modular 'ON' RNA switches using biological components. Nucleic Acids Res. *41*, 10449–10461. 10.1093/nar/gkt787.

97. Wachsmuth, M., Findeiss, S., Weissheimer, N., Stadler, P.F., and Morl, M. (2013). De novo design of a synthetic riboswitch that regulates transcription termination. Nucleic Acids Res. *41*, 2541–2551. 10.1093/nar/gks1330.

98. Yu, D., and Breaker, R.R. (2020). A bacterial riboswitch class senses xanthine and uric acid to regulate genes associated with purine oxidation. RNA *26*, 960–968. 10.1261/rna.075218.120.

99. Ren, A., Wang, X.C., Kellenberger, C.A., Rajashankar, K.R., Jones, R.A., Hammond, M.C., and Patel, D.J. (2015). Structural basis for molecular discrimination by a 3′,3′-cGAMP sensing riboswitch. Cell Rep. *11*, 1–12. 10.1016/j.celrep.2015.03.004.

100. Keller, H., Weickhmann, A.K., Bock, T., and Wöhnert, J. (2018). Adenine protonation enables cyclic-di-GMP binding to cyclic-GAMP sensing riboswitches. RNA *24*, 1390–1402. 10.1261/rna.067470.118.

101. Moon, M.H., Hilimire, T.A., Sanders, A.M., and Schneekloth, J.S. (2018). Measuring RNA–Ligand Interactions with Microscale Thermophoresis. Biochemistry *57*, 4638–4643. 10.1021/acs.biochem.7b01141.

102. Hedstrom, L. (2009). IMP Dehydrogenase: Structure, Mechanism, and Inhibition. Chem. Rev. *109*, 2903–2928. 10.1021/cr900021w.

103. Ballut, L., Violot, S., Kumar, S., Aghajari, N., and Balaram, H. (2023). GMP Synthetase: Allostery, Structure, and Function. Biomolecules *13*, 1379. 10.3390/biom13091379.

104. Abrams, R., and Bentley, M. (1959). Biosynthesis of nucleic acid purines. III. Guanosine 5′-phosphate formation from xanthosine 5′-phosphate and l-glutamine. Arch. Biochem. Biophys. *79*, 91–110. 10.1016/0003-9861(59)90383-2.

105. Hamal Dhakal, S., Kavita, K., Panchapakesan, S.S.S., Roth, A., and Breaker, R.R. (2023). 8-oxoguanine riboswitches in bacteria detect and respond to oxidative DNA damage. Proc. Natl. Acad. Sci. *120*, e2307854120. 10.1073/pnas.2307854120.

106. Kellenberger, C.A., Wilson, S.C., Hickey, S.F., Gonzalez, T.L., Su, Y., Hallberg, Z.F., Brewer, T.F., Iavarone, A.T., Carlson, H.K., Hsieh, Y.-F., *et al.* (2015). GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP. Proc. Natl. Acad. Sci. U. S. A. *112*, 5383–5388. 10.1073/pnas.1419328112.

107. Tan, Z., Chan, C.H., Maleska, M., Jara, B.B., Lohman, B.K., Ricks, N.J., Bond, D.R., and Hammond, M.C. (2022). The Signaling Pathway That cGAMP Riboswitches Found: Analysis and Application of Riboswitches to Study cGAMP Signaling in Geobacter sulfurreducens. Int. J. Mol. Sci. *23*, 1183. 10.3390/ijms23031183.

108. Winkler, W.C., and Breaker, R.R. (2003). Genetic Control by Metabolite-Binding Riboswitches. ChemBioChem *4*, 1024–1032. 10.1002/cbic.200300685.

109. Roth, A., and Breaker, R.R. (2009). The Structural and Functional Diversity of Metabolite-Binding Riboswitches. Annu. Rev. Biochem. *78*, 305–334. 10.1146/annurev.biochem.78.070507.135656.

110. Weinberg, Z., Nelson, J.W., Lünse, C.E., Sherlock, M.E., and Breaker, R.R. (2017). Bioinformatic analysis of riboswitch structures uncovers variant classes with altered ligand specificity. Proc. Natl. Acad. Sci., 201619581. 10.1073/pnas.1619581114.

111. Hengge, R., Gründling, A., Jenal, U., Ryan, R., and Yildiz, F. (2016). Bacterial Signal Transduction by Cyclic Di-GMP and Other Nucleotide Second Messengers. J. Bacteriol. *198*, 15–26. 10.1128/JB.00331-15.

112. Cancino-Diaz, M.E., Guerrero-Barajas, C., Betanzos-Cabrera, G., and Cancino-Diaz, J.C. (2023). Nucleotides as Bacterial Second Messengers. 10.20944/preprints202310.1900.v1.

113. Hengge, R., Pruteanu, M., Stülke, J., Tschowri, N., and Turgay, K. (2023). Recent advances and perspectives in nucleotide second messenger signaling in bacteria. microLife *4*, uqad015. 10.1093/femsml/uqad015.

114. D'Argenio, D.A., and Miller, S.I. (2004). Cyclic di-GMP as a bacterial second messenger. Microbiology *150*, 2497–2502. 10.1099/mic.0.27099-0.

115. Corrigan, R.M., and Gründling, A. (2013). Cyclic di-AMP: another second messenger enters the fray. Nat. Rev. Microbiol. *11*, 513–524. 10.1038/nrmicro3069.

116. Commichau, F.M., Dickmanns, A., Gundlach, J., Ficner, R., and Stülke, J. (2015). A jack of all trades: the multiple roles of the unique essential

second messenger cyclic di-AMP. Mol. Microbiol. *97*, 189–204. 10.1111/mmi.13026.

117. Nelson, J.W., Sudarsan, N., Furukawa, K., Weinberg, Z., Wang, J.X., and Breaker, R.R. (2013). Riboswitches in eubacteria sense the second messenger c-di-AMP. Nat Chem Biol *9*, 834–839.

118. Smith, K.D., and Strobel, S.A. (2011). Interactions of the c-di-GMP riboswitch with its second messenger ligand. Biochem. Soc. Trans. *39*, 647–651. 10.1042/BST0390647.

119. Smith, K.D., Shanahan, C.A., Moore, E.L., Simon, A.C., and Strobel, S.A. (2011). Structural basis of differential ligand recognition by two classes of bis-(3′-5′)-cyclic dimeric guanosine monophosphate-binding riboswitches. Proc. Natl. Acad. Sci. *108*, 7757–7762. 10.1073/pnas.1018857108.

120. Launer-Felty, K.D., and Strobel, S.A. (2018). Enzymatic synthesis of cyclic dinucleotide analogs by a promiscuous cyclic-AMP-GMP synthetase and analysis of cyclic dinucleotide responsive riboswitches. Nucleic Acids Res. *46*, 2765–2776. 10.1093/nar/gky137.

121. Shanahan, C.A., and Strobel, S.A. (2012). The bacterial second messenger c-di-GMP: probing interactions with protein and RNA binding partners using cyclic dinucleotide analogs. Org. Biomol. Chem. *10*, 9113–9129. 10.1039/c2ob26724a.

122. Smith, K.D., Lipchock, S.V., Livingston, A.L., Shanahan, C.A., and Strobel, S.A. (2010). Structural and Biochemical Determinants of Ligand Binding by the c-di-GMP Riboswitch,. Biochemistry *49*, 7351–7359. 10.1021/bi100671e.

123. Smith, K.D., Lipchock, S.V., and Strobel, S.A. (2012). Structural and biochemical characterization of linear dinucleotide analogues bound to the c-di-GMP-I aptamer. Biochemistry *51*, 425–432. 10.1021/bi2016662.

124. Shanahan, C.A., Gaffney, B.L., Jones, R.A., and Strobel, S.A. (2011). Differential analogue binding by two classes of c-di-GMP riboswitches. J. Am. Chem. Soc. *133*, 15578–15592. 10.1021/ja204650q.

125. Kulshina, N., Baird, N.J., and Ferré-D'Amaré, A.R. (2009). Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch. Nat. Struct. Mol. Biol. *16*, 1212–1217. 10.1038/nsmb.1701.

126. Whiteley, A.T., Eaglesham, J.B., Mann, C.C. de O., Morehouse, B.R., Lowey, B., Nieminen, E.A., Danilchanka, O., King, D.S., Lee, A.S.Y., Mekalanos, J.J.,

*et al.* (2019). Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. Nature *567*, 194–199. 10.1038/s41586-019-0953-5.

127. Ryjenkov, D.A., Simm, R., Römling, U., and Gomelsky, M. (2006). The PilZ domain is a receptor for the second messenger c-di-GMP: the PilZ domain protein YcgR controls motility in enterobacteria. J. Biol. Chem. *281*, 30310–30314. 10.1074/jbc.C600179200.

128. Burroughs, A.M., Zhang, D., Schäffer, D.E., Iyer, L.M., and Aravind, L. (2015). Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. Nucleic Acids Res. *43*, 10633–10654. 10.1093/nar/gkv1267.

129. Li, F., Cimdins, A., Rohde, M., Jänsch, L., Kaever, V., Nimtz, M., and Römling, U. (2019). DncV Synthesizes Cyclic GMP-AMP and Regulates Biofilm Formation and Motility in Escherichia coli ECOR31. mBio *10*, 10.1128/mbio.02492-18. 10.1128/mbio.02492-18.

130. He, L., Kierzek, R., SantaLucia, J.Jr., Walter, A.E., and Turner, D.H. (1991). Nearest-neighbor parameters for G.U mismatches: 5′GU3′/3′UG5′ is destabilizing in the contexts CGUG/GUGC, UGUA/AUGU, and AGUU/UUGA but stabilizing in GGUC/CUGG. Biochemistry *30*, 11005–11192. 10.1021/bi00110a015.

131. Davis, A.R., and Znosko, B.M. (2008). Thermodynamic Characterization of Naturally Occurring RNA Single Mismatches with G-U Nearest Neighbors. Biochemistry *47*, 10178–10187. 10.1021/bi800471z.

132. Chen, J.L., Dishler, A.L., Kennedy, S.D., Yildirim, I., Liu, B., Turner, D.H., and Serra, M.J. (2012). Testing the Nearest Neighbor Model for Canonical RNA Base Pairs: Revision of GU Parameters. Biochemistry *51*, 3508–3522. 10.1021/bi3002709.

133. Wilson, K.S., and Hippel, P.H. von (1995). Transcription termination at intrinsic terminators: the role of the RNA hairpin. Proc. Natl. Acad. Sci. *92*, 8793–8797.

134. Larson, M.H., Greenleaf, W.J., Landick, R., and Block, S.M. (2008). Applied Force Reveals Mechanistic and Energetic Details of Transcription Termination. Cell *132*, 971–982. 10.1016/j.cell.2008.01.027.

135. Weigand, J.E., Gottstein-Schmidtke, S.R., Demolli, S., Groher, F., Duchardt-Ferner, E., Wöhnert, J., and Suess, B. (2014). Sequence Elements Distal to the Ligand Binding Pocket Modulate the Efficiency of a Synthetic Riboswitch. ChemBioChem *15*, 1627–1637. 10.1002/cbic.201402067.

136. Appasamy, S.D., Ramlan, E.I., and Firdaus-Raih, M. (2013). Comparative Sequence and Structure Analysis Reveals the Conservation and Diversity of Nucleotide Positions and Their Associated Tertiary Interactions in the Riboswitches. PLOS ONE *8*, e73984. 10.1371/journal.pone.0073984.

137. Duchardt-Ferner, E., Gottstein-Schmidtke, S.R., Weigand, J.E., Ohlenschläger, O., Wurm, J.-P., Hammann, C., Suess, B., and Wöhnert, J. (2016). What a Difference an OH Makes: Conformational Dynamics as the Basis for the Ligand Specificity of the Neomycin-Sensing Riboswitch. Angew. Chem. Int. Ed. *55*, 1527–1530. 10.1002/anie.201507365.

138. Chyży, P., Kulik, M., Re, S., Sugita, Y., and Trylska, J. (2021). Mutations of N1 Riboswitch Affect its Dynamics and Recognition by Neomycin Through Conformational Selection. Front. Mol. Biosci. *8*. https://doi.org/10.3389/fmolb.2021.633130.

139. Doucet, N., Watt, E.D., and Loria, J.P. (2009). The Flexibility of a Distant Loop Modulates Active Site Motion and Product Release in Ribonuclease A. Biochemistry *48*, 7160–7168. 10.1021/bi900830g.

140. Moesta, A.K., Norman, P.J., Yawata, M., Yawata, N., Gleimer, M., and Parham, P. (2008). Synergistic Polymorphism at Two Positions Distal to the Ligand-Binding Site Makes KIR2DL2 a Stronger Receptor for HLA-C Than KIR2DL31. J. Immunol. *180*, 3969–3979. 10.4049/jimmunol.180.6.3969.

141. Stadinski, B.D., Trenh, P., Duke, B., Huseby, P.G., Li, G., Stern, L.J., and Huseby, E.S. (2014). Effect of CDR3 Sequences and Distal V Gene Residues in Regulating TCR–MHC Contacts and Ligand Specificity. J. Immunol. *192*, 6071–6082. 10.4049/jimmunol.1303209.

142. Hedstrom, L., Szilagyi, L., and Rutter, W.J. (1992). Converting Trypsin to Chymotrypsin: The Role of Surface Loops. Science *255*, 1249–1253. 10.1126/science.1546324.

143. Allemann, R.K., Presnell, S.R., and Benner, S.A. (1991). A hybrid of bovine pancreatic ribonuclease and human angiogenin: an external loop as a module controlling substrate specificity? Protein Eng. Des. Sel. *4*, 831–835. 10.1093/protein/4.7.831.

144. Kellenberger, C.A., Wilson, S.C., Sales-Lee, J., and Hammond, M.C. (2013). RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messengers Cyclic di-GMP and Cyclic AMP-GMP. J. Am. Chem. Soc. *135*, 4906–4909. 10.1021/ja311960g.

145. Palmer, A.E., Qin, Y., Park, J.G., and McCombs, J.E. (2011). Design and application of genetically encoded biosensors. Trends Biotechnol. *29*, 144–152. 10.1016/j.tibtech.2010.12.004.

146. Paige, J.S., Nguyen-Duc, T., Song, W., and Jaffrey, S.R. (2012). Fluorescence Imaging of Cellular Metabolites with RNA. Science *335*, 1194–1194. 10.1126/science.1218298.

147. Lindenburg, L., and Merkx, M. (2014). Engineering Genetically Encoded FRET Sensors. Sensors *14*, 11691–11713. 10.3390/s140711691.

148. Sanford, L., and Palmer, A. (2017). Chapter One - Recent Advances in Development of Genetically Encoded Fluorescent Sensors. In Methods in Enzymology Enzymes as Sensors., R. B. Thompson and C. A. Fierke, eds. (Academic Press), pp. 1–49. 10.1016/bs.mie.2017.01.019.

149. Boussebayle, A., Groher, F., and Suess, B. (2019). RNA-based Capture-SELEX for the selection of small molecule-binding aptamers. Methods *161*, 10–15. 10.1016/j.ymeth.2019.04.004.

150. Truong, L., Kooshapur, H., Dey, S.K., Li, X., Tjandra, N., Jaffrey, S.R., and Ferré-D’Amaré, A.R. (2022). The fluorescent aptamer Squash extensively repurposes the adenine riboswitch fold. Nat. Chem. Biol. *18*, 191–198. 10.1038/s41589-021-00931-2.

151. Shanahan, C.A., Gaffney, B.L., Jones, R.A., and Strobel, S.A. (2011). Differential analogue binding by two classes of c-di-GMP riboswitches. J. Am. Chem. Soc. *133*, 15578–15592. 10.1021/ja204650q.

152. Hallberg, Z.F., Wang, X.C., Wright, T.A., Nan, B., Ad, O., Yeo, J., and Hammond, M.C. (2016). Hybrid promiscuous (Hypr) GGDEF enzymes produce cyclic AMP-GMP (3’, 3’-cGAMP). Proc. Natl. Acad. Sci., 201515287. 10.1073/pnas.1515287113.

153. Pollock, A.J., Choi, P.H., Zaver, S.A., Tong, L., and Woodward, J.J. (2021). A rationally designed c-di-AMP FRET biosensor to monitor nucleotide dynamics. 2021.02.10.430713. 10.1101/2021.02.10.430713.

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

155. Horning, D.P., and Joyce, G.F. (2016). Amplification of RNA by an RNA polymerase ribozyme. Proc. Natl. Acad. Sci. *113*, 9786–9791. 10.1073/pnas.1610103113.

156. Tjhung, K.F., Shokhirev, M.N., Horning, D.P., and Joyce, G.F. (2020). An RNA polymerase ribozyme that synthesizes its own ancestor. Proc. Natl. Acad. Sci. *117*, 2906–2913. 10.1073/pnas.1914282117.

157. Markley, A.L., Begemann, M.B., Clarke, R.E., Gordon, G.C., and Pfleger, B.F. (2015). Synthetic Biology Toolbox for Controlling Gene Expression in the Cyanobacterium Synechococcus sp. strain PCC 7002. ACS Synth. Biol. *4*, 595–603. 10.1021/sb500260k.

158. Hall, B., Micheletti, J.M., Satya, P., Ogle, K., Pollard, J., and Ellington, A.D. (2009). Design, synthesis, and amplification of DNA pools for in vitro selection. Curr. Protoc. Mol. Biol. *Chapter 24*, Unit 24.2. 10.1002/0471142727.mb2402s88.

159. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12. 10.14806/ej.17.1.200.

160. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. 10.1038/nmeth.1923.

161. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008). The Vienna RNA Websuite. Nucleic Acids Res. *36*, W70–W74. 10.1093/nar/gkn188.

162. Yao, Z., Weinberg, Z., and Ruzzo, W.L. (2006). CMfinder—a covariance model based RNA motif finding algorithm. Bioinformatics *22*, 445–452. 10.1093/bioinformatics/btk008.

163. Griffiths-Jones, S. (2005). RALEE—RNA ALignment Editor in Emacs. Bioinformatics *21*, 257–259. 10.1093/bioinformatics/bth489.

164. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLOS ONE *5*, e9490. 10.1371/journal.pone.0009490.

165. Hartmann, R.K. ed. (2009). Handbook of RNA biochemistry 1st student ed. (Weinheim: Wiley-VCH).

166. Simons, R.W., Houman, F., and Kleckner, N. (1987). Improved single and multicopy lac-based cloning vectors for protein and operon fusions. Gene *53*, 85–96. 10.1016/0378-1119(87)90095-3.

167. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of

Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. *2*, 2006.0008. 10.1038/msb4100050.